FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Uncovering and Comparing Large-Scale Art History Narratives in Biographical Datasets

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Dipl.-Ing. Doron Goldfarb
Matrikelnummer 9827080

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Mag. Dr. Dieter Merkl

Diese Dissertation haben begutachtet:

Prof. Vivien Petras, PhD	Ao.Univ.Prof. Dipl.-Ing. Dr. Andreas Rauber

Wien, 13. Oktober 2020

Doron Goldfarb

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Uncovering and Comparing Large-Scale Art History Narratives in Biographical Datasets

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Technischen Wissenschaften

by

Dipl.-Ing. Doron Goldfarb
Registration Number 9827080

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.Prof. Mag. Dr. Dieter Merkl

The dissertation has been reviewed by:

Prof. Vivien Petras, PhD	Ao.Univ.Prof. Dipl.-Ing. Dr. Andreas Rauber

Vienna, 13th October, 2020

Doron Goldfarb

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Doron Goldfarb
Garnisongasse 1/31, 1090 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 13. Oktober 2020

Doron Goldfarb

Danksagung

Meinem Betreuer Prof. Dieter Merkl gebührt mehrfacher Dank. Einerseits für die Gelegenheit, mich im Rahmen des VSEM Projektes intensiv mit dem Themenbereich “Digitales Kulturerbe” auseinandersetzen zu können. Dann für die große Freiheit bei der Erarbeitung meiner Fragestellungen und das stets wohlwollende Feedback, das mich immer konstruktiv dazu angeregt hat, über meine Ideen zu reflektieren und meinen Weg zu finden. Schließlich auch für die langjährige Betreuung, die große Geduld und die vielen Gespräche, bei denen ich viel gelernt habe, sowohl fachlich als auch persönlich. Meinen Kollegen aus dem VSEM Projekt, Max Arends, Josef Froschauer und Martin Weingartner, danke ich für die spannende gemeinsame Zeit und für die Möglichkeit, meine Fragestellungen und Zugänge aus ihren Blickwinkeln zu betrachten. Dem International Office der TU Wien danke ich für die finanzielle Unterstützung für die Teilnahme an Konferenzen.

Prof. David Quigley, Stephen Mathewson, Sebastian Koch, Katja Mayer, Armin Medosch und der Technopolitics Arbeitsgruppe danke ich für ihr Feedback und ihr Interesse aus der Perspektive der Künstler, Sozialwissenschaftler, Kulturtheoretiker und Kunsthistoriker. Durch sie konnte ich über den Tellerrand der Informatik hinausblicken, was mir wichtige Impulse gegeben und weitergehende Kooperationen ermöglicht hat.

Prof. Maximilian Schich und der University of Texas at Dallas danke ich für die Möglichkeit, ein Semester an der School of Arts, Technology, and Emerging Communication (ATEC) zu verbringen, wo ich eine interdisziplinäre, konstruktive und motivierende Umgebung vorfand, meine Arbeit weiterzuentwickeln. Prof. Reynaldo Thompson für die Einladung, meine Arbeit an der Universität von Guanajuato, Mexiko zu präsentieren.

Meiner Lebensgefährtin Susanne Schatz danke ich für die große, inhaltliche wie persönliche, Unterstützung. Ihr Vertrauen und die große Motivation waren außerordentlich wichtig für mich. Meinen Eltern Michael und Nomi Goldfarb sowie ihren LebensgefährtInnen Adelheid Goldfarb und Götz Fritsch danke ich dafür, dass sie immer an mich geglaubt haben, sowie für die immerwährende Unterstützung und die Möglichkeit, meinen Interessen nachzugehen. Meinem Freund Dominik Walch danke ich für die vielen wertvollen Gespräche und das Interesse an meiner Arbeit.

Prof. Vivien Petras und Prof. Andreas Rauber danke ich abschließend für ihre Bereitschaft, meine Dissertation zu begutachten.

Kurzfassung

Die Initiativen der letzten Jahre, institutionelle Sammlungen von Digitalisaten von Objekten des kulturellen Erbes und dazugehörende Metadaten der Öffentlichkeit zugänglich zu machen, treffen auf stetig größer werdende Online-Sammlungen von nutzergenerierten Inhalten. Viele Aspekte der veröffentlichten Kulturerbedaten, wie zum Beispiel Beziehungen zwischen Personen, können als Netzwerke aufgefasst und abgebildet werden. Ihre Analyse und Visualisierung kann für verschiedene Szenarien von Nutzen sein, von der Navigation durch Online-Kunstsammlungen hin zu der Erforschung von strukturellen Aspekten der Netzwerke selbst. Der Vergleich der Inhalte von institutionell gepflegten mit nutzergenerierten Datensammlungen stellt hier ein besonders interessantes Szenario für den Einsatz visueller und anderer analytischer Methoden dar.

Geleitet von diesen Entwicklungen, ist das Ziel der vorliegenden Arbeit die Erforschung von Datensätzen über Personennetzwerke aus dem Bereich Kunstgeschichte. Motiviert von historischen Beispielen, kunstgeschichtliche Entwicklungen in visueller, diagrammatischer Form darzustellen, folgt die Arbeit den Fragestellungen, 1) ob und wie datengetriebene Netzwerkvisualisierungen die Kontextualisierung von Kunstwerken in virtuellen Präsentationen unterstützen, 2) ob die Betrachtung der Gesamtheit der in einer Datensammlung gespeicherten, individuellen historischen Personenbeziehungen kunsthistorische Narrative erkennen lässt und welche das sind sowie, 3) ob und wie solche Narrative sich zwischen verschiedenen Datensammlungen unterscheiden. Die Integration von Informationen über Personenbeziehungen aus einer bestehenden institutionellen Datensammlung, der Getty Union List of Artist Names (ULAN), mit Daten über Kunstwerke aus der Web Gallery of Art (WGA) ermöglicht die Erzeugung einer virtuellen 3D Galerieumgebung, deren Grundriss auf einem automatischen Netzwerk-Layout basiert. Die Evaluation mit Experten zeigt, dass der Zugang eine interessante und neuartige Möglichkeit der Erkundung von Kunstgeschichte im digitalen Raum bietet. Die allgemeine Analyse des ULAN Personennetzwerkes zeigt das Vorhandensein einer zusammenhängenden und mehrere Jahrhunderte umfassenden, chronologischen Struktur, deren Vielzahl an abgebildeten historischen Interaktionen in großem Maßstab ein Narrativ ergibt, dass sich zu einem gewissen Grad mit entsprechenden wissenschaftlichen Sichten deckt. Der Vergleich mit ähnlichen Inhalten aus Wikipedia Biographien zeigt eine bedeutende Übereinstimmung bezüglich der dort abgebildeten Personenbeziehungen, sowie strukturelle Gemeinsamkeiten und Unterschiede zwischen dem ULAN Netzwerk und entsprechenden Hyperlink-Netzwerken in verschiedenen Wikipedia Sprachversionen. Deren separater Vergleich untereinander zeigt wiederum klare individuelle Präferenzen für die der jeweiligen Sprache kulturell

nahestehenden Biographien auf, legt gleichzeitig jedoch auch grundsätzliche Übereinstimmung hinsichtlich der fundamentalen Entwicklungen westlicher Kunst nahe. Wikidata wird als alternative Möglichkeit zur Identifikation kunsthistorisch relevanter Biographien und anderer Wikipedia Artikel vorgestellt, demonstriert anhand der Analyse, Verarbeitung und Visualisierung eines bi-partiten Netzwerkes von Biographien und Artikeln über Kunst- und Architekturstile. Abschließend wird eine weitere Methode der Identifikation von bereichsspezifischen Biographien in Wikidata vorgestellt, die auf der Analyse und dem Clustering von gleichzeitig zugeordneten Berufen in Wikidata Personendatensätzen beruht. Dies wird anhand eines weiteren Künstlernetzwerkes sowie der Visualisierung von Personennetzwerken anderer Berufsgruppen demonstriert.

Die wichtigsten Beiträge dieser Arbeit umfassen verschiedene Aspekte. Einerseits wird demonstriert, dass ausreichend große Netzwerke von Personen aus der Kunstgeschichte sowohl in professionell kuratierten als auch in nutzergenerierten Datensammlungen eine zusammenhängende und chronologische "bottom-up" Struktur aufweisen, deren visuelle Darstellung wesentliche Entwicklungen in der Kunstgeschichte aufzeigt und zum Teil mit entsprechenden wissenschaftlichen Visualisierungen in historischen Diagrammen übereinstimmt. Dies zeigt auch, dass individuelle Beiträge von den verschiedenen Datensätzen sich gegenseitig ergänzen, wodurch ihre Kombination zu einer globaleren Sicht auf die Kunstgeschichte führt. Andererseits werden auch über die Unterstützung der kontextualisierten Präsentation von Artefakten hinausgehende Vorteile demonstriert, die aus der Integration von verschiedenen, aber miteinander in Zusammenhang stehenden Kulturerbe-Datensätzen entstehen können. Implizite Information, die aus einer Datenquelle extrahiert wurde, kann beispielsweise als Referenz für die Identifikation von fehlender Information in einer anderen Datenquelle genutzt werden. Die Integration von professionell kuratierten mit ähnlichen, nutzergenerierten Datensammlungen stellt wiederum ein Beispiel dar, bei dem die umfassenden Metadaten der professionell kuratierten Datenquelle dazu genutzt werden, die nutzergenerierte Datensammlung quantitativ zu analysieren sowie mit ihr zu vergleichen. Schließlich werden neue Wege demonstriert, wie bereichsspezifische Personengruppen in allgemeinen Datensammlungen von Personendatensätzen identifiziert werden können. Methodologische Beiträge bestehen aus einem Rahmenwerk für die Integration und Analyse des Netzwerkes von Wikipedia Biographien anhand von ULAN Personenattributen, verschiedenen Techniken zur Filterung von Wikipedia Netzwerken für die Offenlegung der darin potentiell eingebetteten chronologischen Struktur, sowie zusätzlichen Methoden für die Extraktion aggregierter Netzwerke von Nationalitäten und Berufen.

Abstract

Recent initiatives to open institutional repositories of digitized cultural heritage artifacts and related metadata meet growing collections of related user-generated content. Many aspects of the cultural heritage information that becomes available, such as person relationships, can be perceived and represented as networks, whose analysis and visualization can serve a variety of scenarios, ranging from navigation across online art collections to the study of structural aspects of the networks themselves in more research-oriented settings. The comparison of institutional and user-generated repositories using visual and other analytical means represents a particularly interesting scenario in this regard.

In light of these developments, the main goal of this work is to explore datasets about person-to-person networks in the context of art history. Motivated by historical examples to represent developments in the arts in visual, diagrammatic form, it seeks to explore 1) if and how data-driven network visualizations can support the contextualization of artworks in virtual presentations, 2) if and which large-scale art history narratives are embedded in extensive network data and 3) if and how they differ across multiple datasets. Social relationship data between person records provided by an existing institutional dataset, the Getty Union List of Artist Names (ULAN), are integrated with data about artworks taken from the Web Gallery of Art (WGA) in order to create a virtual 3D art gallery environment using an automatic network layout as floor plan. Evaluation with domain experts shows that the approach provides an interesting and novel way to explore art history in the digital realm. An overall analysis of the ULAN person-to-person network reveals a contiguous chronological structure spanning multiple centuries, whose multitude of historical interactions yields a large-scale narrative that correlates with related scholarly views to a certain extent. The comparison with similar content derived from Wikipedia biographies reveals significant overlap regarding the featured person-to-person relationships as well as structural commonalities and differences between the ULAN network and corresponding hyperlink networks in various Wikipedia language versions, whose separate comparison reveals cultural self-focus bias regarding the coverage of art history biographies but also basic agreement on the fundamental developments in Western art. Wikidata is introduced as an alternative means to identify biographies and other Wikipedia articles relevant to the domain of art history, demonstrated by the analysis, processing and visualization of a bi-partite network of articles about persons and art and architecture styles. Clustering co-occurring occupations in Wikidata persons records is eventually introduced as additional approach to identify domain-specific biographies,

demonstrated by the example of an additional artist network and the visualization of networks from other domains.

The main contributions of this work address different aspects. One the one hand, it demonstrates that large-scale biographical networks extracted from both professional as well as crowd-sourced datasets possess a bottom-up contiguous and chronological structure whose visualization reveals major developments across art history, which partly correspond with related scholarly views expressed in historical diagrams. This also reveals that the individual contributions of the different datasets complement each other in a way that their combination yields a more global view on the history of art. On the other hand it demonstrates the benefit of integrating different but related cultural heritage datasets beyond the goal of contextualized presentation of artifacts, such as using implicit information extracted from one data source as reference data for the identification of missing information in another. The integration of professionally curated datasets with similar, crowd-sourced data provides another showcase in this regard, demonstrating how rich metadata present in the professionally curated resource can be used to quantitatively study and compare the structure of its crowd-sourced counterpart. Last but not least, this work demonstrates new ways to identify domain-specific groups of persons in general-purpose collections of person records. Methodological contributions include the framework to integrate and analyze the network of Wikipedia biographies with ULAN person attributes, different techniques to filter the Wikipedia networks for uncovering their chronological structure and additional methods to extract aggregate networks of nationalities and occupations.

Contents

CHAPTER 1

Introduction

1.1 Digital cultural heritage

In recent years, cultural heritage related datasets have increasingly been made available to the public, including both catalog metadata as well as digitized content. A number of reasons stand behind this development. As pointed out by Terras in [Terras, 2015], activities surrounding the open access movement in academia have been one driving factor, leading to initiatives such as OpenGLAM[1] (GLAM standing for galleries, libraries, archives and museums) which actively encourage cultural heritage institutions to open up their digitized collections to the public. The aim has been to foster content access and reuse in both academic research as well as in commercial settings such as the creative-industries, paralleled by legislative efforts e.g. by the European Commission, whose PSI directive[2] was established to lower the access barrier to publicly owned data collections in order to drive innovation. European considerations and activities regarding the role of digital cultural heritage in the context of the PSI, but also the vision of providing central access to Europe's diverse cultural heritage, were outlined in a report by the European Commission [Niggemann et al., 2011]. Europeana[3], the portal to Europe's digitized cultural heritage, is one of the premier outcomes in this regard, currently offering access to more than 58 Million[4] digitized cultural heritage items from about 4,000 institutions including artworks, books and other media. As a rare occasion at least in the context of the digital realm, these European endeavors also caused effects on the other side of the Atlantic ocean, where the Digital Public Library of America[5] (DPLA) represents an effort similar to Europeana, aggregating content from an increasing number of U.S. cultural

[1] https://openglam.org/, retrieved Sept., 8th, 2020
[2] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information, retrieved Sept., 8th, 2020
[3] https://www.europeana.eu/, retrieved Sept., 8th, 2020
[4] https://pro.europeana.eu/about-us/mission, retrieved Sept., 8th, 2020
[5] https://dp.la/, retrieved Sept., 8th, 2020

heritage institutions. In contrast to Europeana, DPLA was initially based on a private initiative, described by Darnton in [Darnton, 2013], suggesting that the drive towards opening up collections also stems from the institutions themselves to some extent, not necessarily imposed on them via legislation.

Different traditions in CH institutions

Such shifts in institutional policies can be seen as a reaction to a general necessity to redefine cultural heritage institutions' roles in an increasingly digitized and networked world. As discussed by Trant in [Trant, 2009a], however, different types of CH institutions have very different traditions regarding their collections, related audiences and users, and thus very different starting points in this regard. One main distinction is due to the nature of collection items, usually being unique in museums and archives but not in libraries (besides specialized ones dedicated to rare books and manuscripts). Due to this distinction, museums usually present their collections to the public in a highly curated, narrative manner and libraries are in turn more specialized in offering means of efficient access to their whole collection, already having a long tradition of data sharing and public online access. Faced with the transition to the digital and especially to the Web, the differences between these institutions become more blurred, Trant identified a number of common challenges for them. Besides issues regarding digitization and digital collections management, this also includes questions regarding the presentation and use of the resulting digital resources. The latter is arguably a bigger challenge to museums and archives, since by offering online access to their full collection, they would in some way have to give up their monopoly on choosing the narrative within which their items are usually presented. On the other hand, this allows third parties to use these previously unavailable resources to identify new creative uses for them.

Museums in the realm of the virtual

MacDonalds and Alsford already in 1991 suggested to conceive the mission of museums as rather being information providers instead of mere custodians of physical artifact collections, with the aim of ultimately becoming "[...] 'information utilities' [...] available in every home." [MacDonald and Alsford, 1991]. The emergence of the World Wide Web in the early 1990's accelerated related developments and museums started to explore the new medium in a number of different ways and for different purposes. This on the one hand included means to support the visit at the actual "brick and mortar" museum, ranging from simple "how to find us" pages to complex visitor engagement during pre-, on- and post-visit, in later years increasingly also via social media platforms. On the other hand, there was an increasing tendency towards opening up online access to the museum collections by providing searching and/or browsing facilities via the museum Website, built on top of internal collection databases. Providing broad categories for the different approaches, Schweibenz in [Schweibenz, 2004] referred to online sites supporting the physical visit as the "brochure museum", while online means to explore the collections

were distinguished between the "content museum" and "learning museum". Content museums are object-oriented and to a large extent directly reflect the organization of the underlying collection database, they are thus more targeted at users who know what they are looking for. The philosophy of the learning museum is in turn to present the artifacts in a more context-oriented way, targeted at users who are not necessarily familiar with the content, supporting them via different points of access and by didactically enhancing the online visit via links to additional information. The exploration of previously unknown content can also be referred to as "exploratory search", outlined by White and Roth in [White and Roth, 2009] as alternative to traditional "query-response" means of searching online content. According to the authors, exploratory search consists of two main activities "exploratory browsing" and "focused searching" which occur in sequence, users are expected to gradually move from browsing to searching with increasing knowledge about the search domain. This view blends well with Schweibenz' notion of the content and the learning museum, where the former is associated with focused searching and the latter with exploratory browsing.

Browsing was defined by Kwasnik in [Kwasnik, 1992] as "movement in connected space", where the notion of "connected space" encompasses a variety of virtual settings such as pages connected via hyperlinks, image sequences in a digital library or pathways in a virtual world. Such diverse means for exploratory browsing through museum collections are fundamental to the idea of the "virtual museum", an essentially non-physical space populated by digital surrogates of cultural heritage artifacts whose arrangement can, in contrast to real world museums, in principle be constantly rearranged and reconfigured. A virtual museum should ideally be both a content as well as a learning museum, offering direct but also context based access to the virtual objects. As far as surrogates to the actual objects are concerned, the idea of a virtual museum already emerged before the onset of electronic media, mainly through the invention and proliferation of photography. Famous examples for the use of this medium as means to transcend physical arrangements of museum artifacts are the 63 panels of Aby Warburg's "Mnemosyne Atlas", discussed by Zöllner in [Zöllner, 2010], and André Malraux's "Musée Imaginaire", the "Museum without walls" [Malraux, 1949]. Even more than their analog predecessors, digital virtual museums enable to completely leave behind the constraints of the physical world[6], allowing to arrange a potentially unlimited collection of digitized surrogates in a wide variety of settings in which virtual visitors can choose their own paths across the provided content.

According to Huhtamo in [Huhtamo, 2002], pioneering approaches to virtual museums in the early 1990's were strongly driven by the then emerging Hypertext paradigm, although at that time mainly published in form of CD-Roms. While only few of these approaches sought to provide 3D-simulations of physical premises, the majority presented carefully narrated presentations of collection highlights. Around the year 2000, increasing bandwidth and technological means finally led museums to experiment with means to providing access to their collections online, Cameron in [Cameron, 2003] described different generations of related attempts. While the first steps in this direction were still very much based on presentations of digital surrogates of museum objects in relatively

[6] Besides physical limitations to computing power, memory, storage and bandwidth

fixed sequences using "traditional museum devices such as object labels, images and didactic text panels [...]" [Cameron, 2003], later approaches already acknowledged that users might want to explore collections according to individual pathways between the featured objects via metadata connections which were sometimes also directly rendered as semantic map on the user screen, supporting navigation.

2D browser based environments were soon also accompanied by 3D approaches such as virtual walkthroughs e.g. using Quicktime VR or VRML. While the former usually provided virtual walkthroughs through existing exhibition spaces, the latter also opened up possibilities to create purely virtual environments. 3D multi user environments such as Second Life, described by Hazan in [Hazan, 2010], extended the available possibilities by providing the opportunity for mutual interaction between virtual visitors, such as for example on the — now disbanded — Dresden gallery island[7] offering a full 3D replica of the historical building and its complete image gallery.

Another recent development is the application of methods from information visualization to exploring cultural heritage collections online, with the aim of providing Web visitors with means to encounter new content in serendipitous ways. Such views on information seeking behavior call for new frameworks for understanding user needs in this regard, such as the concept of the "Information Flaneur" sketched by Dörk et al. in [Dörk et al., 2011]. With this concept, the authors sought to turn the search scenario from a "deficiency-oriented" approach, characterized by rather negative notions such as "search problems" and "knowledge gaps" towards a more positive view on information seeking based on curiosity and open mindedness.

Besides desktop based means of exploring CH content online, mobile approaches are usually realized on handheld devices such as Smartphones or Tablet computers. Their often present GPS functionality can be used to present cultural heritage artifacts "in the field", outside of institutions and potentially in their historical geographical context. There is, however, also a long tradition of handheld devices within institutions, ranging from the well established Audioguide to multimedia guides serving as companion for museum visitors. Augmented reality environments represent the most recent development in the mobile domain, potentially fusing virtual with real experiences.

While developments such as the ones outlined above strongly contribute to the notion of Schweibenz' "learning museum", only the highest level in his categorization of different approaches to online museum presences from [Schweibenz, 2004] was reserved to his concept of the virtual museum. For Schweibenz, the virtual museum only becomes reality once it leaves behind the individual institution, once it is able to provide access to the combination of different collections, interlinked through as many contextual links as possible. This can only happen if institutions decide to open up their collections not only via portals on their own Website, but also via allowing direct access to (parts of) the underlying data in a machine readable form.

[7] Dresden Gallery in Second Life: `https://youtu.be/YtNT-ZAH1n4`, retrieved Sept. 8th, 2020

From portal to platform

From a collection data point of view, the presentation of museum collections via platforms such as Second Life can be perceived as the beginning of such a paradigm shift in online museum presences, requiring to move (parts of) the collection data outside of the institutions own IT infrastructure to another, external platform which takes over the task of presentation. While Second life related approaches were mainly based on manual work in this regard, more automated procedures would call for institutions to provide their data to the outside world via dedicated APIs. Sherratt in [Sherratt, 2013] referred to this paradigm shift as "from portal to platform", i.e., institutions do not provide means of discovery by themselves but rather serve as infrastructure for third parties in this regard.

Related considerations, however, already took place much earlier, Bearman in [Bearman, 1995] for example anticipated a near future where broadband-connected users would be catered by information services arranging contextualized information according to their own requirements. He envisioned cultural heritage institutions to be a potentially valuable source of information in this regard and even stressed that providing full access to collections and programs would be vital for institutions to succeed in the 21^{th} century. As fundamental prerequisite for interlinking and combining cross-institutional cultural heritage information, Bearman identified the need for collection and content data standards, which had, besides in the library world, rarely been considered before in a cross-institutional cultural heritage context. Standardized cultural heritage data could be provided via dedicated APIs and re-used in many different ways. This vision was challenged by the extensive heterogeneity of cultural heritage collection information, not only between but also within institutions, where registrars, curators, scientists and visitors usually have quite differing requirements in this regard.

Semantic Web

The idea of the Semantic Web, proposed by the WWW's inventor Berners-Lee in [Berners-Lee et al., 2001] as means to interlink heterogeneous sources of structured information via existing Web infrastructure, one specific implementation proposal for which is well known as the Linked Data approach [Berners-Lee, 2006], appeared as the ideal solution to dealing with the present heterogeneity in cultural heritage data collections. As discussed by Parry et al. a few years later in [Parry et al., 2008], however, the gap between the vision and the reality of the Semantic Web remained large, the authors stated widely different conceptions of it to be one reason in this regard, to a large extent divided between positions calling for a monolithic, unified "hard" approach to its realization where all stakeholders agreed on e.g. which ontologies to use as a foundation versus a distributed "soft" conception of it being a collection of loosely connected "islands" based on individual community or even user-generated solutions. While it must be expected that there will never be the "one and only" Semantic Web, the underlying ideas nevertheless keep resonating and are applied in various settings.

One such application scenario is the semantic enrichment of collection objects with links to contextual entities which can provide added value for searching and browsing, described for example by Manguinhas et al. in [Manguinhas et al., 2016] in the context of Europeana. According to this view, digital cultural heritage objects can be perceived as being embedded within a multi-layered web of knowledge, also called "Web of Data", in which artifacts are linked with persons, places and other physical or abstract entities, which are themselves interlinked amongst each other and ideally also with artifacts from other collections. One example are links from objects to concepts from a thesaurus, the latter themselves related to each other via hierarchical, equivalence or associative links, allowing inferential connections to related objects linked to thesaurus concepts on different but connected positions in its hierarchy. Connections between objects and social networks related to their socio-historical context, with links leading from artworks to their creators who are themselves directly connected via recorded teaching or family ties are another example which enables the contextual navigation through object collections. While the thesaurus links may rather support semantically enhanced search e.g. via inference on broader/narrower terms, links to a snapshot of the historical social network behind an artifact can enhance browsing unknown collections via multiple pathways.

CH objects in context

The contextual view on cultural heritage data requires special data sources that go beyond immediate artifact metadata and the cultural heritage domain has already produced plenty of them. Existing in form of, amongst others, thesauri, taxonomies or authority files, so-called Knowledge Organization Systems (KOS) mainly serve as controlled vocabularies for annotating metadata records with consistent terminology, such as consistent keywords for subject indexing or unique identifiers for related persons. Well known examples from the library domain are the Library of Congress Subject Headings[8] and Name Authority File[9]. As discussed by Zeng and Mayr in [Zeng and Mayr, 2018], KOS play a fundamental role for the idea of Linked Data, providing a standardized "value vocabulary" which, when commonly used, acts as the glue between datasets published on the Web of Data. Some cultural heritage institutions identified this as an opportunity to place their own KOS prominently there, with the aim of supporting the evolution of this emerging data ecosystem, having the welcome side-effect of improving their presence and visibility there. Besides pioneering contributions from the library domain, including the previously mentioned Library of Congress authorities[10] and the "Gemeinsame Normdatei"[11], serving similar purposes for the German speaking library world, this increasingly includes contributions from other cultural heritage domains. One notable example in this regard are the Getty Vocabularies[12] (AAT, TGN, ULAN and

[8] https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html, retr. Sept. 8th, 2020
[9] http://id.loc.gov/authorities/names.html, retrieved Sept. 8th, 2020
[10] https://id.loc.gov/, retrieved Sept. 8th, 2020
[11] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html, retrieved Sept. 8th, 2020
[12] http://www.getty.edu/research/tools/vocabularies/, retrieved Sept. 8th, 2020

CONA), important data sources for digital collection management in the context of art history, maintained by the Getty foundation[13] in Los Angeles and used and contributed to by numerous institutions around the world. The process and challenges of transforming them to and publishing them as Linked Data were described by Baca and Gill in [Baca and Gill, 2015], mentioning the importance of multi-linguality for the success of the Semantic Web and the expected significance of publishing the Getty vocabularies there. The authors also highlighted the internal interlinking of Getty vocabulary records, such as present in their Arts and Architecture Thesaurus (AAT) or as associative links between person records in the Union List of Artist Names (ULAN), representing a snapshot of a historical network of persons relevant to art history, as outlined above, inviting for further exploration.

User-generated content

Besides professionally curated, institutional KOS published as Linked Data, additional contributions originated from the Internet users themselves. So-called user generated content became a topic during the emergence of the Web 2.0 paradigm and is in the meantime considered as an important source of information for cultural heritage related online activities. Trant in [Trant, 2009b] for example described an experiment called steve.museum, where Internet users participated in subject indexing of museum artifacts online using their own terminology. This activity is generally referred to as social tagging and when applied large scale, it can result in so-called folksonomies, i.e. user generated taxonomies with potentially differing views on the particular domain compared to e.g. the Getty AAT. As far as steve.museum was concerned in this regard, a significant "semantic gap" between expert and non-expert terminology was identified via the observation that about 86% of the user assigned terms were not present in existing museum documentation for the same tagged artifacts, suggesting that folksonomies might serve as valuable entry point for non-expert search activities.

Another community driven source of information is Wikipedia. Since its inception in 2001, the free encyclopedia has become one of the most visited sites on the Web (Alexa Rank[14] 12 as of September 2020). Due to its encyclopedic nature, its articles are usually clearly delineated thematically, focusing on specific entities such as persons, places or abstract concepts. Besides serving its original purpose of providing a human readable online encyclopedia, it therefore increasingly also plays an important role as structured knowledge base: DBpedia[15], described by Auer et al. in [Auer et al., 2007] represented the first large scale attempt to extract structured information from Wikipedia and publish it as Linked Data. DBpedia especially focuses on data provided via so-called infoboxes, i.e. templates reused in multiple Wikipedia pages about specific entities such as persons, with the aim to provide structured information for common attributes such as birth/death date and place, occupation etc. As of June 2018, DBpedia provided semantic extracts

[13] Not to be mistaken with the Getty Images stock photography company
[14] `https://www.alexa.com/topsites`, retrieved Sept. 8th, 2020
[15] `https://wiki.dbpedia.org/`, retrieved Sept. 8th, 2020

for about 4.58 Million entities corresponding to articles in the English Wikipedia (based on a Wikipedia data dump from April 2016 consisting of about 5.1 Million[16] articles) as a subset of 38.3 Million entities for 125 language versions in total.

A closely related and more recent contribution is Wikidata, described by Vrandečić and Krötzsch in [Vrandečić and Krötzsch, 2014]. This project aims at providing a database of common facts, such as person birth dates, to the Wikipedia community in order to act as central repository for the integration of such information in multilingual Wikipedia pages, thus potentially reducing ambiguous or conflicting entries in different languages. The process of contributing information to this knowledge base is based on a community effort similar to the one in Wikipedia, and the present content is growing rapidly, contributed by individuals, but increasingly also by institutions, as outlined by a dedicated project page[17]. As of June 2018[18], Wikidata contained about 49.1 Million entities, many of which linking to corresponding Wikipedia articles in one or more languages. Recent initiatives such as a dedicated project to represent structured information about Bulgarian Icons in Wikidata, described by Alexiev et al. in [Alexiev et al., 2020], suggest that this open repository will become an important source for cultural heritage research data in the near future.

Wikipedia articles and their structured counterparts get increasingly interlinked with other related data sources about similar entities. As discussed by Klein and Kyrios in [Klein and Kyrios, 2013], especially the German Wikipedia community has put significant effort into assigning authority control information to Wikipedia articles about persons. This information has been integrated into both DBpedia and Wikidata, the latter has moreover been continuously updated with new external authority links via tools such as Mix'n'Match[19], resulting in an increasing integration of Wikipedia/-data articles and records with related, to a large extent professionally curated, structured information from other sources. Besides getting enriched with links to external entities in other data collections, DBpedia and Wikidata increasingly got referenced from external datasets as well. On the one hand, such as described by Oomen and Aroyo in [Oomen and Aroyo, 2011], Wikipedia links have increasingly been added to cultural heritage collection item records to provide virtual visitors with additional contextual information from the free encyclopedia via articles sometimes created in close cooperation between the institutions' own cultural heritage experts and so-called "Wikipedians in residence". On the other hand, more and more digitized or born digital free text resources - news articles, history textbooks, letters, etc. - have been semantically annotated with links to respective Wikipedia articles in a process called "Wikification", described by Mihalcea and Csomai in [Mihalcea and Csomai, 2007]. This increasing two way interlinking of structured extracts of Wikipedia/-data content with external data sources had the effect that both DBpedia and Wikidata have become important central hubs in the current Web of Open

[16] https://stats.wikimedia.org/EN/TablesWikipediaEN.htm, retrieved Sept. 8th, 2020
[17] https://www.wikidata.org/wiki/Wikidata:WikiProject_Cultural_heritage, retrieved Sept. 8th, 2020
[18] https://www.wikidata.org/wiki/Wikidata:Statistics/en, retrieved June 26th, 2018
[19] https://tools.wmflabs.org/mix-n-match/#/, retrieved Sept. 8th, 2020

Data, connecting many otherwise unrelated data sources. Their growing importance also becomes visible in their increased application for the calculation of different metrics for various scenarios, ranging from measures of Information Quality, such as the validation of contradictory sources of attribution of artworks described by Daquino in [Daquino, 2020], to general social indicators such as Gender Inequality, as described by Konieczny and Klein in [Konieczny and Klein, 2018].

Besides acting as hub between data collections and providing structured facts from Wikipedia info-boxes, DBpedia also makes available extracted hyperlinks between individual Wikipedia articles in different language editions, thus providing machine readable representations of the article network structures of different Wikipedia language versions. Considering specific subsets of articles yields hyperlink networks between descriptions of similar typed entities, which potentially provide valuable data for exploration. Considering for example the subset of Wikipedia biographies, their mutual hyperlinks on the one hand represent a highly complex and multi-layered network of family, professional and other types of social relationships embedded in the individual life narratives and on the other hand, especially as far as highly influential historical persons are concerned, networks of relationships based on the interpretation of their influence on other historical or contemporary persons.

Network narratives in Cultural Heritage data

Both professionally curated and user generated data collections usually provide multiple pathways across the featured entities. As far as cultural heritage artifact collections are concerned, such pathways can be chains of inter-object connections based on common features between them, such as same author/creator, genre, topic, etc., but, as discussed above, also be based on more complex relationships, such as trajectories through historical social networks of the creators of these objects and other related persons. The hyperlinks between Wikipedia articles as provided by DBpedia for example represent structured extracts of parts of the "stories" present in the individual articles, in this case the explicit connections between different entities as they occur throughout the narrated course of events the entities are involved in. In case of biographical Wikipedia articles, the extracted links represent inter-person relationships based on direct interaction or influence, in case of articles about artists and artistic styles/movements, they represent the respective affiliations as declared by the artists themselves or as ascribed to them by art historians. The same is the case for explicit social relationships present in the Getty ULAN, resulting from documented relationships between relevant persons which have been recorded by art historians and related researchers.

Manovich in [Manovich, 2002] referred to the sum of all such pathways between entities in a data collection — database in his terms — as a hyper-narrative. Taking the ULAN person records as an example, the sum of all person relationships and all possible paths connecting any two persons therein are the "ULAN hyper-narrative", all connections between Wikipedia articles in a specific language version are e.g. the "English Wikipedia hyper-narrative", only connections between artist biographies in English Wikipedia would

be the "English Wikipedia artist hyper-narrative", etc. Cultural heritage presentation environments using such data sources as a basis for browsing or searching thus operate within these hyper-narratives by allowing users to choose their individual paths through them. Visualizing such hyper-narratives as a whole could serve as a map for users to locate their current position within the data-collection and to possibly identify interesting regions for further discovery.

Canons

Considering that many CH data collections such as the Getty ULAN stem from a specific scientific domain, in this case art history, it can be assumed that their respective hyper-narratives to some extent represent large-scale "bottom-up" representations of specific overall aspects of the underlying scientific knowledge, composed of numerous individual "bits" such as the person relationships in the ULAN. Given the many debates that emerged for example in art history in the second half of the 20th century, such as Post-Colonial or Feminist studies questioning grand narratives and revealing traditions of bias in art history research, it is also of interest to study the overall structure of such hyper-narratives and to search for any large-scale patterns in them. Such a "distant reading" approach, as coined by Moretti in [Moretti, 2013], could help to make visible any specific trends or biases present in large-scale structures of cultural heritage data collections which otherwise might remain hidden from their users.

Such an analysis could also contribute to the question to what extent existing domain narratives "survive" the transition to the digital realm and shed light on if and how existing canons, understood here as "a small number of normative and formative texts, places, persons, artifacts, and myths which are meant to be actively circulated and communicated in ever-new presentations and performances [... and have ...] passed rigorous processes of selection [...]", defined by Assmann in [Assmann, 2008] and representing the "cultural working memory" of a society, are also embedded in digital cultural heritage data collections.

1.2 Problem statement

As outlined in section 1.1, cultural heritage related contextual data sources, such as the Getty ULAN, increasingly become available online and offer various opportunities for exploration, both with respect to their application, as well as to their analysis. Being embedded in the VSEM[20] research project, this work focuses on the art history domain and the main aspects of the research can be outlined as follows:

1. **Combining digitized CH content & metadata with contextual data about historical networks of persons related to the arts**
   The emerging web of data potentially connects collections of digitized CH content,

[20] https://web.archive.org/web/20191126200412/http://vsem.ec.tuwien.ac.at/, retrieved Sept. 8th, 2020

i.e. digitized surrogates of artworks, etc. including metadata such as title, creator, date, genre, etc. with contextual historical social network data, such as the Getty ULAN. It is of high interest to explore the potential of the integration of these diverse datasets, on the one hand regarding the information gain associated with the data integration itself and on the other hand for the usability of such data mashups in the context of virtual presentation environments for cultural heritage data, assuming that social network data can be used as "floor plan" or overview map to present artworks embedded in their historical context, thus providing an "exploratory search" approach to navigating art collections.

2. **Analysis of historical network data for "big pictures" of cultural processes**
   It can be assumed that large numbers of structured information "snippets" representing historical facts, such as individual person relationships, provide bottom-up views on large-scale historical processes. It is thus of interest to create such views and describe the resulting structures. Historical material, such as diagrams about developments in art history created by scholars, could serve as reference material in this regard.

3. **Comparison of different data sources for differences and commonalities**
   Besides the growing availability of institutional sources, increasingly comprehensive offerings based on user generated content, such as Wikipedia, often provide information about similar entities, such as the many multilingual Wikipedia biographies about artists also featured in the ULAN. This provides high incentives to compare the mutual coverage of these diverse resources, assessing the extent of information they have in common and where they diverge. Comparing the links between entities in the different resources offers itself as useful quantiative approach in this regard.

1.2.1 Research questions

Based on the aspects outlined above, the following research questions form the basis for the work presented in this thesis.

- **Research Question 1 (RQ1)** What potential exists in integrating object-level metadata with related social network data for exploratory search based virtual presentations of art history content?

- **Research Question 2 (RQ2)** What kind of "large-scale" historical narratives are embedded in networks of interconnected persons related to art history?

- **Research Question 3 (RQ3)** How do the structures of the person-to-person networks found in institutional and/or crowd-sourced art history data collections differ and what do they have in common?

1.3 Methodological approach

This section covers the methodological aspects of this thesis. The first subsection discusses basic assumptions related to network data in the context of the problem statement. The second subsection embeds the research work and its results within the Design Science Framework.

1.3.1 Initial assumptions

As outlined in Section 1.2, the main research focus is on network data derived from different sources related to art history. The overall assumption behind this approach is that datasets featuring sufficiently large numbers of relationships between art historically relevant entities, i.e. teacher/student, influential, professional or family ties between artists, patrons and other important historical persons, but also more complex forms such as ties between such persons and art styles, contain contiguous networks that interconnect the involved entities across multiple centuries, resulting in structures that, translated into visual form, represent data-driven maps of art history. Networks between art historically relevant persons are for example expected to feature numerous parallel but mutually interconnected chains of successively interconnected persons, merging and branching across time like complex genealogies whose visualization would ideally represent a seamless "relational" narrative from the onset of the history of art throughout the present days.

Throughout the history of the scientific domain of art history, however, there have been shifting conceptions of what constitutes relevant developments across time. As outlined by Hatt and Klonk in [Hatt and Klonk, 2006], there were shifting perceptions whether individual artistic genius, overall socio-historical context, independent evolutionary-like development or even some metaphysical entity such as Hegel's Weltgeist determined stylistic advance. This was reflected in a variety of approaches ranging from biographical historiography such as Vasari's "Lives", providing a — not always historically accurate — genealogy of Renaissance artists and their mutual relations, over Connoisseurship, attributing stylistic features to individual artists, to Formalism, seeking to systematically and formally describe and track stylistic developments based on visual features only, disregarding individual names. More recent developments starting around the second half of the 20^{th} century re-introduced the social aspects of art history via views on power and exclusion, such as Marxist, Feminist or Post-Colonial approaches, which also questioned the existence of an all-encompassing grand narrative of art history.

Since it is out of scope of this work to participate in these discussions, is is acknowledged that person-centered views as present for example in social network settings provide only one specific lens on developments related to the arts, but at the same time assumed that these views nevertheless represent an important aspect amongst the many individual factors contributing to such developments: On the one hand, social relationships provide the societal foundation for art to happen at all, since artistic production depends on resources which have to be provided by someone. On the other hand, specific person relationships, such as teaching, also imply at least some stylistic connection between

the œuvres of the involved artists and biographical narratives of artists often mention stylistic influence from individual artworks created by others. Moreover, it can also be argued that the presence of a recorded relationship between two persons to some extent correlates with their historical importance, i.e. acts as evidence for other relevant processes. Such a notion of importance, however, is to be taken with care, since, as for example discussed by Eisler in [Eisler, 1987], collections of artist biographies celebrating the great achievements of their protagonists were often created in order to act as halls of glory especially for the emerging nation states during the 19th century. It is therefore of importance to interpret quantifiable data extracts of such information accordingly.

1.3.2 Design Science Research framework

The work presented in this thesis follows the "Design Science in Information Systems Research" (DSR) approach as outlined by Hevner et al. in [Hevner et al., 2004]. This research paradigm is centered around the creation and application of new and innovative design artifacts for achieving knowledge and understanding of a problem domain. The concept of a design artifact refers to different aspects of the overall design process, which include constructs ("vocabulary and symbols"), models ("abstractions and representations"), methods ("algorithms and practices") and instantiations ("implemented and prototype systems"). The design process is characterized by its open ended nature, centered on the iterative refinement of artifacts based on continuous assessment.

Figure 1.1 shows the DSR framework adapted to the context of this work. The problem domain of this thesis is the growing digital cultural heritage data universe and especially large-scale data about networks of art historically relevant entities, whose availability is a very recent phenomenon offering many opportunities for original research. The research undertaken in this thesis can therefore be considered exploratory, using design artifacts as "probes" for revealing interesting and relevant aspects in the encountered data suitable for further research. As outlined in the research questions stated in Section 1.2.1, one aspect of the work is about studying data mashups of metadata about digitized artworks with contextual data sources, in this case the Getty ULAN, for possible information gains and the applicability in the context of the creation of virtual museums (RQ1), while the other aspect is about the analysis and comparison of the content of contextual data sources themselves, seeking to identify possible large-scale structures embedded therein (RQ2 + RQ3). This results in a set of different artifacts which mainly fall into the "methods" and the "instantiations" categories outlined above.

As far as RQ1 is concerned, the design artifact **A1** (Figure 3.6), is an experimental software prototype for the presentation of digitized artworks and their metadata retrieved from the online resource "Web Gallery of Art" (WGA) combined with relevant contextual data, in this case social network data of related persons from the Getty ULAN and additonal information sourced from DBpedia. The prototype makes use of Semantic Web technologies for retrieving and integrating the data sources and of network visualization paradigms and algorithms for their presentation. It served as testbed for assessing the possibilities arising from the combination of these complementary types of data based on the continuous addition of experimental features, with the aim of providing an exploratory

search environment for fine art. The most recent version of the prototype, described in Chapter 3, is the result of a continuous assessment/refinement process where the iterative addition of features revealed interesting insight on data level. This particularly refers to the extraction of alternative social network information based on creator-subject relations in portrait paintings found in the WGA, which can be considered as separate artifact **A2** (Figure 3.8) supporting the identification of information gaps in the ULAN, motivating further research outlined in RQ2 and RQ3. Gathering the different sources and combining them in 3D space using an adaptation of the well known "dot" graph layout algorithm can be considered as separate method **M1**.

In contrast to RQ1, the research work performed in the context of RQ2 and RQ3 is of descriptive/analytical nature and the created artifacts can be ascribed to the "methods" and the "instantiations" categories. Network visualizations of the Getty ULAN social network data represent design artifact **A3** (Figure 4.24), revealing insight into the overall art history narrative embedded there and motivating the comparison of the ULAN data with related data from Wikipedia. The infrastructure for comparing these two data sources represents a methodological framework **M2**, which includes the construction of a combined dataset of ULAN person records with their respective metadata and associative links and corresponding Wikipedia articles and their mutual hyperlinks in various language versions, mapped via ULAN mappings in Wikidata. The latter can be considered as an approach to select a domain-specific set of Wikipedia content via an external data source. Additional sources such as artifact **A2** and information from scholarly literature were used as reference data for comparing the ULAN and the Wikipedia networks. Design artifact **A4** (Figure 5.19), a visualization of the network of biographical Wikipedia articles identified via the ULAN, can be considered as the outcome of a method **M3** that involves chronological filtering of the network of inter-article hyperlinks based on temporal features. The differences encountered between the Getty ULAN and the Wikipedia visualizations motivated an additional comparison between the ULAN mapped networks in individual Wikipedia language versions, one of the outputs being artifact **A5** (Figure 5.25) representing the juxtaposition of visualizations of networks of ULAN-mapped Wikipedia biographies in several language versions. Another aspect of the Wikipedia person-to-person networks was explored via artifact **A6** (Figure 5.27), representing a network of nationalities derived from aggregating the Wikipedia network by the persons' nationalities, revealing clusters of nationalities based on geographical, cultural as well as colonial ties. The filtering approach to create artifact **A6** is based on the application of standardized Pearson residuals in two-way tables and can be considered a separate method **M4**.

The outcomes of the visualizations of Wikipedia person-to-person networks motivated further exploration: Artifact **A7** (Figure 6.8) represents the visualization of a bi-partite network of hyperlinks between Wikipedia articles about persons and articles about art movements as identified in Wikidata, inspired by a famous historical scholarly visualization. The creation of **A7** involved the construction of the bi-partite network, representing a modification of method **M2**, **M5**, which identifies persons relevant to art history on Wikipedia without using external data sources such as the ULAN. The subsequent

chronological filtering of the extracted bi-partite network represents an extension of method **M3**, **M6**, which requires the prior definition of beginning and end dates for artistic movements which were identified via life-dates of linked persons. Artifact **A8** (Figure 6.10) represents an evolution of Artifact **A7** based on the application of a statistical filtering method called FDSM to the extracted bi-partite network, resulting in a filtered and projected network of interlinked art styles. The application of the FDSM approach in the context of the bi-partite artist-style network represents the method **M7**.

The continued exploration of approaches to identify Wikipedia biographies relevant to art history without using external sources such as the ULAN led to method **M8** which uses co-occurrences of occupations in Wikidata person records to identify domain-specific clusters of related occupations, which can be used to select subsets of domain-specific person biographies. The clustering of co-occurring occupations and its visualization yielded artifact **A9** (Figure 6.12) and the visualization of the set of identified artist biographies resulted in artifact **A10** (Figure 6.14). The generalizability of method **M8** to other domains was demonstrated by the creation of additional visualizations of interlinked biographies relevant to other fields, resulting in artifact collection **A11** (Figures 6.16, 6.17 and 6.18).

1.4 Main contributions

The exploratory nature of the research undertaken in this thesis has led to a number of results than can be considered as contributions to the digital cultural heritage domain, but also beyond. The following list provides an overview on the different aspects involved:

- **Combination of complementary data sources for the virtual art exhibits**
  The software prototype **A1** (Figure 3.6) and the accompanying method **M1** highlight the benefit of the integration of complementary data sources. It provides multiple layers of information which represent a "whole which is greater than the sum of its parts", yielding the derived network of portrait creators and subjects **A2** (Figure 3.8) which contribute to the identification of missing information in the ULAN and its comparison with Wikipedia content.

- **Find new, data driven ways to visualize developments in art history**
  The network visualizations of the ULAN person-to-person network data **A3** (Figure 4.24) and of the corresponding Wikipedia network **A4** (Figure 5.19), as well as the "ULAN independent" visualizations of a) art styles and movements interconnected via person articles (**A7**, Figure 6.8), b) directly interconnected styles and movements based on the projection and filtering of the network behind a) (**A8**, Figure 6.10) and c) interconnected Wikipedia person records identified via artistic occupations (**A10**, Figure 6.14) explicitly reveal and display narratives corresponding to the succession of major art historical epochs. Together with the applied methods **M3**, **M6** and **M7** they can thus be considered data-driven updates of historical approaches to visualize large-scale developments in art history and also as contributions to the visualization of large-scale biographical networks in general.

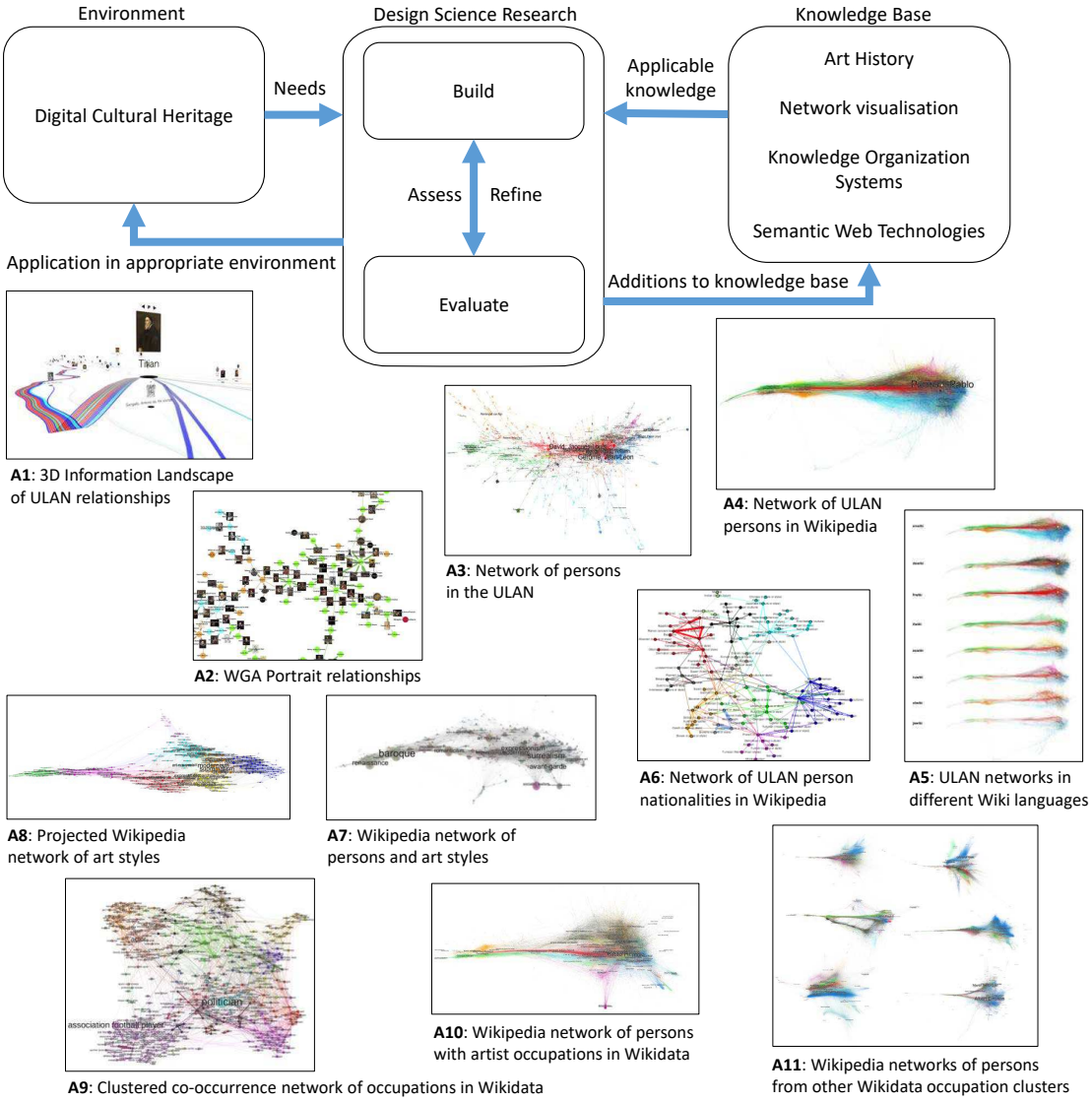


Figure 1.1: Design Science Research (DSR) framework and design artifacts **A1-A11**

- **Compare and analyze different datasets about networks of persons relevant to art history**
  This contribution mainly consists of the framework **M2** for comparing the networks present in the ULAN and in Wikipedia. It is centered on a hybrid dataset combining ULAN records and their mutual associative links with multilingual Wikipedia hyperlink information between Wikipedia biographies mapped to the ULAN via Wikidata. This provides a rigorous infrastructure for the systematic analysis of overlaps and gaps between these two data sources and also between individual Wikipedia language versions. Additional outcomes are the visual juxtaposition of

hyperlink network structures of ULAN biographies in different Wikipedia language versions in artifact **A5** (Figure 5.25) and the analysis of a high-level aggregation of the hyperlinks between Wikipedia biographies from the perspective of person nationalities, yielding artifacts **M4** and **A6** (Figure 5.27).

- **Find new ways to identify domain specific subsets in large-scale data collections**
  While artifacts **A4** and **A5** are based on Wikipedia biographies identified via the ULAN via framework **M2**, artifacts **A7**, **A8**, **A10** are purely derived from crowd-sourced content available via Wikidata and Wikipedia. The two respective methods **M5** and **M8** can thus be considered as contributions to the problem of identifying relevant content in large-scale cross-domain data collections. As far as **M8** is concerned, the co-occurrence analysis of Wikidata occupations yielding artifact **A9** (Figure 6.12) is an additional contribution in its own right, while design artifact **A11** (Figures 6.16, 6.17 and 6.18) serves as proof-of-concept for the approach to identify domain specific person groups via clustered occupations by visualizing exemplary person networks from other domains identified that way.

Figure 1.2 places the created artifacts and methods within the DSR knowledge contribution framework. The visualizations **A3**, **A4**, **A5**, **A7**, **A8** and **A10** as well as methods **M3** and **M6** can be considered new solutions to known problems and are thus represent improvements to existing approaches. The integration prototype **A1** and its accompanying method **M1** and artifact **A2** are based on the application of existing solutions to new problems, the same is the case for methods **M2**, **M4**, **M5**, **M7** and **M8** and artifacts **A6** and **A9**, they can thus be considered to fall into the "exaptation" category.

1.5 Structure of this work

The remainder of this thesis is structured as follows. Chapter 2 provides an overview on historical and contemporary applications of the network metaphor in the context of cultural heritage, art history in particular, and also discusses a few historical "non-network" examples which are of relevance as well. Chapter 3 proceeds with the work done in the context of RQ1 which resulted in design artifact **A1** (Figure 3.6), the accompanying method **M1** and the artifact **A2** (Figure 3.8). Motivated by the results obtained in the context especially of the latter, Chapter 4 follows RQ2 and provides an analysis of the content found in the ULAN, focusing on the analysis and visualization of its network of associative person relationships, yielding design artifact **A3** (Figure 4.24). Following RQ2 and RQ3, Chapter 5 introduces Wikipedia as alternative data source and provides overview on related research. It continues with the creation of a hybrid dataset combining ULAN person metadata and associative links with the link structure derived from mapped Wikipedia biographies identified via Wikidata, which forms the heart of the framework **M2** for comparing the two network structures. The analysis highlights several properties of the Wikipedia network of biographies in terms of ULAN

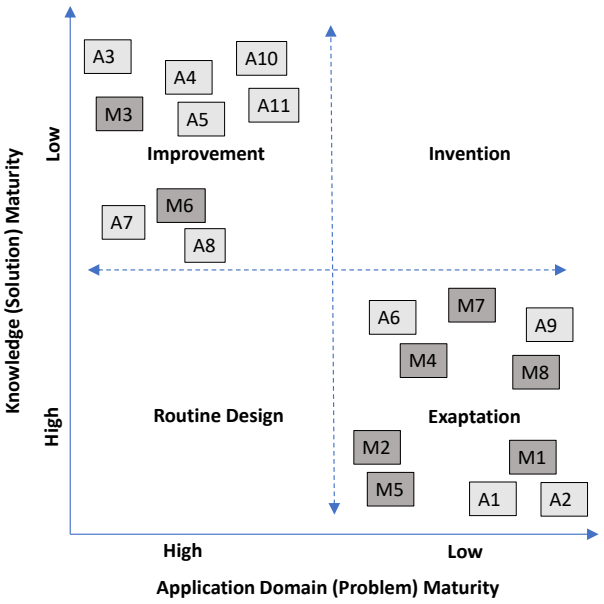


Figure 1.2: DSR knowledge contribution framework

coverage and overall structure, the latter leading to artifacts **A4** (Figure 5.19), **A5** (Figure 5.25) and **A6** (Figure 5.27) and related methods **M3** and **M4**. Continuing RQ2 and RQ3, Chapter 6 introduces the two additional approaches to analyze and visualize biography-based art history content on Wikipedia beyond the ULAN, using Wikidata as main metadata source instead. A discussion of the analysis and visualization of the bi-partite network of persons and styles and the related artifacts **A7** (Figure 6.8) and **A8** (Figure 6.10) and the underlying methods **M5** and **M7** is followed by a description of the final experiment based on occupation co-occurrence in Wikidata and the respective artifacts and methods **A9** (Figure 6.12), **A10** (Figure 6.14), **A11** (Figures 6.16, 6.17 and 6.18) and **M8**. Chapter 7 reflects on the contributions of this thesis and provides an outlook for future work.

This work is based on the following peer-reviewed papers:

- [Arends et al., 2009] Arends, M., Goldfarb, D., Merkl, D., and Weingartner, M. (2009). Interaction with art museums on the web. In Proceedings of the IADIS Int'l conference WWW/Internet 2009, Rome, Italy, pages 117–125. IADIS Press.

- [Arends and Goldfarb, 2010] Arends, M., Froschauer, J., Goldfarb, D., Merkl, D., and Weingartner, M. (2010). Interaction with Museum Content in Web3d. In Konferenzband EVA Berlin 2010, pages 161–167, Heidelberg. arthistoricum.net.

- [Arends et al., 2010] Arends, M. and Goldfarb, D. (2010). Social interaction with cultural heritage on the web. In International Conference on Web Engineering. Vienna, Austria, volume LNCS 6189, pages 587–592. Springer.
- [Goldfarb et al., 2011b] Goldfarb, D., Arends, M., Froschauer, J., Merkl, D., and Weingartner, M. (2011). Combining cultural heritage related web resources in 3d information landscapes. In Proceedings of the 2011 International conference on Electronic Visualisation and the Arts, pages 73–80. BCS Learning & Development Ltd.
- [Goldfarb et al., 2011a] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2011). Revisiting 3d information landscapes for the display of art historical web content. In Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, page 46. ACM.
- [Goldfarb et al., 2012] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2012). Art history on wikipedia, a macroscopic observation. Extended Abstract presented at the ACM Web Science Conference (WebSci 12), Evanston, Illinois.
- [Goldfarb et al., 2013] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2013). Comparing Art Historical Networks. Leonardo, 46(3):279–279.
- [Goldfarb et al., 2014] Goldfarb, D., Arends, M., Froschauer, J., Weingartner, M., and Merkl, D. (2014). Collectivizing the Barr Model. Leonardo, 47(3):270–270.
- [Goldfarb et al., 2015] Goldfarb, D., Merkl, D., and Schich, M. (2015). Quantifying cultural histories via person networks in Wikipedia. Poster presented at the International School and Conference on Network Science (NetSci 2015), Zaragoza.
- [Goldfarb and Merkl, 2016] Goldfarb, D. and Merkl, D. (2016). Bilder der Kunstgeschichte in der Getty Ulan und in Wikipedia. In Konferenzband EVA Berlin 2016, pages 33–42, Heidelberg. arthistoricum.net.
- [Goldfarb and Merkl, 2018] Goldfarb, D. and Merkl, D. (2018). Visualizing Art Historical Developments Using the Getty ULAN, Wikipedia and Wikidata. In Proceedings of the 22nd International Conference Information Visualisation (IV), Fisciano, Italy, pages 459–466. IEEE.
- [Goldfarb and Merkl, 2019] Goldfarb, D. and Merkl, D. (2019). Data-Driven Maps of Art History. International Journal of Art, Culture and Design Technologies (IJACDT), 8(1):1–15.

CHAPTER 2

Networks and Art History

In the context of big data and large-scale data analysis, the concept of networks and their visualization/analysis is often immediately associated with online social networks and other "born digital" data collections. The basic notation and the mathematical foundations for studying social and other networks in a formal way, however, were already laid in the first third of the 20th century. Even earlier than that, generations of genealogists were involved in constructing networks of kinship in order to track lineages mainly of powerful and influential persons. Such genealogies also played a role in the context of art history, laying an early foundation for a long tradition of "network thinking" also in this discipline. This chapter is dedicated to providing an overview in this regard, starting from historical "analog" examples and ending with current approaches. Additional examples from related domains or representing broader approaches are discussed as well.

2.1 Historical examples

Giorgio Vasari's famous 16th century "Vite", the story about the "Lives of the Most Excellent Painters, Sculptors, and Architects" ([Vasari, 1991]), can be considered as one of the first artist genealogies, not represented visually but embedded in a collection of cross-referencing biographies of famous Italian Renaissance artists. According to Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005], his work has been explained to be rooted in the ambition of a number of wealthy families such as the Medici to document their social advancement in form of a genealogical tree and must rather be considered as socio-historical literature than as historiographic source. It nevertheless had strong impact on following generations, which increasingly also sought to represent their derivatives of Vasari's foundation in visual form. As described by Schmidt-Burkhardt, explicit visual representations of networks of interconnected artists, artist groups or artistic styles started to appear in larger numbers during the 19th and 20th centuries, serving quite different purposes. On the one hand, they were increasingly used alongside other

visual diagrams as educational and communicational means to explain perceptions of developments in art history, usually serving as visual compression of scholarly reasoning in this regard. On the other hand, they were increasingly also used by artists themselves in order to embed, and thus legitimate, their work in a broader art historical context. A notable shift away from genealogies of genius individuals towards more formal conceptions of developments in the arts, such as progressions of style, happened in the early 20th century, with strictly formalist positions even proclaiming an "art history without names", although the social context of art and especially the connection between the object and its creator nevertheless remained to be relevant.

2.1.1 Genealogies of artists and styles

Figure 2.1 shows Arcisse de Caumont's "Tableau figuratif des variations de l'architecture religieuse" from 1831, identified by Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005] as an early example for a genealogical representation of stylistic development. It appeared in form of a genealogical tree of sacral French architecture styles whose trunk represented main stages from Greek antiquity to the time of its creator while the branches showed influences from other cultures and regions as they occurred throughout the centuries. One item on Caumont's agenda was to lay ground to a French national identity by following its continuous cultural traces into the past and many other approaches had similar motivations, turning such representations into political devices themselves.

Genealogical trees were employed not only from the historian's point of view but also directly by artists, using such trees as monuments for themselves. The German romantic artist group of the Brotherhood of St. Luke ("Lukasbund") for example, which later dissolved into the Nazarene movement, sought to reestablish a true German art based on the admiration of Albrecht Dürer as its founding father and the rejection of the then prevailing classicism. According to Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005], their "Stammbaum der neudeutschen Kunst", shown in Figure 2.2, was an attempt to visualize their conceived heritage which also represented an important step towards the autonomization of the fine arts since the artists started to make their own art history. The trunk of the represented tree carried a miniature reproduction of Albrecht Dürer's "The Resurrection" from which spreading branches held little plates with the names of 63 artists, linking them to their self-declared common ancestor.

The construction of the pedigree relating the brotherhood's members to Albrecht Dürer is a genealogical operation Zerubavel in [Zerubavel, 2011] called "Pasting", i.e. the connection of noncontiguous symbolic lineages into one seemingly coherent whole. References to antiquity or to the Renaissance, ubiquitous in art history, are another example for this operation. The identification of common ancestors, referred to by Zerubavel as "Lumping", is another important aspect of such genealogical constructs: The vision of a common ancestry, such as seeking the German identity in Dürer's art, helps to tie together members of nations, but also the notion of a European culture is strongly associated especially with the quasi-convention of Greek antiquity as unifying element.

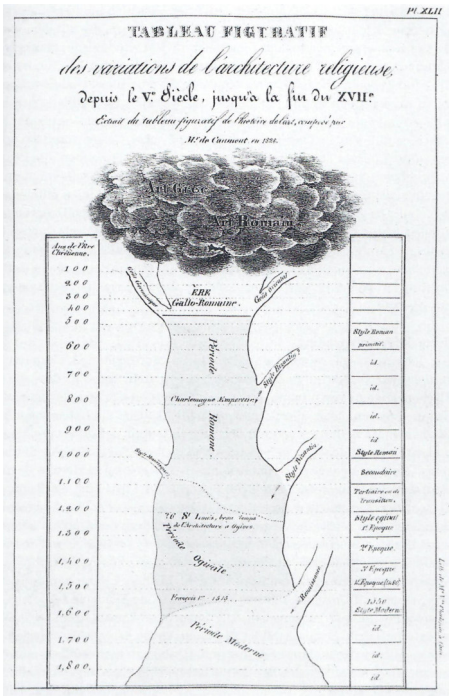


Figure 2.1: Tableau figuratif des variations de l'architecture religieuse, A. de Caumont, 1831

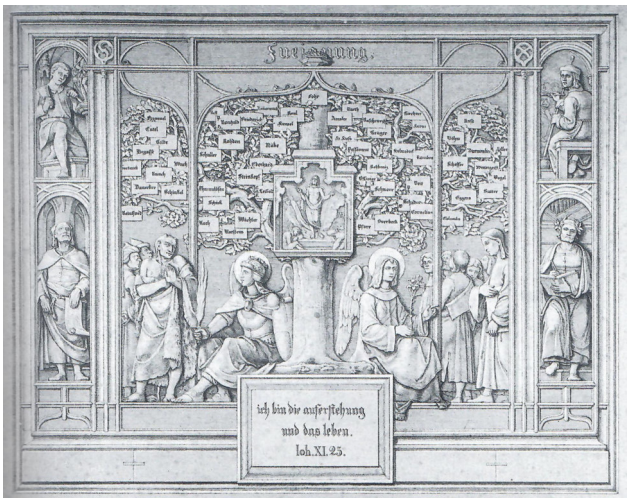


Figure 2.2: Stammbaum der neudeutschen Kunst, Ferdinand Olivier, 1823

Systems and developments

Influenced by other scientific disciplines and their methods, which increasingly included scientific diagrams as means of reasoning and communication, a number of 19th century art history scholars sought to create sophisticated diagrams in order to reason about and explain developments within their own field of study, some of which also using means of network visualization. A remarkable example, shown in Figure 2.3, was published by the Swiss art and architecture historian Heinrich von Geymüller in [Geymüller, 1898] and discussed by Heck in [Heck, 2014]. The chart sought to give a bird's eye view on the development of the Renaissance style in French architecture from its introduction through influx from Italy at around 1500 until Geymüllers age. It was based on a stylistic analysis over that period focusing on how the influence of "rigorous and objective" antique elements (red) imported from Italy interacted with the "free and subjective" Gothic style (green) that had prevailed in France until around 1500. Geymüller conceived the latter to subsequently have lived forth (blue) but transformed based on other "subjective" Italian and protestant influences, interplaying with the antique style in a periodical manner. This aggregate process display was fed by two machine like pipe systems at the top and at the bottom of the chart, where the pipes connected various chambers which represented

a mixture of about 50 important persons and 50 architectural works. The chart thus presented Geymüller's general conception a historical process of stylistic development in visually aggregated form as a result of underlying individual processes, whose detailed visualizations acted as supporting evidence. Focusing on the development of French architecture, Geymüller's diagram seems like a sophisticated extension of Caumont's tree shown in Figure 2.1, in which the Renaissance influx into France is shown as entering from the bottom-right branch.

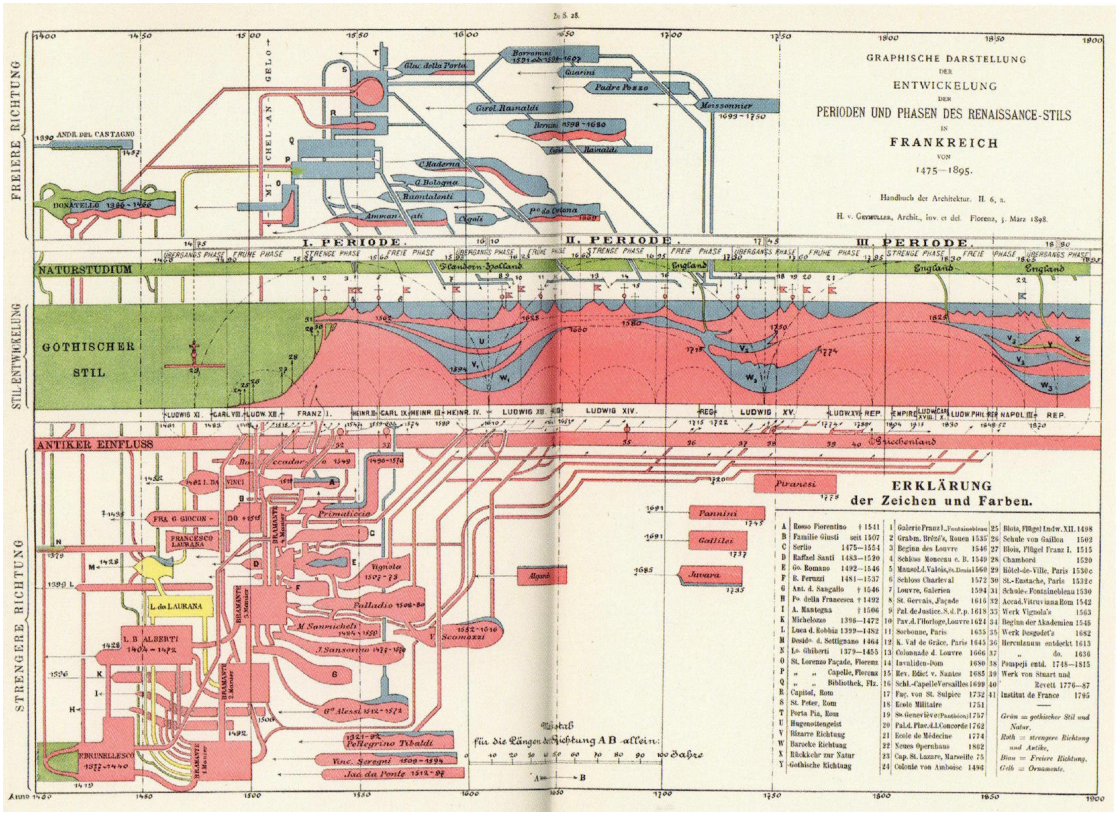


Figure 2.3: Graphische Darstellung der Entwickelung der Perioden und Phasen des Renaissance-Stils in Frankreich 1475-1895, Heinrich von Geymüller, 1898

Chronological networks of style

One of the most famous visualizations of artistic developments, a genealogy of modern art styles drawn on a regulating chronological grid, was created by Alfred H. Barr, Jr., American art historian and inaugural director of the Museum of Modern Art in New York City. His diagram for the 1936 Cubism & Abstract Art exhibition at the MoMA, shown in Figure 2.4 and published in the exhibition's catalog in [Barr, 1936], represented the interrelationships between 14 modern art movements from 1890 until 1935 and became an

important element in modern art history. On the one hand it reoresented a blueprint for the show curated by Barr and, as part of the exhibition catalog and as poster shown in each of the exhibition rooms, served as navigational guide for visitors. On the other hand, as outlined by Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005], it also represented Barr's specific conception of a history of modern art, an art history not any longer represented by individual artists but by artistic movements, the so-called "-isms" which were presented as irreversibly driving towards abstraction, bifurcating between two main strands of geometrical and non-geometrical abstract art along the way, with Cubism as watershed. Seven artists were nevertheless still individually mentioned by name in the diagram, most of them from before 1900, including Van Gogh, Gauguin, Cézanne and Seurat, which were considered by Barr as the founding fathers of modern art. One particularly new aspect was the explicit inclusion of "non-Western" influences such as "Negro Sculpture", "Near-Eastern Art" or "Japanese Prints", the latter for example were an important input for Van Gogh, who collected them extensively[1]. Barr's evolutionary model of the progress of modern art became the standard way to conceive modern art developments and how to present them at the museum.

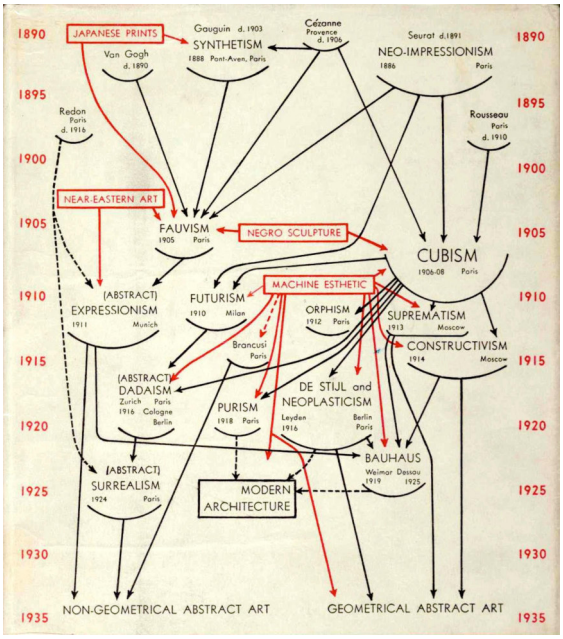


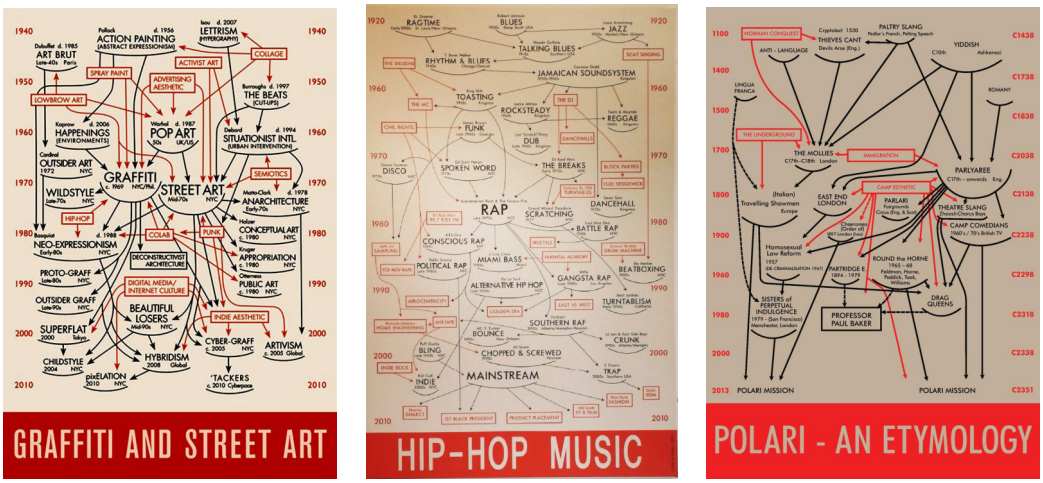
Figure 2.4: Diagram of Stylistic Evolution from 1890 until 1935, Alfred H. Barr, Jr., 1936

Although Barr was a Formalist believing in the autonomy and political neutrality of art, his model of modern art history was not always conceived as unpolitical or neutral. According to Schmidt-Burkhardt, it was challenged for leaving out modern art movements which did not fit the authors idea of progress towards abstraction and its underlying Formalist approach moreover criticized by Marxist-oriented art historians

[1] https://www.vangoghmuseum.nl/en/japanese-prints, retrieved Sept. 8th, 2020

to be detached from social processes, which was later joined by Feminist positions questioning the very limited presence of female artists in the artist movements featured by Barr (about 5%). Unfortunately, the latter fact holds true for most art historical approaches until the second half of the 20th century. Other art historians questioned Barr's focus on French art as root for the presented developments, as well as that the focus on abstraction later suggested that the successor to the achievements of the school of Paris was to be identified in American post WWII Abstract Expressionism, at the expense of other modern art achievements around the world [Schmidt-Burkhardt, 2005]. This was particularly interesting in the context of increasing rumors, such as discussed by Sooke in [Sooke, er 4], that Abstract Expressionism was actively promoted with the support of the Central Intelligence Agency as antagonist to Socialist Realism, with the aim to spread the idea that artists had absolute individual freedom in the West.

Regardless of the critique of Barr's history of modern art, his approach to visually convey developments in the arts became very influential and iconic, referenced from both within and outside art history. As far as its visual impact is concerned, Rosenberg and Grafton in [Rosenberg and Grafton, 2010] explicitly mentioned the combination of genealogical and chronological elements to be the foundation for its power, which led to many variations also in other domains, such as the examples shown in Figure 2.5. Whether as attempt to to describe developments in Graffiti and Street Art [Feral, 2011] (a), to interrelate music genres related to Hip-Hop [Mccauley, 2010] (b) or to show the development of a specific language of gay subculture [Dolan, 2012] (c), its many replications demonstrate its far-reaching influence even into contemporary popular culture.



(a) Grafitti and Street Art (b) Hip-Hop Music (c) Polari - An Etymology

Figure 2.5: Variations of Barr's diagram

Another diagram combining genealogy with a chronological grid, mentioned by Rosenberg/Grafton in [Rosenberg and Grafton, 2010], was Eric Newton's 1941 "Chart of Art History" shown in Figure 2.6, published in [Newton, 1941], an art historical survey. Newton sought to give a broad overview on the history of art from its common roots in antiquity until his own time, focusing on nine national schools and about 68 notable representatives thereof. The diagram appeared as an attempt to contine the "history of old art" into the realm of modern art, but already ended at its very beginning, featuring a small group of famous French Impressionists. Although created a few years after Barr's diagram, it appeared like an immediate predecessor to it, featuring Cézanne and Picasso as linking points in this regard, but still sticking to the concept of national schools rather than any "-isms".

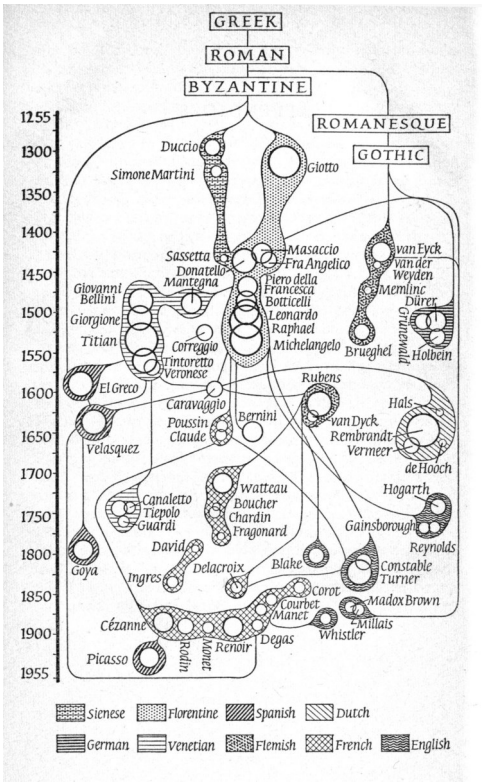


Figure 2.6: Chart of Art History, Newton, 1941

Newton's diagram featured a number of additional layers of information compared to Barr's. Size played a much more prominent role for representing importance and moreover, individual persons were clustered into higher level structures representing the schools. The importance of each school was represented by the size of the respective shaded area, which was pierced through by circles of varying radius representing the importance of individual artists, respectively. While both schools and artists were connected through explicit, so-called "threads of influence", the clustering of the individual persons into

schools and their chronological "flow" provided an additional, more implicit layer of connectedness. In a note about the diagram, Newton stated that his chart was based on his personal opinion and that any other person would create a different version of it. Especially with respect to the ranking of schools and artists by importance he admitted the lack of a general underlying principle, for example explaining the higher importance credited to "old artists" with the more favorable conditions offered to them at that time.

A more recent example for a chronological and, although rather implicit, genealogical view on developments in the arts was the "Tate Artist Timeline" from 2006, a commissioned work created by the graphic designer Sara Fanelli for the Tate Modern gallery in London, shown in Figure 2.7. Integrated into the museum architecture as monumental mural spanning 40 meters across two museum floors, about 65 modern art movements were chronologically arranged from 1900 to the 21st century. Each of the movements was surrounded by a subset of in total about 417 related artists which floated around like satellites and formed an implicit network between each other. It is worth noting that this visualization was described by art historian Tiampo in [Tiampo, 2010] as "less deterministic, [...], more open, and more international", update of Barr's diagram providing a clear British perspective on recent art history, although it itself again only featured a fraction of artists of all those active throughout the 20th century, based on yet another specific selection of the ones deemed to be important enough to be shown.

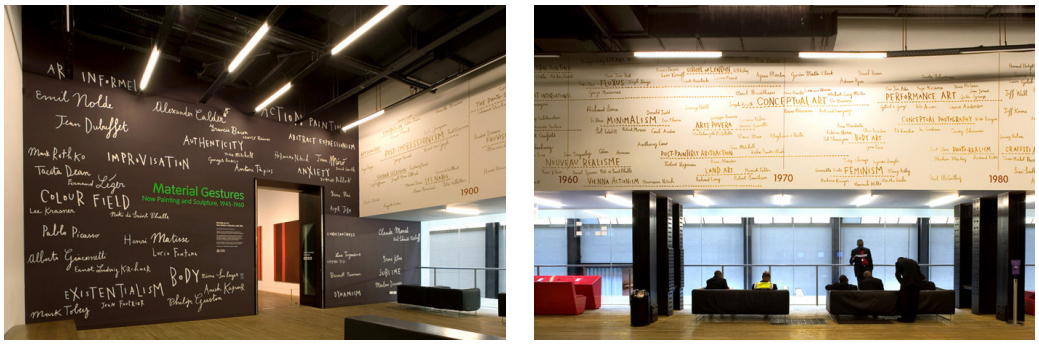


Figure 2.7: Tate artist timeline, Sara Fanelli, 2006

Borrowing from the visual style of subway maps, another recent approach to chronologically represent the interplay between both artistic styles and artists was created by Gompertz as book illustration in [Gompertz, 2012]. Shown in Figure 2.8, the visualization featured art movements and artists connected via differently colored lines. Art movements appeared to be placed chronologically from left to right and were represented as interchanges, while individual artists were placed in between them. Although the placement of artists in between different movements suggested their affiliation with both, they were also assigned to single styles via color coding, which introduced some ambiguity in this regard.

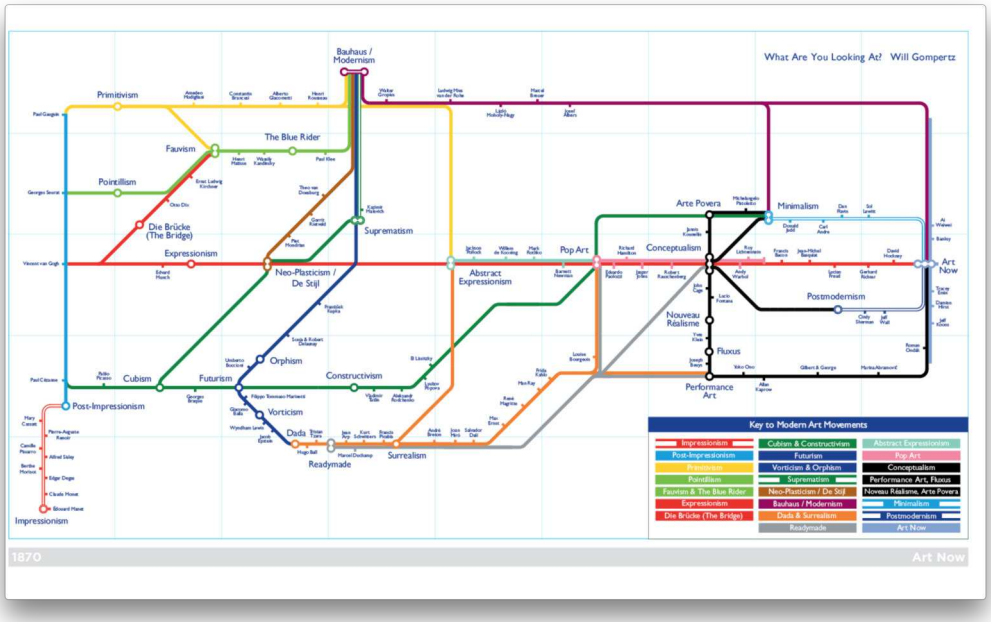


Figure 2.8: What are you looking at?, Gompertz, 2012

2.1.2 Examples from other domains

Attempts to display historical developments in visual form were of course not limited to the domain of art history. Some broader approaches covered a variety of disciplines, which also included the arts. One "classic" example, the 1765 "Chart of Biography" by British polymath Joseph Priestley, shown in Figure 2.9, was presented by Rosenberg and Grafton in [Rosenberg and Grafton, 2010]. Amongst its chronological display of lifespans of notable persons from various domains, it included a dedicated row for artists (painters, sculptors, printers, engravers and architects, but also actors and musicians), highlighted by a red rectangle in the Figure. Chronologically placed along the horizontal axis, artist lifespans on the left side concentrated at the time of antiquity and then again at the right side from around 1200AD, leaving in between a remarkable gap of more than 1,000 years, suggesting that no notable artist had lived during this period, which was not the case in any other featured domains.

Priestley provided an accompanying description of this chart in [Priestley, 1765] where he acknowledged this gap in comparison to the much more dense "Statemen, Heroes, and Politicians" section and also mentioned that the chart might have looked very different if it was based on other criteria. Like Eric Newton almost 200 years later for his "Chart of art history" in [Newton, 1941], he admitted that this would have been the case if someone else would have created it and particularly stated that his choice to include a person was on the one hand based on fame rather than merit, on the other hand also

based on the intended audience of his chart, which he believed to be mainly English, leading to a higher proportion of persons of that nationality. The middle-age gap in the artist section, however, was not due to Priestley's individual selection and also appeared elsewhere. Schich et al. in [Schich et al., 2014] for example explained it with a general lack of art historical coverage for the period between 300 and 600 AD and also mentioned that artist signatures only started to appear on artworks created after 1200AD.

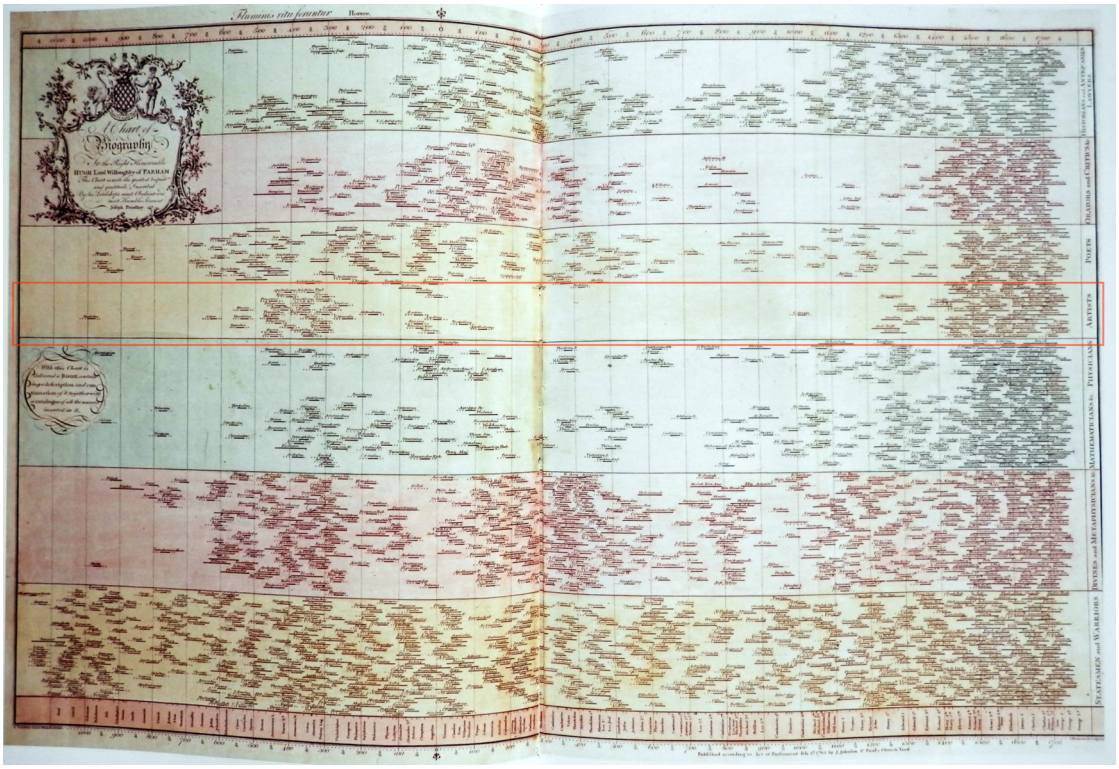


Figure 2.9: New Chart of Biography, Joseph Priestley, 1765

Another remarkable historical example presented in [Rosenberg and Grafton, 2010] is Friedrich Strass's chart "Strom der Zeiten" (Stream of time), the first version of which created in 1804, a later edition shown in Figure 2.10. It provided an aggregated view on history based on the growth, decline and interactions of various peoples of the world starting at around 3000BC, visualized in form of parallel rivers, bifurcating and joining across the flow of time from top to bottom, ending at around 1800AD. As visible in the Figure and also mentioned in [Rosenberg and Grafton, 2010], the time axis in this chart did not represent time in a linear fashion as it was the case in Priestley's visualization, but increased in resolution towards more recent times.

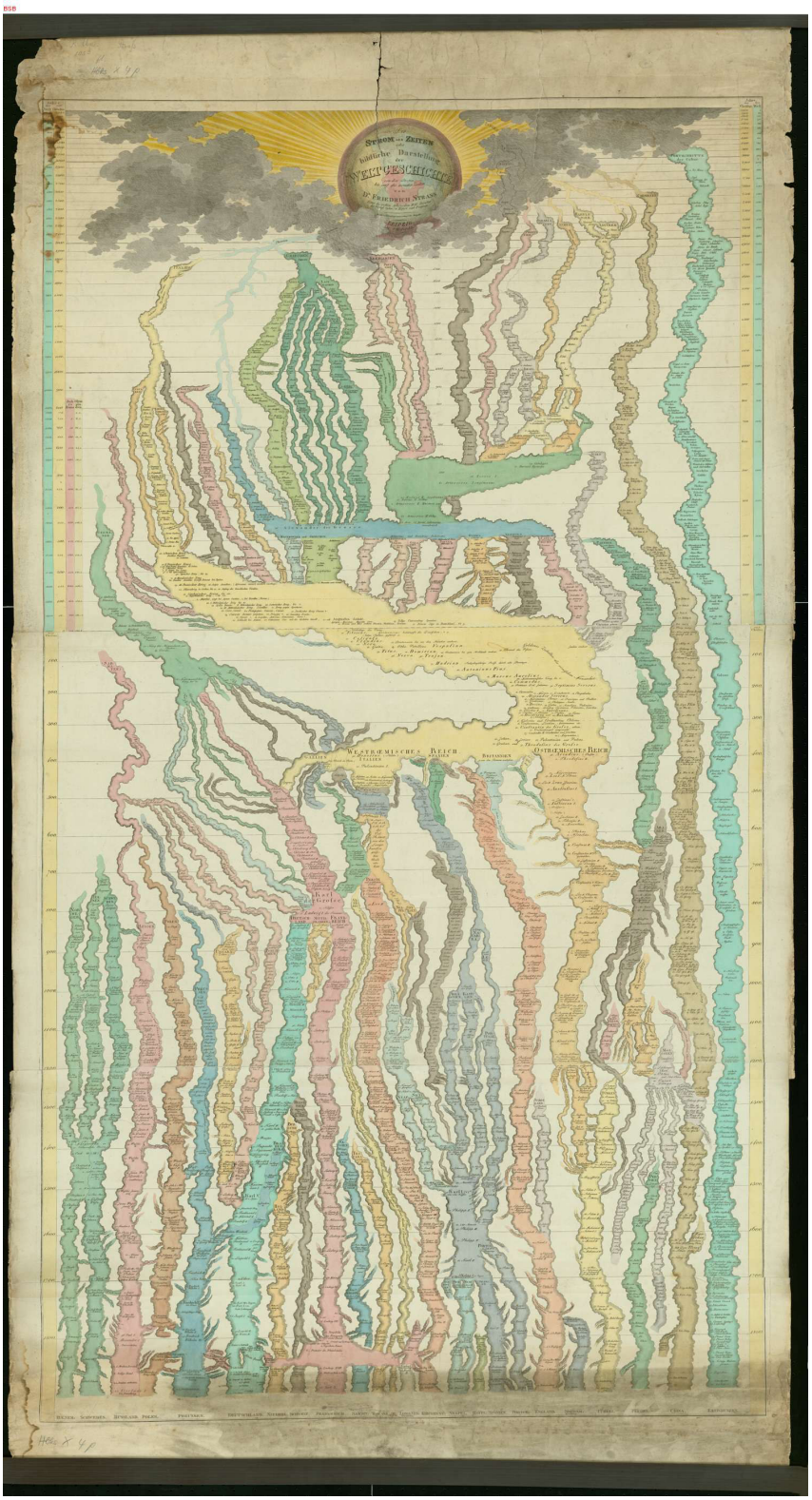


Figure 2.10: Strom der Zeiten, Friedrich Strass, 1828, urn:nbn:de:bvb:12-bsb00003029-3

2.2 Art History networks in the realm of the digital

The discussion of historical approaches to represent networked narratives in art history revealed a long tradition of related visual artifacts which were used in different settings, targeting both scholarly peers as well as a more general audience. In recent years, the increasing availability of low-cost computing equipment and powerful graphics hardware resulted in a shift of related efforts towards embracing new possibilities within the realm of the digital. Emerging tools for data analysis and visualization brought new means for original cultural heritage research and the communication of results, and increasingly interactive modes of presentation enabled new approaches to cultural education.

The revolution of the Internet and the World Wide Web added another dimension in this regard, allowing to provide more dynamic access to fast growing collections of content, as well as means to combine different sources for more contextualized interpretation. The mere idea of such a global network of information, and especially more recent developments such as the emergence of online social networks or the vision of an all encompassing, machine readable Web of Data, put the network metaphor even more into the center of attention. While historical approaches mainly used network representations for communicating knowledge gained via other, more traditional means, contemporary digital approaches are used for a variety of purposes in communication, education, online collection access as well as the search for new findings. Beginning with a number of basic definitions, this Section provides important examples for these different applications of networks.

2.2.1 Basic definitions

Considering digital representations of networks, they are usually represented as *graphs*, which are defined by Newman in [Newman, 2003a] to consist of a set of *nodes*, also called *vertices*, which are pairwise connected via *links*, also referred to as *edges*. Throughout this work, they are referred to as nodes and links. Links between pairs of nodes can be *directed* or *undirected* and are called *self-links* or *loops* if they connect a node with itself. In some cases, pairs of nodes can be connected by more than one link, which are then referred to as *multi-links*.

Both nodes and links can be assigned with *attributes*. Node attributes encode specific features of the described entity, which usually at least includes a *label*, often also a *type* and additional type-specific attributes such as birth date and place for a person. Link attributes encode information about the relationship between pairs of nodes, often including the *weight*, i.e. a usually numeric attribute describing the strength of a tie, and other specific attributes such as if a tie between persons is based on a teaching or on a family relationship.

Networks featuring only one type of node, such as, e.g., friendship networks between persons, are called *uni-partite*, such as shown in Figure 2.11 a). Networks featuring two or more different node types and the additional constraint that only nodes of different type can be connected with each other are called *multi-partite*. If there are only two

types of nodes, the latter are also known as *bi-partite* or *affiliation networks*, such as for example connecting authors with publications, sketched in Figure 2.11 b).

Links between nodes can be based on direct observation, such as teacher/student relationships, or be derived from more "low-level" information. One example for the latter are aggregated networks, such as shown in Figure 2.11 c), where the original nodes are grouped according to similar features, resulting in a reduced, weighted version of the original version. Another important example for derived networks, in this case of bi-partite networks, are so-called *bi-partite projections*, shown in Figure 2.11 d) and e), where the bi-partite network from Figure 2.11 b) is split into two uni-partite networks, each of which consisting only of nodes of one type which are pairwise connected via weighted links based on the number of "other-type" nodes they are mutually connected with in the original network.

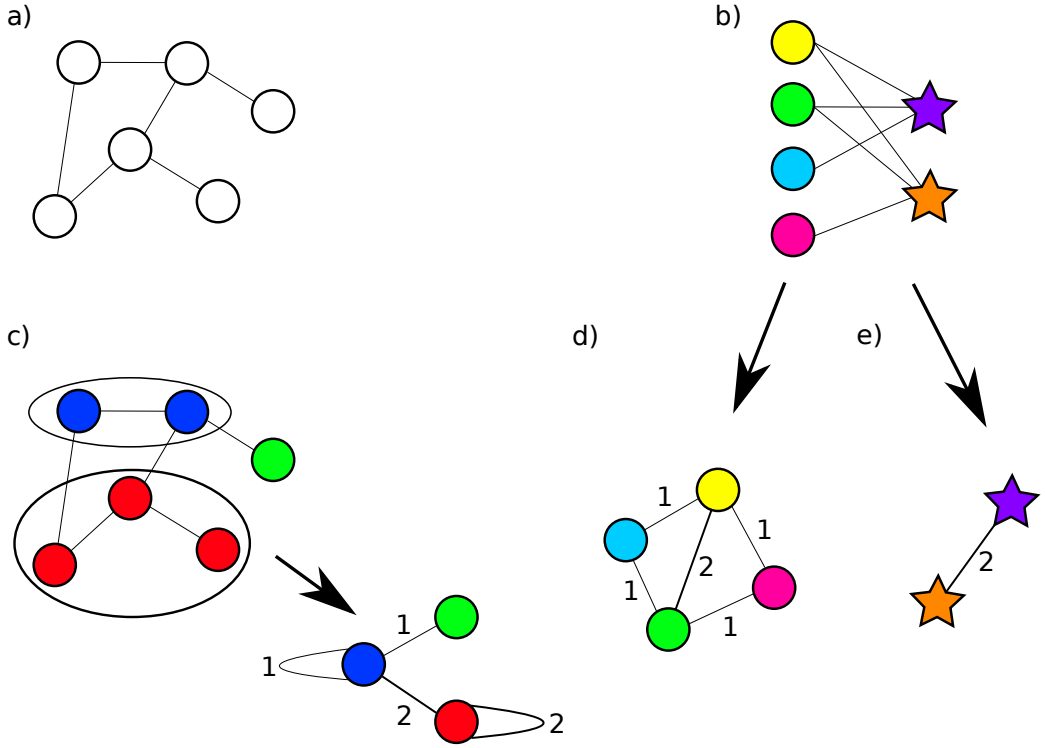


Figure 2.11: Uni- (a) and bi-partite (b) , aggregated (c) and projected networks (d, e)

Network/Graph Visualization

The visualizations shown in Figure 2.11 are based on the mapping of the abstract network data into a visual form. As discussed in [Fleischer and Hirsch, 2001], the fundamental building blocks for visualizing networks are the choice of the visual representation

of the network itself, i.e. choosing visual representations for its nodes and links, and the choice of an appropriate *layout* of these representations in 2D or (virtual) 3D space.

Nodes are usually displayed using 2D or 3D shapes or symbols, while the links between them are drawn as lines or tubes. The visual representations can moreover be used to encode node or link attributes via visual variables such as shape, size or color. Nodes can also be used as canvas for image data such as digital reproductions of artworks.

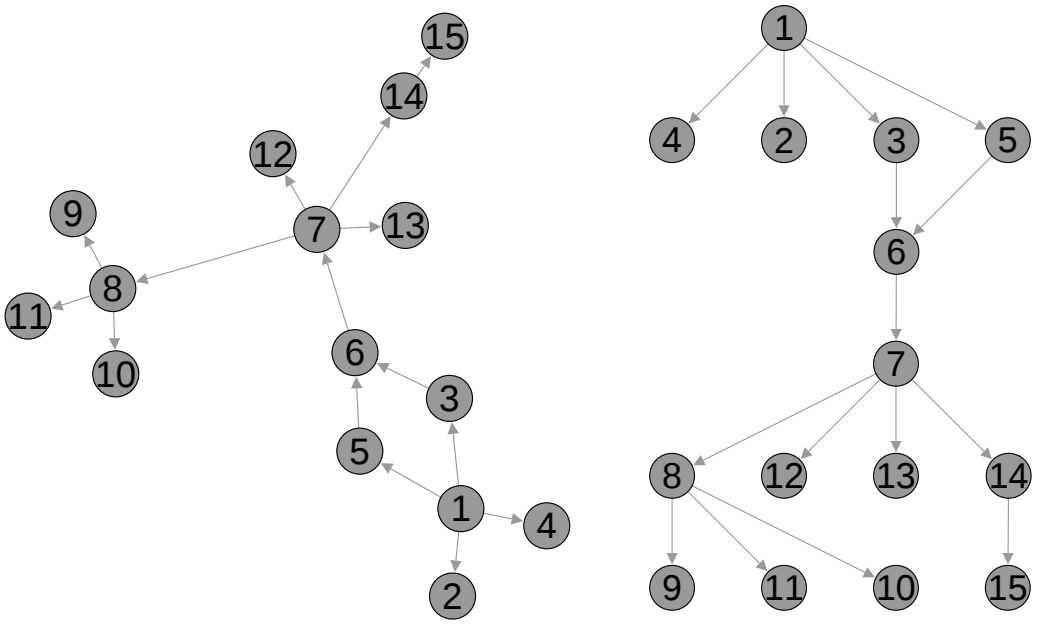
While attribute mappings to shape, color and size usually immediately reflect the respective values, either directly or via a transfer function, the locations of the elements are usually determined via a specific layout strategy. Although such a strategy can as well be based on direct mappings of node attribute values to positional coordinates, the aim is often rather to emphasize specific characteristics of the underlying network structure, such as the visual identification of densely interconnected subgroups of nodes. This is usually achieved by using a dedicated *layout algorithm*.

According to [Gibson et al., 2013], approaches to computing graph layouts can be divided across two main strands. So-called *force-based layout* approaches tend to focus on the topological aspects of the network where only its structure, i.e. nodes and links, the latter sometimes weighted, serve as input for the layout and usually no additional attributes are taken into account. The classical approach is to simulate a physical system of mutually repulsing nodes held together by spring-like forces representing the strength of the inter-node links. Such simulations seek to iteratively reach an equilibrium in which repulsion and attraction cancel each other out to a maximum possible degree. This usually results in topologically clustered layouts where nodes which are part of densely interconnected groups tend to be positioned closely together while the groups themselves are visually more separated from each other, as visible in the example shown in Figure 2.12 (a).

Other layout methods in turn do take specific node attributes into account for determining node positions, although not necessarily via direct mappings. Gibson et al in [Gibson et al., 2013] referred the latter to be only one of three different approaches in this regard, for example used when assigning geographic coordinates to nodes for displaying them on a world map. The second approach, so-called *constraint-based layouts*, employs node attributes to impose specific restrictions on node placement, for example by fixing one coordinate dimension according to some node feature, while determining the other dimension(s) algorithmically. The third approach is to consider potential group memberships, such as affiliations with political parties, to determine node placement.

A well known example for constraint-based methods are layered graph drawing approaches where nodes are assigned to specific layers, which can either be based on the constraint that directed links should always point into one direction, or on a specific node attribute such as a persons's birth date. Within each layer, nodes are arranged in a way that that the number of crossing links beween the nodes in different layers is minimized, which, according to Purchase in [Purchase, 2004], is amongst the most important aesthetic criteria for the effectiveness of a network visualization. Figure 2.12 (b) shows a layered version of the same graph as shown in (a). This visualization is based on the constraint that the directed links of the graph should always point into one

direction, thus highlighting the flow of information present in the network structure. As seen in the comparison between the two methods in the Figure, this comes at the expense of the readability of the node clustering which is also inherent to the observed network.



(a) Force-based (b) Layered

Figure 2.12: Force-based and layered layout of the same graph

2.2.2 Presentation of arts and related developments/collection access

Concerning the application of networks for the digital presentation of art and related developments or as means for providing access to collections, a distinction must be made regarding the level at which the network concept is applied. On the one hand, network visualizations can be used at the interface level as contemporary continuation of the historical approaches discussed in Section 2.1, with added possibilities such as interaction and dynamic access to an underlying database. On the other hand, and especially in the context of collection access, networks can also be used at "lower" levels which are less visible to the user. One example for this are recommendation systems which use networks to suggest items related to the currently viewed one via various pathways.

Interface level

An example for a virtual presentation of art based on a social network approach was realized for the 2012 exhibition "Inventing Abstraction 1910-1925" at the Museum of

Modern Art in New York City. According to the show's curator Leah Dickerman, it was based on the conception that the drive towards abstraction was based on social relationships between artists, "pushing each other to ever more radical positions"[2]. Coined by Artnews[3] as "Facebook for Abstractionists", a visualization of a network of 83 persons[4], shown on the left hand side of Figure 2.13, emerged from the curators' effort to investigate and record the historical relationships between them. As a clear reminiscence to Barr's famous 1936 Cubism and Abstract Art exhibition, a large static print of the visualization, whose visual design closely resembled Barr's iconic diagram, was put at the entrance of the exhibition space and featured in the exhibition catalog in [Dickerman and Affron, 2012].

An interactive version of the diagram was presented via a dedicated website[5]. Web visitors were presented with a layout similar to the static version, but provided with means to explore each person's ego network (i.e. the immediate network neighborhood of one person) by switching to a related detail view upon click. In the detail view, clicking on one of the featured artist nodes revealed life dates, featured works and a brief description, shown on the right in Figure 2.13. Explicitly created for the exhibition, the visualization and its interactive counterpart were based on highly curated data solely focused on early 20th century modern art and its related actors.

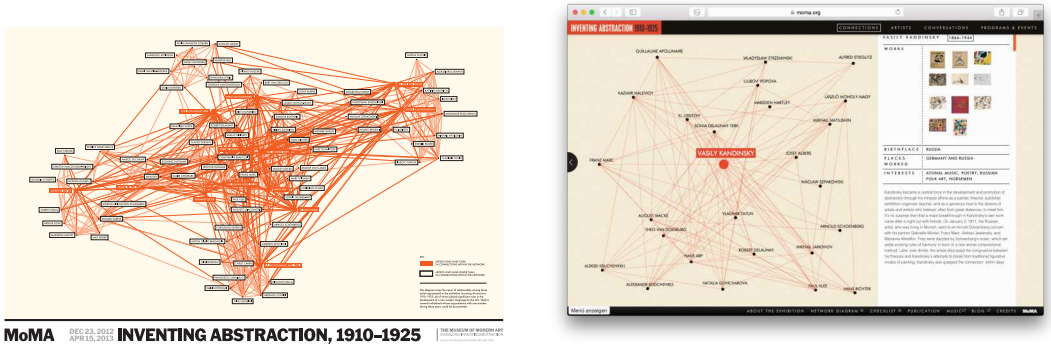


Figure 2.13: MoMA Inventing Abstraction network views, [Dickerman and Affron, 2012]

Focusing more on broad semantic interrelations, Hinrichs, Schmidt and Carpendale used network visualization as means to communicate contextual information within a museum installation about a single artist, described in [Hinrichs et al., 2008]. The contextual network, shown on the right side of Figure 2.14, consisted of different interconnected types of entities such as art movements, general topics, publications and persons. In this

[2] https://www.moma.org/explore/multimedia/videos/250/1200, retr. Sept. 8th, 2020
[3] http://www.artnews.com/2012/10/02/momaabstractionfaceboo/, retr. Sept. 8th, 2020
[4] https://www.moma.org/interactives/exhibitions/2012/inventingabstraction/?page=artists, retr. Sept. 8th, 2020
[5] https://www.moma.org/interactives/exhibitions/2012/inventingabstraction/?page=connections, retr. Sept. 8th, 2020

regard, it resembled the notion of a semantic map providing multiple pathways across collections, as expressed by Cameron in [Cameron and Robinson, 2007].

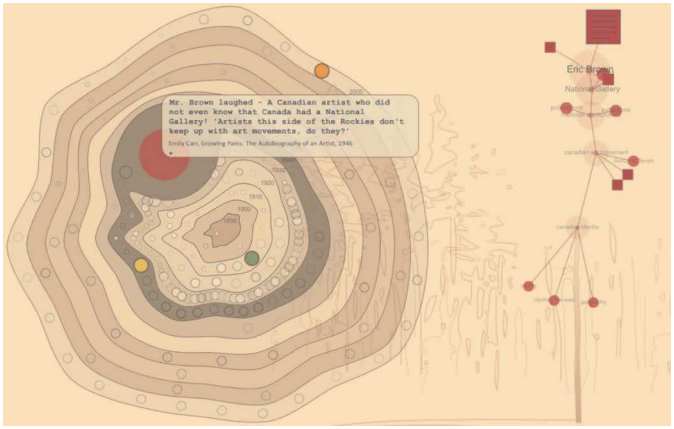


Figure 2.14: EMDialog, [Hinrichs et al., 2008]

An example for a projection of a multi-partite network was described by Dörk, Pietsch and Credico in [Dörk et al., 2017], featuring a network visualization of persons and organizations based on projection of person-artifact links, shown in Figure 2.15. Sizes of nodes were based on the number of artifacts associated with the respective person or organization while the thickness of the connection between them was based on the number of artifacts referenced in common. The network was not displayed as a whole but presented via time slices, each of them showing the sub-network for a given period of time.

Figure 2.15: Persons and organizations interrelated via artifacts, [Dörk et al., 2017]

3D Network Visualization

While the approaches mentioned above employed 2D network visualization methods for presentation, related 3D approaches have also been applied in the context of cultural heritage and art history. Transcending beyond the 3D replication of existing architecture, Ruffaldi et al. in [Ruffaldi et al., 2008] presented a 3D information visualization approach to explore graphic collections, shown in Figure 2.16. The system allowed curators to place texts and images in 3D space using different layouts such as "cube", "corridor", "hierarchical map" and "solar system". An important aspect was the possibility to visually interlink individual items in the presentation with explicit relationships. The authoring of the environment and the creation and placement of the links, however, had to be performed in a manual fashion using a dedicated editor and were thus not "data-driven" in this regard.

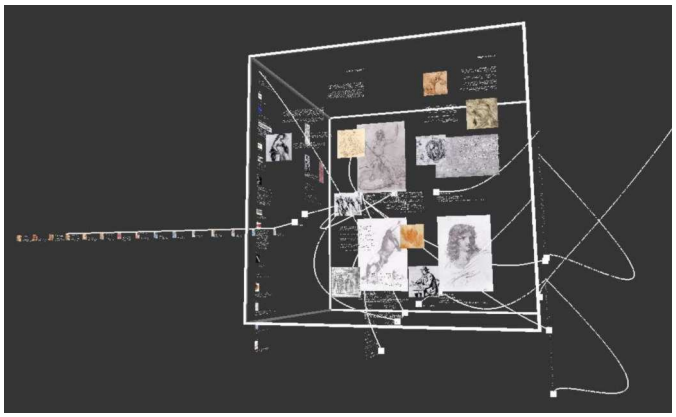


Figure 2.16: 3D Information Landscapes, [Ruffaldi et al., 2008]

As far as genealogical networks are concerned, a 3D map of artist relationships was presented as a static rendering in the context of the East Art Map project, shown on the left of Figure 2.17. The project was dedicated to map a view on developments of post 1945 fine arts in Eastern European countries, which were rather neglected regarding their role in the canon of modern art history due to the political division of the world at that time. The interactive Website[6] shown on the right of Figure 2.17, however, did not feature any network visualization feature, although it allowed users to follow the artists network via "related artists" links present in some of the featured artist records.

[6] http://www.eastartmap.org/, retrieved Sept. 8th, 2020

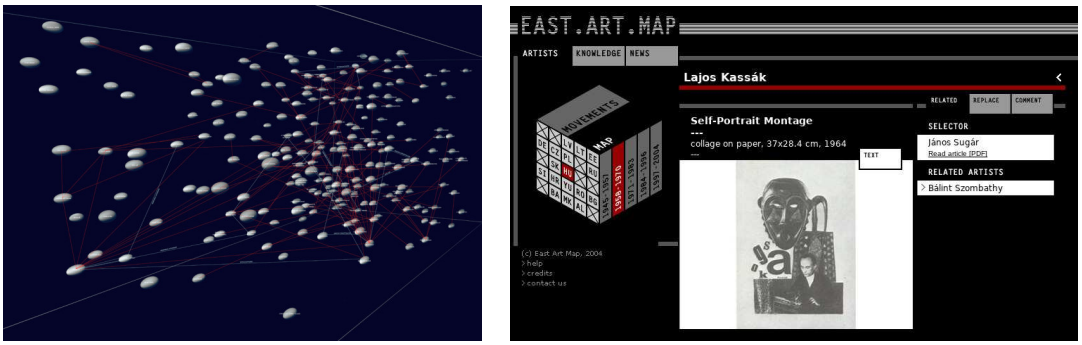


Figure 2.17: East Art Map, `http://www.eastartmap.org/`

Time

The existing approaches presented so far have in common that their network views did not encode time visually, i.e. no visible chronological flow was present in the ordering of the displayed nodes. Only the system presented by Dörk, Pietsch and Credico in [Dörk et al., 2017] took chronological aspects into account by allowing to sweep across network slices taken at different points in time. A number of additional approaches, however, explicitly included chronological layouts in their presentations and are discussed below.

Another refresh of a static visual representation of art related developments across time was the interactive update of the Tate artist timeline already discussed in Section 2.1. Realized in 2015 as a local installation on a large interactive video-wall[7][8] and shown in Figure 2.18, the display featured a similar timeline view with art movements surrounded by related artists, with the added functionality to study individual artworks and related texts upon touch. Besides the implicit bi-partite network linking artists with styles, no further relational aspects between any of the displayed entitites could be explored there.

This aspect was for example approached by Dörk et al. in [Dörk et al., 2012], where one visualization mode allowed users to explore ego networks of famous painters across a timeline view. While this mode, shown in Figure 2.19, effectively highlighted the temporal aspects intrinsic to such data, it was less suited for the simultaneous display of multiple ego networks and thus rather limited to displaying the neighborhood of one person after the other.

[7] `https://www.youtube.com/watch?v=2WT1P8Botek`, retrieved Sept. 8th, 2020
[8] `https://www.tate.org.uk/visit/tate-modern/bloomberg-connects-timeline-modern-art`, retrieved Sept. 8th, 2020

Figure 2.18: Interactive Tate modern art timeline, [Framestore, 2015]

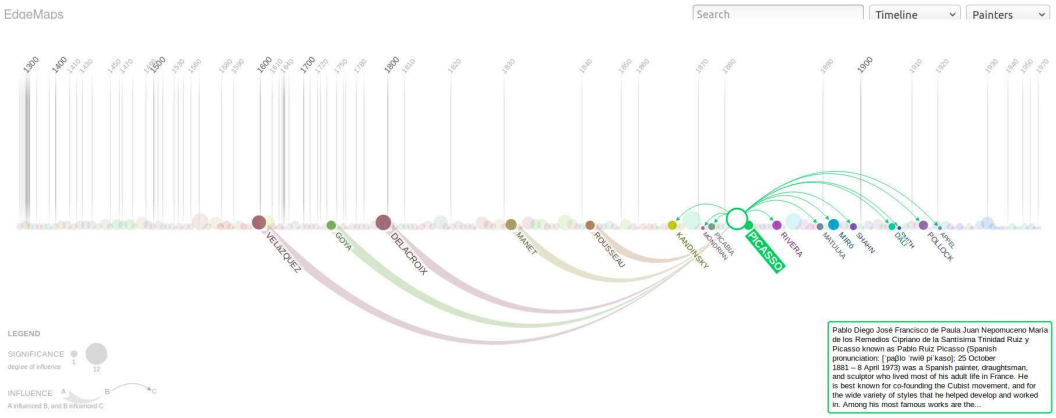


Figure 2.19: Edgemaps,[Dörk et al., 2012]

Another approach to rendering networks across chronological constraints was presented by Albrecht et al. in [Albrecht et al., 2016]. The visualization showed a network of movies[9] and influential relationships between them by mapping nodes to coordinates, where the Y-axis was used to represent time and the X-axis used to encode different structural network measures such as connectivity or clustering. In contrast to the previously mentioned approaches, this mode of visualization allowed the simultaneous display of a relatively large network consisting of tens of thousands of nodes, enabling the visual identification of large scale trends potentially embedded in such data collections. Using attributes and derived measures to locate nodes on the 2D plane, however, did put less focus on the actual topology of the network.

[9] `http://www.culturegraphy.com/`, retrieved Sept. 8th, 2020 also features a version for a network of persons

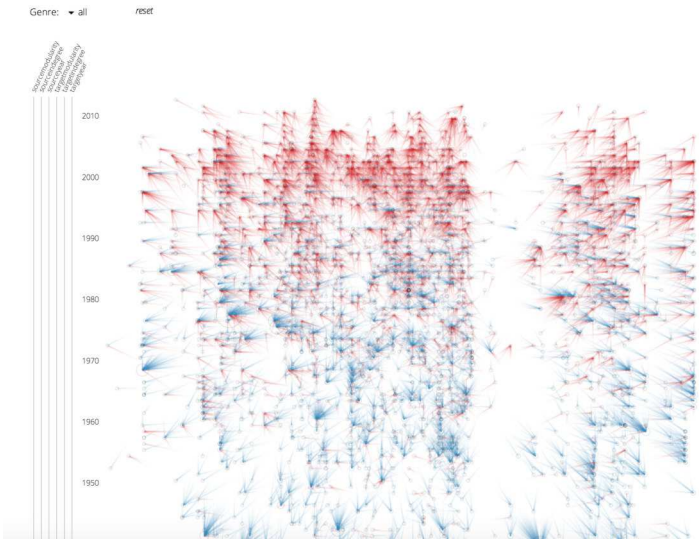


Figure 2.20: Culturegraphy,[Albrecht et al., 2016]

Semantic search and recommendation

The approaches described in the first part of this Subsection used means of visualization to explicitly display direct or derived relations between different entities such as artists, artworks or art movements. These visualized network topologies were mainly used as means to navigate collections of artworks at different levels. As stated in the introduction to this Subsection, networks have also been used "behind the scenes", as means to support search and for recommendation. Concerning art history related systems, the ULAN played an important role as data source in this regard.

Kurki and Hyvönen presented an approach in [Kurki and Hyvönen, 2009] which used an RDF representation of the ULAN and its relationships to drive the "relational search" facility in the Finish cultural heritage portal CultureSampo[10], allowing users to query for potential connections between featured artists and other persons present in the database. The authors showed how person relationships could be used to create recommendation rules suggesting e.g. students of the currently viewed artist. One interesting aspect of the service was a simple network visualization feature allowing users to explore the ego-networks around one selected person. The visualization only featured person names, no other information or additionally linked data such as images were provided and the authors mentioned its usefulness for disambiguating historical persons having similar name, thus suggesting it for rather analytical than exploratory purposes.

Artwork recommendations based on ULAN person relationships were also reported by Schreiber et al in [Schreiber et al., 2008] in the context of a semantic search demonstrator where users were provided with suggestions for artworks created by artists which were

[10] `http://www.kulttuurisampo.fi/`, retrieved Sept. 8th, 2020

professionally related to currently viewed artist. This system also featured the possibility to create visualizations of the underlying search graph, which included not only the social relationships emanating from the currently viewed person but also all the other data paths represented in the database. Provided as a feature of the database server infrastructure, these visualizations therefore served as administrative tool for data maintenance and were not targeted at end users.

Wang, Stash, Aroyo et al. in [Wang et al., 2009] showed the overall usefulness of social relationship data as contextual information for artworks by analyzing different types of semantic relations for the semantic recommendation of art related concepts. The authors found that while artists recommended via teacher/student relationships were rated to be of average accuracy by users, they were considered to be amongst the most interesting recommendations. Searching for useful navigation patterns in art recommenders, the authors identified patterns of the form "artwork -> creator -> teacher/student -> style" to be amongst the most promising ones, concluding that vocabularies such as the ULAN or the AAT were particularly useful for recommendations in the context of the arts, thus also highlighting the importance of data about relationships between historical persons in this regard.

2.2.3 Art history education and research

The examples provided in Subsection 2.2.2 were mainly targeted at a broad audience, using networks as means of exploratory navigation or item recommendation across collections of artworks. Many additional applications of the network metaphor in art history, however, are used in the context of higher education and research. This Subsection discusses a number of important examples in this regard.

An example for the application of networks in art history education was described by Ross in [Ross, 2013]. Students attending a university-level art history course were assigned to extract data about personal networks of 20^{th} century female artists from literature and to explore them by means of network visualization. The approach was based on the question if female artists formed similar social webs of mutual artistic influence as it was so common amongs their male peers. The resulting visualization is shown in Figure 2.21, reflecting the conclusions of the exercise that only few truly independent female networks could be identified. This was in part attributed to an existing gender bias especially in secondary literature about female artists, where, according to the authors, more focus was put on documenting ties to famous male artists than to other female peers.

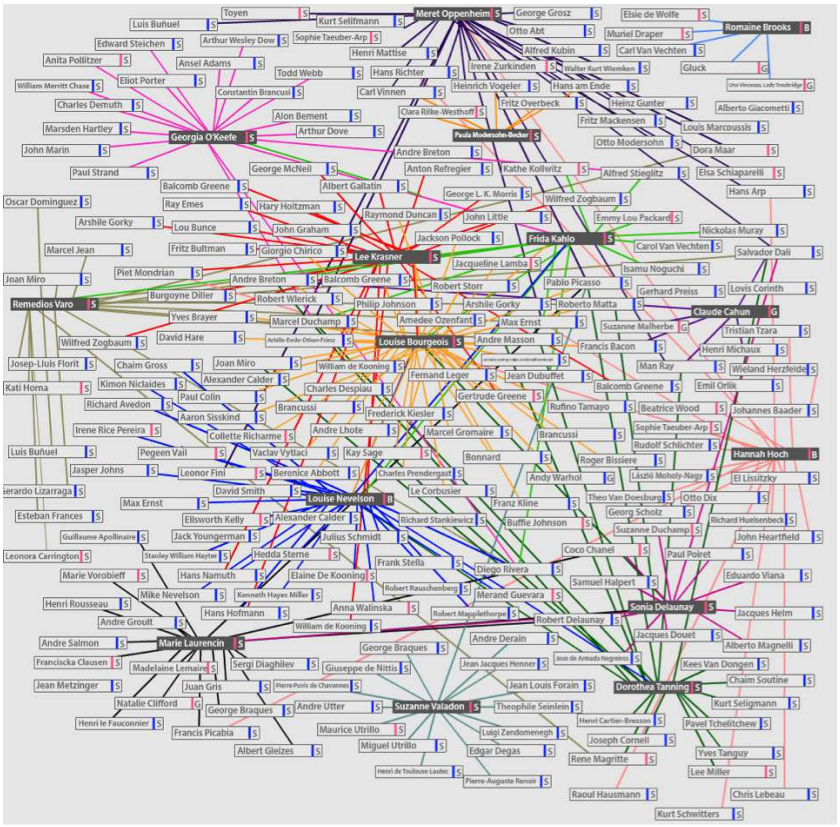


Figure 2.21: Social networks of 20th century female artists, [Ross, 2013]

An analytical view on the information present in the ULAN was presented by Schich et al. in [Schich et al., 2014], the main author also in general being an early pioneer in studying networks in the context of art history. Analyzing the migration patterns of notable persons based on their birth and death places, part of which were derived from the ULAN, the authors were able to show the usefulness of biographical data present in such databases for quantitative exploration of human mobility throughout the ages. This work, however, constructed a network of places based on birth and death information present in ULAN person records and did not consider the explicit social relationships available there. Deriving a link between the birth and death places stored in a person record can be considered as a bi-partite projection.

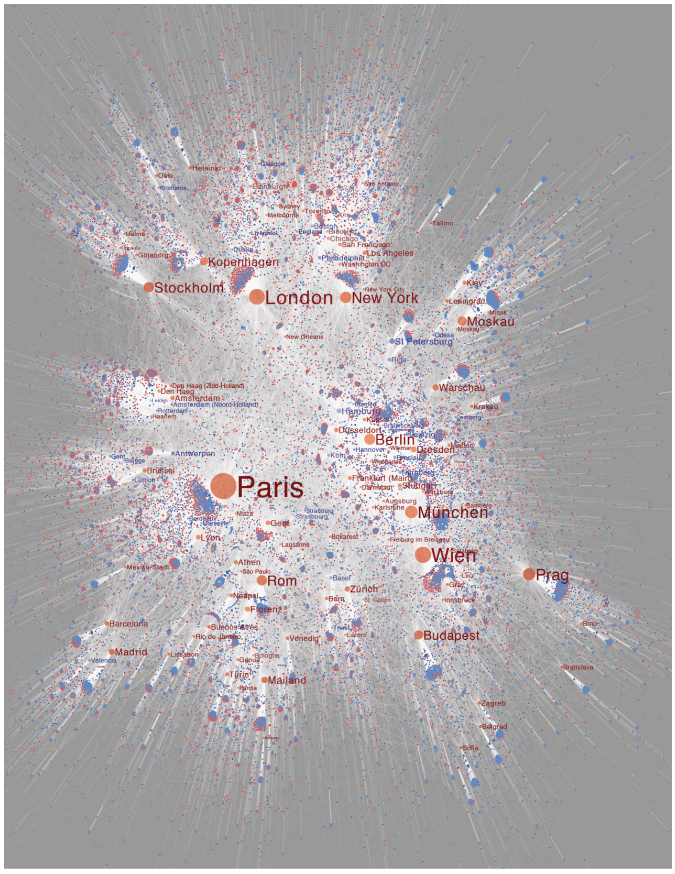


Figure 2.22: Birth to death migrations of artists, [Schich et al., 2014]

Another example for the application of networks in the context of art history is the analysis of buyer-seller interactions, focusing on the social context of the art market. Schich et al. in [Schich et al., 2017] provided an analysis of provenance data acquired from the Getty Provenance Index (GPI)[11], one of the most extensive collections of such data. The authors analyzed a multidimensional network of human or institutional "social actors" such as buyers or sellers, attributed artists, sales locations and dates along social, spatio-temporal and conceptual dimensions. Figure 2.23 shows a snapshot of buyer-seller interactions in art auctions, featuring many actors active in both roles, which were referred to as brokers by the authors. The network view revealed a clear regional clustering with concentrated local markets shown for Great Britain and Ireland, France and Belgium and the Netherlands, although the authors explicitly also pointed out that the regional clusters were all connected with each other and the European art auction market thus had to be considered as a single integrated system.

[11] `http://www.getty.edu/research/tools/provenance/search.html`, retrieved Sept. 8th, 2020

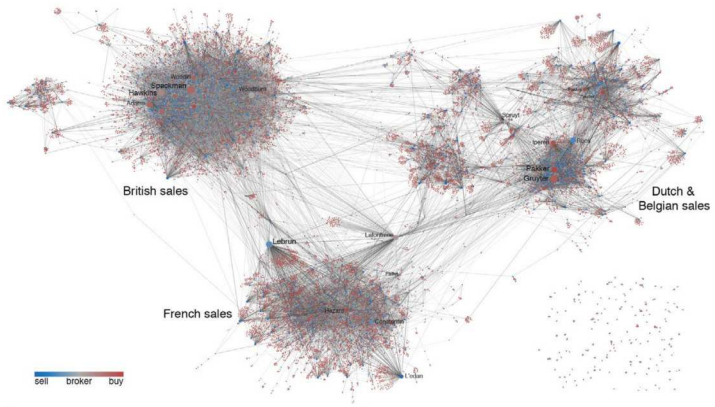


Figure 2.23: Network of art auction sellers, buyers, and brokers, [Schich et al., 2017]

A view on on social dynamics in the Austrian art scene centered around 1900 was described by Kaiser in [Kaiser, 2017]. This approach was based on bi-partite networks between artists and artists' associations, a snapshot of which shown in Figure 2.24. Related questions were to find groups of persons within specific institutions such as the Viennese Künstlerhaus and to ask for the dominant factors for their emergence, such as family relationships, affilition with certain academic schools or with influential collectors.

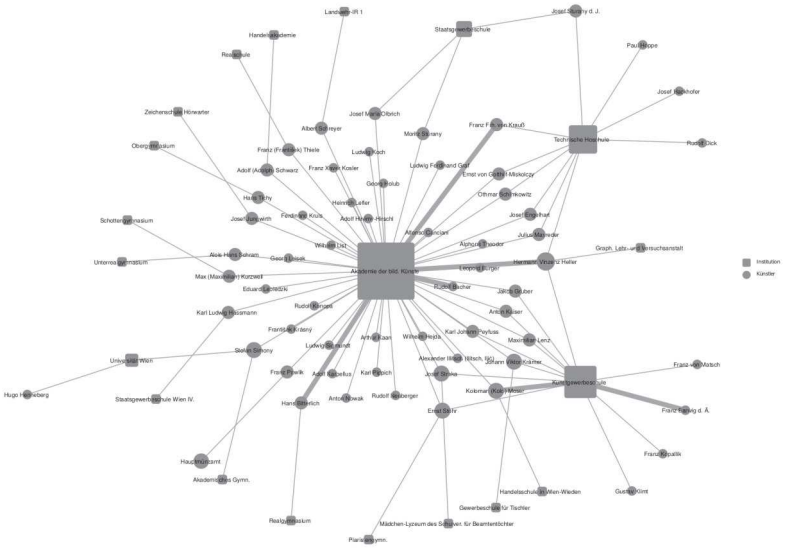


Figure 2.24: Network of Austrian artists and associations around 1900, [Kaiser, 2017]

Social dynamics were also the topic of Lincoln's study of 16th-18th century dutch printmaking networks, described in [Lincoln, 2016]. The author constructed networks between designers, printmakers and publishers based on their mutual involvement in the creation and dissemination of fine art prints, with the incentive to track how such artistic collaboration networks were subject to centralization, i.e. the tendency to be organized around highly influential individuals who maintained powerful positions allowing them to dominate and control the printmaking market. Besides showing interesting fluctuations in centrality over time, the author highlighted the importance of considering different data sources for analysis in order to mitigate specific foci of individual collections, shown in Figure 2.25.

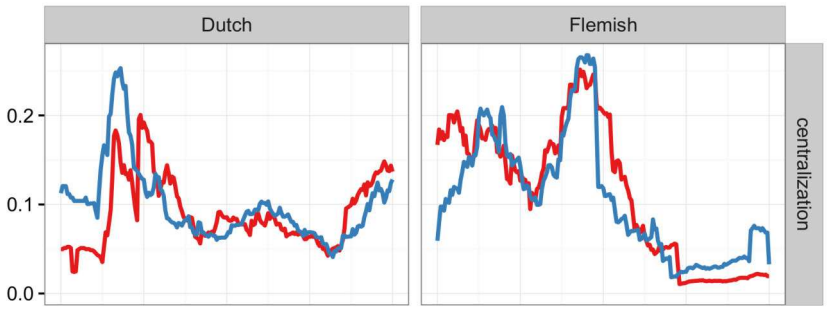


Figure 2.25: Changing network centralization in printmaking networks, [Lincoln, 2016]

In [Lincoln, 2020], Lincoln argued that computational network analysis can support art historians especially when they are dealing with amounts of evidence that exceed what individual scholars can process in their mind and when complex, unintuitive relationships are involved. He pointed out that many art historians are still hesitating to consider quantitative methods for their research, since such approaches require significant abstraction and structured description, standing in contrast to the classical means of studying unique objects of art and their multi-faceted contexts. Lincoln stated that the power of network analysis lies in its ability to bridge such, as he called them, "micro-scales" with "macro-scales", by which he referred to the study of larger-scale developments, by allowing the analysis on multiple levels of aggregation which can always be traced back to the individual connections they are composed of.

2.3 Summary and motivation

The examples discussed in this Chapter revealed a long tradition of network thinking in the context of art history, starting with the conception of artist genealogies at the time of the Renaissance and developing into large-scale contemporary network visualization and analysis approaches using digital technologies as tools for presentation and original research. Section 2.1 provided an overview on the most prominent historical examples for visual representations of such networks, ranging from early visual approaches to

describe scholarly interpretations of developments in the arts to more abstract models of interrelationships between artistic styles. Especially the visualization of developments across time played a dominant role in this regard and famous visualizations such as Barr's diagram of stylistic interrelations in Cubism and Abstract art demonstrated that this mode of visual communication caused significant resonance within art history and also across other domains.

As far as digital counterparts presented in Section 2.2 were concerned, the presentation-oriented approaches discussed in Subsection 2.2.2 and network-based art history research presented in Subsection 2.2.3 also took other aspects of the featured networks into account. This especially included topological aspects such as clustering, which were either studied for large-scale historical networks spanning multiple centuries or for individual snapshots taken at specific intervals or points in time. The research-oriented applications of networks showed a clear tendency towards bi-partite networks of various kinds, which usually always involved persons linked to different entities such as places, organizations or individual artworks. Especially the latter were used to derive relationships between persons, revealing patterns of trade as well as collaboration. As the example of the network between birth and death places showed, however, persons could equally well also serve as source for deriving links between other entities.

Most of the digital approaches used datasets collected for specific purposes, covering the exhibition or collection content in case of presentation environments or the underlying question in case of research work. Already existing, more "general-purpose" datasets such as the ULAN were mainly used for their originally intended use, such as improving collection search and recommendation, although Schich et al. in [Schich et al., 2014] for example had showed the potential of using such existing large-scale data sources in the digital art history research context. Interestingly, however, besides for collection search and recommendation, the explicit person relationships present in the ULAN had not yet been considered as network source for virtual presentations nor for original digital art history research. Moreover, only few approaches, such as for example Schreiber et al. in [Schreiber et al., 2008] who had used different Getty Thesauri including the ULAN and data from six different collections, followed a Linked Data approach by exploring the potential of interlinking heterogeneous datasets into hybrid, contextualized resources, which motivated to conceptualize further approaches in this regard.

CHAPTER 3

Combining Person Networks and Collection Data

This chapter is dedicated to design artifact **A1**, the experimental platform for integrating digitized artworks and their metadata with social network data from the Getty ULAN as exploratory fine art search environment. It also covers design artifact **A2**, the network of creator-subject relationships, which emerged during the design process. The conception and the design of the experimental platform was based on an analysis of existing approaches as presented in Section 2.2. Many of the approaches discussed there, however, did not exist/were not published yet at the time when the research prototype **A1** was conceived and considerations around the design artifact thus informed by approaches from before and around the period 2009-2011.

3.1 A 3D virtual art museum based on a historical social network

The survey of existing approaches to digital presentation of art history content discussed in Subsection 2.2.2 revealed that, in spite of the historical examples presented in Section 2.1, only few digital approaches focused on chronologically organized network views of developments in the arts and no existing one combined them with digital representations of artworks in an interactive setting. This suggested to explore the potential of such environments based on the idea that chronologically ordered network representations could serve as a "floor plan" for virtual museums, in a similar fashion to Barr's physical organization of his famous Cubism and Abstract art exhibition in 1937 which was expressed in his famous diagram. The availability of the Getty ULAN as data source providing many relationships between persons relevant to art history and their successful application as contextual data for collection search and recommendation suggested to consider their use for such an environment, which would in this case be based on

individual ties between persons rather than between artistic styles. In such a setting, person nodes would serve as anchor points to present individual artistic output which would be interrelated with those of others via visual representations of social ties such as teacher/student relationships which, besides serving as architectural element, would provide users with visual pathways along which they could move through the history of art, giving them guidance and a specific narrative for supporting their exploration.

A virtual 3D presentation environment was chosen due to the then current proliferation of 3D online approaches to virtual museums especially in the context of the Second Life platform ([Henning et al., 2007] [Hazan, 2010]), with the aim of transcending the then prevailing notion of recreating or replicating existing museum architecture there. The majority of existing approaches in this regard, however, were less targeted at abstract information visualization applications such as explicitly visualizing related networks or metadata pathways but rather dedicated to using the virtual 3D space for the simulation of real world environments. The latter can be used for the presentation of three dimensional objects such as sculpture and other physical artifacts and for the (re-)creation of virtual museum architecture for virtual exploration. Virtual museums realized via the platform Second Life for example, as discussed by Hazan in [Hazan, 2010], were usually created manually in a relatively tedious process. A remarkable example for such an art museum presence in Second Life was the Dresden Gallery island, described by Henning et al. in [Henning et al., 2007] offering a complete replica of the museum premises and all of the 750 exhibits on display there. As the authors stated, the presence was created by a team of up to 14 programmers and designers, revealing that only large institutions are able to support such endeavors.

One alternative to such hand built replicas are procedurally generated virtual museums, lowering the effort for creation and offering the possibility to dynamically change the presented content, although usually at the cost of rather generic looking virtual premises. An example for such a system which employed metadata to semi-automatically create room-by-room based 3D virtual museum environments was described by Amigoni and Schiaffonati in [Amigoni and Schiaffonati, 2009]. Curators were offered with means to cluster artifacts based by various metadata attributes which could subsequently be automatically assigned to dedicated virtual rooms which were represented using the Virtual Reality Modeling Language (VRML)[1]. Due to its rather generic nature, the automatically created 3D gallery space was only shown from the inside, acting more like a contemporary "white-box" gallery space putting the exhibits into the spotlight.

Information Landscape

While procedurally generated virtual museums such as described by Amigoni and Schiaffonati still sought to reproduce realistic museum architecture, approaches such as the work by Ruffaldi et al. described in [Ruffaldi et al., 2008] and already presented in Subsection 2.2.2 used the 3D space in a more abstract way. They allowed to use the

[1] https://www.web3d.org/x3d-vrml-most-widely-used-3d-formats, retrieved Sept. 8th, 2020

provided dimensions to arrange artworks and texts beyond the limitations of replicas of real world architecture, essentially using the content itself as architectural element. Pages of text and virtual image canvases served as walls, arranged sequentially or grouped via spherical or cuboid placements, allowing the virtual visitors to navigate this virtual landscape. Movement was either free or led via visualized associative links between the individual elements.

Ruffaldi et al. referred to the overarching concept behind their virtual museum approach as *Information Landscape*, thus introducing an already existing metaphor in the context of virtual presentations of visual arts content. Initially coined by Muriel Cooper et al. in a MIT Media Lab presentation video in [Cooper, 1994] Information Landscapes were introduced as new paradigm of representing and making navigable complex information in virtual 3D environments. One early application of this metaphor was presented by Wise et al. in [Wise et al., 1995] in the context of a system to visualize the content of document corpora. The "ThemeScape" view proposed there encoded the main topics present in a document corpus as mountains whose elevation correlated with each theme's "strength", while the location of each mountain was determined by the similarity between different topics. Shortly after, Andrews et al. in [Andrews et al., 1996] applied this metaphor to hierarchically and associatively linked information in a system called "Harmony Information Landscape". An early application of the Information Landscape concept in the context of cultural heritage was presented by Pampalk et al. in [Pampalk et al., 2002], whose "Islands of Music" system represented an approach comparable to the above mentioned ThemeScape view to exploring music collections.

In the context of the envisioned prototype, the idea was that the visualized topology of the network of ULAN person records should constitute the Information Landscape, put into shape by visual representations of the many relationships between the featured persons. Similar to the ThemeScape and Islands of Music approaches, the landscape's elevation was considered to be defined by a measure of importance of the individual persons, raising their individual nodes to form peaks of imaginary hills of different height. Virtual representations of individual artworks should then be grouped around the nodes of their creators.

Creating a virtual environment based on network visualization required the choice of an appropriate layout algorithm and as outlined in Section 2.2, there were a number of different fundamental approaches in this regard. Due to the design consideration to order the network's nodes chronologically, a constrained, layered graph drawing technique appeared the method of choice in this regard. In the case of the intended Information Landscape, the node placement along one of its two horizontal dimensions would thus be determined based on the chronological ordering, while the placement along the other one would be based on the minimization of edge crossings. The vertical dimension would then be independently determined by some measure of node importance. One advantage of choosing a layered layout was the range of existing implementations, amongst them Gansner et al.'s. well known `dot` layout algorithm described in [Gansner et al., 1993], available via the Graphviz package[2].

[2] `https://www.graphviz.org/`, retrieved Sept. 8th, 2020

In order to avoid performance issues, it was decided to limit the visualized layout to so-called ego-networks around a chosen seed person. The nodes of such networks consist of the seed node (ego) and all other nodes (alters) connected to the ego via pathways of predetermined maximum length (steps), while its links are the set of all ties between the ego and its alters and all the ties present between alters. Ego-networks only featuring alters immediately connected with the seed node are called one-step ego networks, while two-step ego network contain the ego, all its directly connected alters and all the alters of each alter, etc. Two-step networks were considered to represent reasonably sized network subsets in the context of the prototype.

3.2 Data sources and initial data processing

The person records from the Gerry ULAN with their associated metadata and mutal links served as source for the social network layer constituting the Information Landscape. As source for digitized artworks and relevant metadata, the Web Gallery of Art[3] (WGA) could be identified. It featured 24,000 digitized artworks by more than than 2,900 artists covering a timespan between approximately 1000 AD - 1900 AD. At the time of the initial conception of the prototype, this collection represented one of the few online repositories of artworks offering a downloadable representation of structured image metadata and links to reasonably high resolution digital surrogates.

Merging the ULAN with the WGA

In order to be integrated, the ULAN artist data first had to be matched with the WGA creator information. This procedure, however, was not straight-forward: While each ULAN record featured distinct fields for artist preferred name, alternative name(s), birth and death dates, etc., the WGA data only featured an "author" field featuring a name and a "born-died" field consisting of life date information string of the form `(<birthdate>, <birthplace>, <deathdate>, <deathplace>)`. As often encountered with manually edited date and place information, the terminology for entering the appropriate information for each artist was unfortunately not fully consistent, featuring many question marks and acronyms expressing uncertainty. The Google Refine (now Open Refine[4]) tool was used to extract and clean all different patterns encountered. Matching the WGA artist name with the ULAN name (preferred or alternative) and corresponding birth/death dates, it was possible to match more than 80% of the WGA artists with the ULAN.

Data representation

In order to stick to emerging digital cultural heritage standards and not to re-invent the wheel, an existing data architecture was re-used to represent the integrated data:

[3] `https://www.wga.hu/`, retrieved Sept. 8th, 2020
[4] `http://openrefine.org/`, retrieved Sept. 8th, 2020

The MultimediaN E-Culture[5] demonstrator, described by Schreiber et al. in [Schreiber et al., 2008], offered a semantic search prototype based on the semantic integration of the Getty vocabularies with object metadata from a variety of Dutch museums and provided an existing model to integrate the ULAN data with metadata about artworks. This work closely supported the emerging philosophy of the Web of Data by consistently using Semantic Web Standards such as the Resource Description Framework (RDF)[6], whose network-like structure aligned well with the intended application for representing the artist network. Following this approach, the WGA image metadata was transformed to an RDF representation of the VRA Core[7] standard used there. The proprietary XML representations of the Getty ULAN were converted to RDF using a freely available tool[8] provided by the authors. As of today and described by Baca and Gill in [Baca and Gill, 2015], however, the Getty vocabularies can now be retrieved directly from the Getty foundation in form of RDF[9].

3.3 Design iterations

The evolution of the prototype was comprised of an initial proof of concept version followed by a significantly redesigned second version. The different stages are described below. Due to its flexible scriptability based on the C# language and the availability of a free version, the Unity3D[10] game engine was chosen for developing the 3D environment.

3.3.1 Proof of concept

The first version of the prototype served as proof of concept implementation and re-used the Prolog based ClioPatria[11] Semantic Web server infrastructure developed and used by Schreiber et al. in [Schreiber et al., 2008] as backend for storing and retrieving the integrated RDF data in their MultimediaN E-Culture demonstrator. The approach to visualize the ULAN network was realized using a ClioPatria feature which was dedicated to data management purposes and not directed at end users. This feature allowed data managers and server maintainers to visualize parts of the structure of the RDF graph stored in a ClioPatria instance by converting a section of the RDF graph to the GraphViz `dot` representation language, subsequently calling the associated `dot` application to

[5] `http://multimedian.project.cwi.nl/`, retrieved Sept. 8th, 2020
[6] `https://www.w3.org/RDF/`, retrieved Sept. 8th, 2020
[7] `https://www.loc.gov/standards/vracore/`, retrieved Sept. 8th, 2020
[8] `http://multimedian.project.cwi.nl/software/gettyconvert/`, retrieved Sept. 8th, 2020
[9] `http://www.getty.edu/research/tools/vocabularies/lod/`, retrieved Sept. 8th, 2020
[10] `https://unity3d.com/`, retrieved Sept. 8th, 2020
[11] `https://cliopatria.swi-prolog.org/home`, retrieved Sept. 8th, 2020

calculate a network layout using the layered graph layout algorithm described in [Gansner et al., 1993] and to visualize its output.

Since ClioPatria's built-in visualization feature was targeted at representing all the stored RDF statements stored there, the ClioPatria Prolog implementation had to be patched in order to limit the featured information to the person-to-person network information present in the integrated dataset. The patched version therefore filtered the visualization of the RDF graph to contain only the ULAN relationships and also added constraints so that the resulting `dot` representation layered the included person nodes chronologically by birth date. Artwork metadata, only the image URLs at this first stage, were attached as node attributes in the resulting `dot` output, which was subsequently laid out via the `dot` application and stored in a Scalable Vector Graphics (SVG)[12] file. This format was found to be suitable for storing both the network itself and additional information attached to its individual elements in form of XML, which could easily be processed by the visualization client.

The proof of concept prototype client was designed to receive the SVG XML file from the ClioPatria backend and to extract the network layout embedded there. The person relationships were subsequently drawn on the virtual plane, elevating the individual person nodes based on the number of relationships they had with others, using it as a measure of person importance. For each featured person, digitized representations of his or her artworks were downloaded via the attached image URLs and placed in a circle around the respecticve network node. A screenshot of the proof of concept output is shown in Figure 3.1, centered around the elevated node representing the Italian Renaissance artist Perugino. The view is directed into the future, revealing important artists of succeeding generations such as Perugino's pupil Raphael, shown center left in the background. Raphael's importance in art history becomes visible through his high elevation compared to the other nodes in the scene, resulting from the many relationships present in his ULAN record.

The proof of concept environment was informally discussed with potential users. Initial feedback was positive regarding the possibility to "fly through art history", although the approach to visualize a social network by showing artworks around generic node shapes at each person's location appeared less convenient. One user moreover stated that the layout with its many parallel lines resembled a printed circuit board, calling for a "more organic" layout. The most requested additional feature was an overview map for easier navigation and a possibility to reset the view after getting lost, which happened due to the complete freedom granted to users to move in all directions.

[12] `https://www.w3.org/TR/SVG11/`, retrieved Sept. 8th, 2020

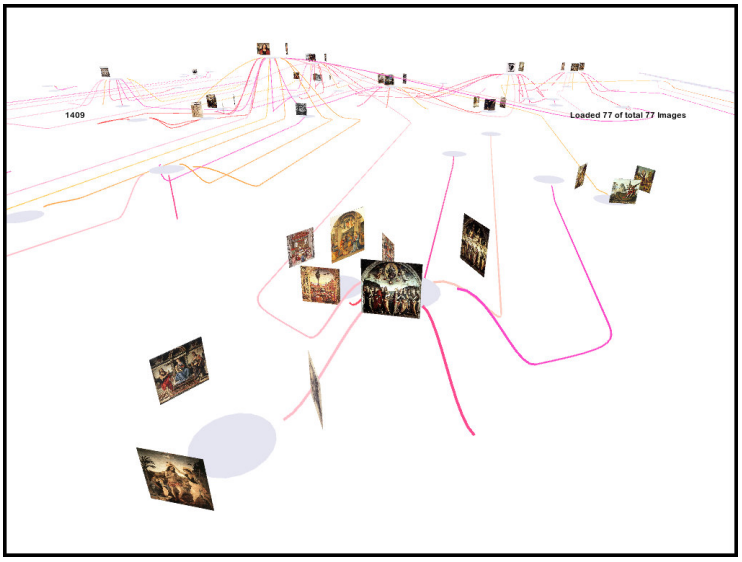


Figure 3.1: Proof of concept prototype

3.3.2 Redesign

The feedback gathered from discussing the proof of concept prototype suggested significant changes for its next iteration. This included to rethink how the different actors in the network could be presented in a more individual manner, to reconsider how the overall network could be drawn in a more "organic" way and to find a usable approach to providing an overview map for better navigation. Moreover, evolving from the proof of concept stage prompted for a more flexible, state of the art system architecture.

Based on the planned changes, the redesign thus involved multiple aspects of the prototype, which included a full redesign of the technical implementation of the client-server infrastructure and a number of significant changes to the way how the information was presented to the users. Since the approach to use Graphviz tools for generating the network layout on the server-side did not allow a close control over the layout process, it was decided that these processing steps should take place client-side instead. This decision in turn negated the initial reason to use the rather arcane Prolog based ClioPatria server, which required deep knowledge of the Prolog ecosystem to modify or add new features there. The main changes introduced to the new prototype iteration on infrastructure level were thus to establish a new data backend based on a more generic database engine and to find ways to implement the network layout on the client-side.

New database backend

As far as the data backend was concerned, the free version of the Openlink Virtuoso[13] server was identified as a flexible platform for storing RDF based data and making them available either via standard SPARQL queries or by encapsulating such queries behind Web API protocols such as SOAP. Already available in form of RDF, it was straightforward to migrate the integrated ULAN and WGA data to the new data backend. Virtuoso's transitive SPARQL extensions — in the meantime to a large extent obsolete due to the wide spread adoption of SPARQL 1.1 — were used to set up a query allowing users to input a "seed person" and to retrieve its ego network of a predetermined depth. Although it would have been technically possible to directly retrieve the data via SPARQL, it was decided to encapsulate the respective call behind a stored procedure exposed via SOAP in order to restrict possible queries against the server to the predetermined functionality. A second SOAP request was set up triggering a SPARQL query to retrieve all stored artworks for a given creator. These two requests formed the basis for the data retrieval from server to client. A diagram representing the described architecture is shown in Figure 3.2.

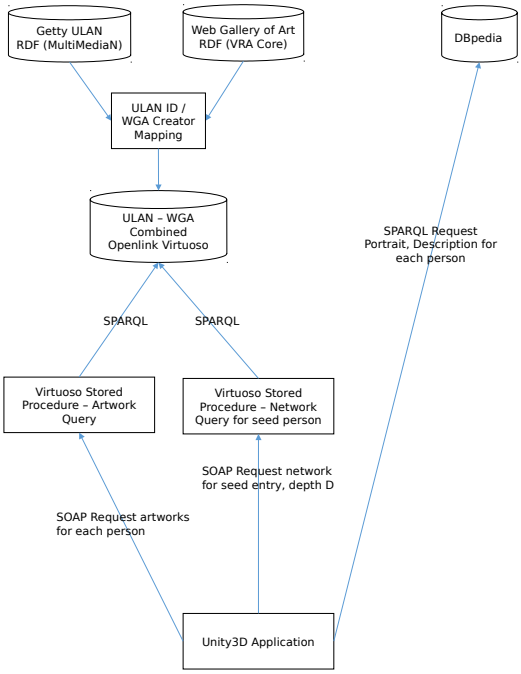


Figure 3.2: Data flow

[13] https://virtuoso.openlinksw.com/, retrieved Sept. 8th, 2020

Portraits

The data architecture view shown in Figure 3.2 reveals that the redesigned prototype issued one more query for each person featured in the network. This was due to the design decision to change the appearance of person nodes in the visualization: Instead of ordering the artworks associated with the person around a generic, anonymous node representation, the redesigned network visualization should place portraits of the represented persons on virtual canvases located at the person nodes in order to give them a more individual appearance. It was thus necessary to identify a source for portraits, since the ULAN records did not feature links to visual representations of the featured persons.

A search for possible sources for portraits led to Wikipedia biographies. This data collection appeared promising because it was freely available, featured many person depictions in biographical articles which were retrievable via standardized SPARQL queries in a structured way due to the efforts of the DBpedia project introduced in Section 1.1. A study by Clough et al. presented in [Clough et al., 2009] had furthermore identified about 15,000 biographical records in the English DBpedia version to cover persons also present in the ULAN. Together with the previously mentioned advantages, this finding led to the inclusion of DBpedia as data source for portraits.

For each of the persons present in the network retrieved from the Virtuoso server, the client therefore issued a respective SPARQL query against the English DBpedia SPARQL endpoint in order to retrieve the persons's portrait image url provided via the optional `foaf:depiction` property there. A very simple matching strategy was chosen for identifying ULAN persons in DBpedia via exact lowercase matches between ULAN preferred or alternative names and the respective DBpedia label, taking wrongly identified matches into account at this stage. In addition to the depiction, a short biographical sketch was queried from DBpedia via the `dbo:abstract` property.

A screenshot of the updated prototype is shown in Figure 3.3. The view is centered on the famous Venetian Renaissance artist Titian, featuring a portrait of him retrieved via DBpedia as described above. As visible in the background, such depictions could also be identified for a number of other persons featured throughout the scene, while dummy images with person silhouettes and a question mark were assigned to persons where this was not possible. Immediately below the Titian node there is a node for the Renaissance architect Antonio da Sangallo the Younger, whose DBpedia portrait link pointed to a woodcut portrait illustrating his biography in Vasari's famous Vite. Portrait illustrations from this important work were found to be quite often assigned as portraits to Renaissance artists and architects in Wikipedia and thus often appeared in the virtual scenery when it was spawned around persons from that time.

Updated, client based layout

As it becomes visible in 3.3, the layout of the redesigned prototype was significantly updated compared to the proof-of-concept prototype shown in Figure 3.1. The changes to the layout were made possible by implementing and using a custom layered graph

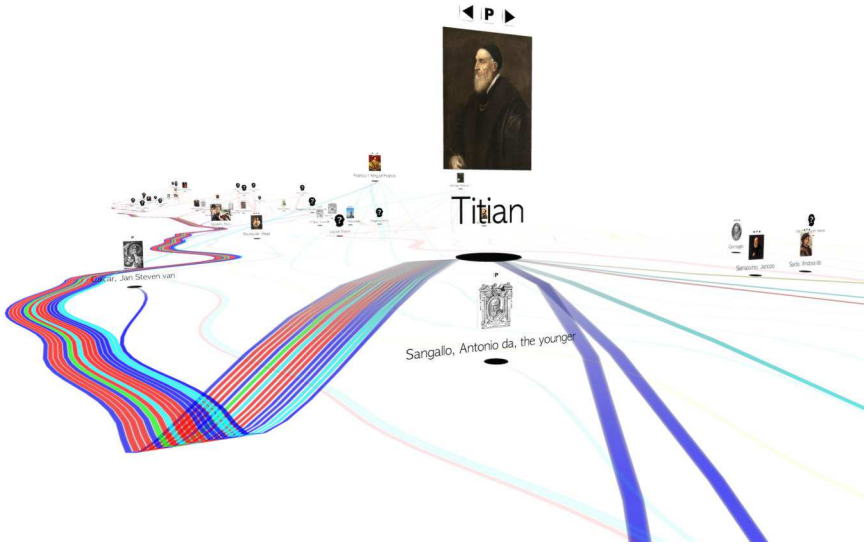


Figure 3.3: The 3D landscape around Titian

drawing algorithm instead of the Graphviz `dot` algorithm used previously. The basis of the custom layout algorithm was nevertheless based on a simplification of `dot` and implemented some of the steps described in [Gansner et al., 1993].

From the four pass layout processing proposed by Gansner et al., the first pass could be omitted, since all nodes were to be placed in distinct layers determined by their birth date. The second pass dedicated to ordering the nodes for minimizing edge crossings was implemented as described in the paper. The third pass was strongly simplified by center aligning the nodes in each layer, while the fourth pass for drawing the links was again implemented as described.

The main addition to the custom algorithm was the introduction of link bundling. This approach seeks to minimize visual clutter in graph layouts by bundling together segments of the visual representations of links according to different criteria, such as having similar start- or endpoints. A recent proposal for a link bundling approach in the context of layered graph drawing was made by Pupyrev et al. in [Pupyrev et al., 2010], based on a minimization of "virtual ink" used to draw the graph layout. In the context of the social network of persons related to art history, it was decided to follow a different approach to link bundling, based on the conception that it was of interest to visually show the amount of influence a person had on following generations by bundling together all of his or her connections pointing into the future. In the context of the Information Landscape metaphor, persons with many connections to younger persons would be the sources of "rivers of influence" emanating from them, bifurcating along the way every time when one of the bundled connections reaches its destination.

Figure 3.4 provides a bird's eye view on the updated Information Landscape shown in 3.3, the time axis flows from left to right. The modifications in the custom layout algorithm become clearly visible, including the results of the link bundling process and the simplified, centrally aligned vertical positioning of the nodes in each timestep layer, resulting in a much more compact representation of the graph, although at the cost of losing straightened links. The "river" emanating from Titian is visible in the top center of the Figure. This representation raised an interesting association with the ThemeRiver approach to visualize the temporal development of topics in large document collections, proposed by Havre et al. in [Havre et al., 2002]. The bundeled and compacted graph representation of the ULAN network appeared like the "network version" of the ThemeRiver approach, taking the number of "active" connections emanating from a person, represented through the respective bundle's thickness, as a measure of strength of his or her influence at each timestep.

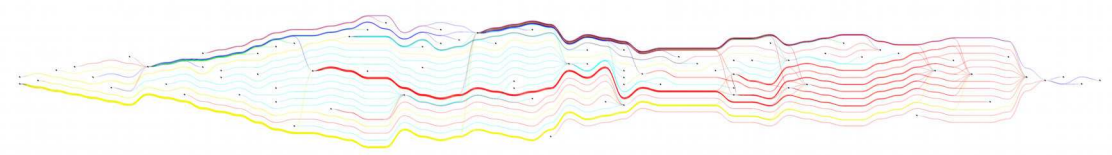


Figure 3.4: The 3D landscape around Titian, bird's eye view

In order to test a representation of the updated version without the simplified, center-aligned placement of nodes within each timestep layer, the within-layer coordinate assignment pass of the custom layout algorithm was re-implemented according to a procedure described by Brandes and Köpf in [Brandes and Köpf, 2001]. The result is shown in Figure 3.5, visualizing exactly the same network as shown in Figure 3.4. Although the effect of the link bundling remains visible, the layout features the same "circuit board"-like appearance which was mentioned rather negatively for the proof-of-concept prototype. It was thus decided to stick with the compact, non straightened version of the layout.

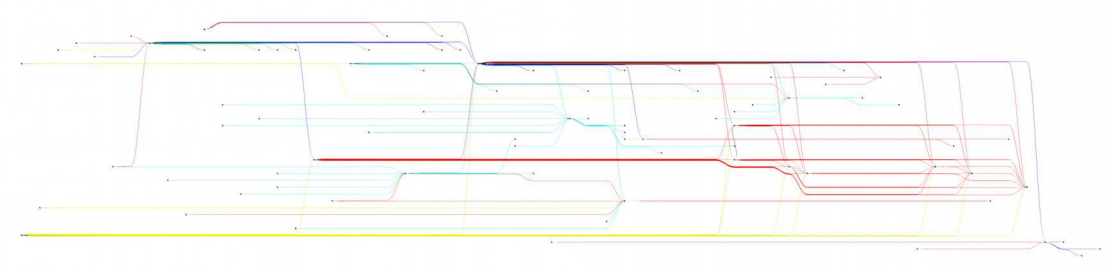


Figure 3.5: The 3D landscape around Titian, bird's eye view, straightened layout

UI features

A number of UI features were added during the redesign, they are shown in Figure 3.6. Additions included an area on the left side of the screen featuring the DBpedia description of the currently visited person, the title of the currently displayed image and the number of artworks available. A legend for the color code used to visualize connection types was provided on top of the screen, clicking on each connection type enabled users to disable/re-enable the display of connections of the respective type. Another set of buttons was located at each image canvas placed for each of the featured persons. This set of buttons enabled users to switch the canvas between portrait and artwork mode, the latter allowing them to cycle through the WGA artworks associated with the currently visited person.

A much requested addition was an overview map, visible on the right side of the screen. The map provided an abstracted bird's eye view on the network layout with omitted links, featuring letter symbols at the positions of the persons present in the current scene, representing their role in art history as encoded in the ULAN ("a" for artist", "p" for patron, "s" for sitter, etc.). Only the labels of the currently visited person and all its immediate connections were rendered on the map, the latter again colored by connection type. A slider allowed users to change the temporal resolution of the map, indicated via the year scale shown on its right hand side. One significant feature of the overview map was the ability to "teleport" to another location by clicking on one of the person symbols or labels rendered there. Upon click, the first person camera user view was moved to the new destination in an automated flight-like sequence and the map view scrolled accordingly.

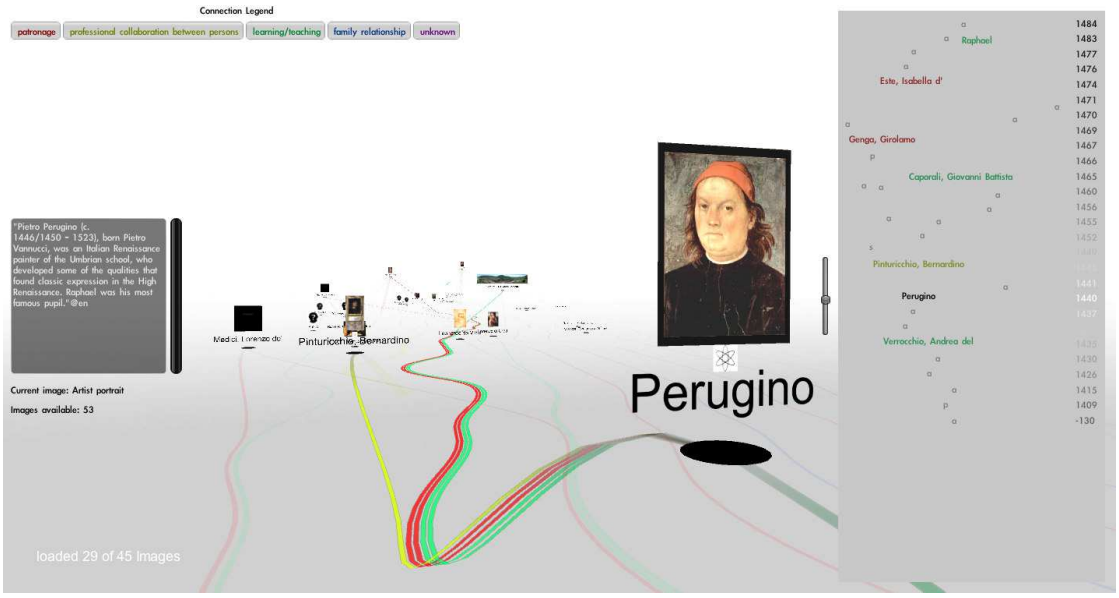


Figure 3.6: The 3D landscape around Perugino, including UI features (**A1**)

Navigation

The mouse/trackpad served as main input device. The default mouse mode enabled users to rotate the virtual camera horizontally and vertically and thus to "look around", while pressing the button served to interact with the GUI elements such as buttons and the map. Movement through the scene was made possible by simultaneously keeping the space button pressed, during which the left mouse button resulted in forward, the right button in backward movement of the first person view. An important feature requested during the informal feedback-gathering sessions was the ability to reset the view in case a user got lost. A dedicated button was now provided for that purpose, resetting the view to the last visited person upon click.

Entry point

Before the scene was generated, users were provided with a search box where they could enter the name of the seed person whose ego-network should be visualized. An auto-completion feature supported the identification of persons available in the ULAN. Moreover, a number of entry suggestions for famous Renaissance and Baroque artists were placed as buttons around the search box. Once a person was selected, the generation of the scenery was initiated, which included the ad-hoc retrieval of the network and artwork information from the system server and the retrieval of portrait information and person descriptions from DBpedia. The download of the sometomes hundreds of images took place in an asynchrous manner which allowed users, whose view was initially placed in front of the representation of the chosen seed person, to start navigating the scene without having to wait for the image download to finish.

3.3.3 Portrait relationships

A significant number of ULAN person records was found to represent non-artist individuals such as patrons, emperors and other members of aristocracy or the clergy, usually connected to ULAN artist records either via patronage or family links. Being part of the ego-networks of many artists, such persons thus often became part of the Information Landscape as well, which was conceived to be a valuable addition providing important historical ("Learning Museum" like) context not available in standard modes of presenting digital cultural heritage. Similar as for artist persons, the redesigned prototype searched DBpedia for matching portraits and short person descriptions, displaying the results accordingly. Not being creators of WGA artworks themselves, however, limited the displayed information for these persons to the content derived from DBpedia.

As it turned out, many of the Wikipedia portraits downloaded for such non-artists via DBpedia were actually created by artists with whom they also were directly or indirectly connected via ULAN relationships. The creators of their portraits were thus usually also part of the same ego-network and thus simultaneously present in the currently viewed Information Landscape scene. Since in many cases, the retrieved Wikipedia portraits

were also present in the WGA artwork collection, this motivated to identify portraits of non-artist persons in the WGA artworks of the artists simultaneously present in the currently viewed ego-network.

A straight-forward approach was to search for ULAN non-artist names in WGA artwork titles upon loading of the scenery. This simple strategy already returned numerous results: Figure 3.7 shows four different WGA artworks featuring the name of the famous Renaissance patron Isabella d'Este in their titles, created by four different creators all present in the same scenery. In such cases, the non-artist person canvases in the scenery were augmented with buttons allowing visitors to scroll through the associated artworks, highlighting the respective creator on the overview map by using an enlarged font there. Using this approach to associate portraits with non-artist persons allowed a striking discovery regarding the completeness of ULAN relationships: The top right view of Figure 3.7 shows a portrait sketch of Isabella d'Este made by no other than Leonardo da Vinci, who is also highlighted in the map overview. While the creators of the other three artworks were all associated with her via patronage links in the ULAN, there was[14] no recorded relationship there between Isabella and Leonardo, which is remarkable due to their historical importance to the field of art history. The approach to associate two persons via artworks which were created by one of them and featured the other thus turned out to enable the identification of gaps in the ULAN, highlighting the benefit of integrating the ULAN with the WGA.

Figure 3.7: Various portraits of arts patron Isabella d'Este

[14] And still is, see http://www.getty.edu/vow/ULANFullDisplay?find=isabella+d+este&role=&nation=&prev_page=1&subjectid=500115132, retrieved Sept. 8th, 2020

3.4 Evaluation

Besides informal feedback gathered from colleagues and other potential users during and between the different design iterations, the prototype was more formally tested by three art professionals and their impressions collected based on semi-structured interviews. Moreover, the prototype was presented to domain experts at various conferences, one of which, Computers and the History of Art (CHART[15]), was explicitly dedicated to digital technologies and art history. In addition to the results from user testing, the art history domain experts in the audience of the CHART conference provided valuable feedback for further research directions.

3.4.1 User evaluation

Three test subjects professionally involved in the arts, two artists and one art theorist/art historian, were invited and agreed to test the prototype. The test scenario was set up on the basis of a semi-structured interview[16] performed before, during and after the interaction with the prototype.

Interview structure

- **Introductory questions**
  The subjects were first introduced to the background and motivation of the interview, then they were asked about the relevance of art history to them and what sources they usually used to gather relevant information in this regard. After that, they were asked if they ever had used the Internet for searching information about art history and if yes, which Websites they used for that purpose.

- **Test session**
  The test session was initiated with an introduction to the prototype and its intended purpose. The subjects were guided through the different features of the prototype and then left alone to explore the environment but provided with support whenever needed. Two rounds of interaction were planned, one for exploring the ego-network of a Renaissance artist due to the rich data available for that period, another round based on free choice of a seed person.

- **Post-questions**
  Subjects were asked if the software helped them to learn new aspects about the displayed artist and/or his or her ego network and if anything they knew about was missing. Then they were asked if their interest was also drawn to the individual relationships between the persons present in the scenery.

[15] https://web.archive.org/web/20170612044025/http://www.chart.ac.uk/, retrieved Sept. 8th, 2020

[16] The transcripts of the interviews are provided online at https://github.com/d0rg0ld/thesis

After that they were asked if they liked using the software and what aspects they found particularly interesting/the software was particularly suited to convey. Moreover, users were asked whether they felt comfortable with the user interface, how well they were able to navigate the 3D scene and if they had suggestions for changes or extensions to the system. The interview was closed with the question if users could imagine using the prototype to search for information related to art history and if they had any additional remarks.

Results

- **Summary of answers to introductory questions**
  Given their professional background, it was no surprise that art history was relevant to all three users, although the artists and the art historian revealed differing perspectives in this regard. The art historian mentioned that he was especially interested in the context of artworks, including biographies, while the artists were mainly interested to find relevant artworks. One of the artists stated that his interest was more directed at contemporary artists but that it was always important to be able to research older works as well, since "artists always reference past things", especially image material, that he was always looking for artists creating visually/technically similar artworks. The other artist preferred to "think of his own art history" but nevertheless acknowledged its importance as far as things/events from the past are concerned.

  Asked about the sources they usually consulted for art history information, all three subjects mentioned books as primary source, although one of the artists went further and referred to other persons as primary source, taking book recommendations from them. As far as the Internet was concerned, the subjects made a distinction between using it to find information and to retrieve images. Regarding the latter, there was a general agreement that the amount of available relevant imagery was very limited and that the quality of the few available reproductions was generally very low, one subject mentioned Google image search as tool to find relevant pictures, but was rather disappointed by the results, describing them as "muddled". All subjects also mentioned to use Wikipedia for both information and pictures, one artist stated that he used it because it was usually amongst the first search results but was often disappointed that articles did not contain enough pictures, while the other artist used Wikipedia when he already knew the name of the artist he was interested in. The art historian said he looked up Wikipedia "whenever he did not remember something" but generally referred to it as "not deep enough". Overall, he described the Internet to support "Rhizome thinking"[17] which he characterized as "not deep" and superficial.

[17] The term Rhizome was coined by Deleuze and Guattari in [Deleuze and Guattari, 1992], referring to their proposal to conceive thinking and the organization of knowledge to rather follow rhizome-like structures than strict taxonomies

Internet search played a role for all subjects, all of them mentioned the Google search engine as main entry point, the art historian mentioned Google Books as convenient way to find relevant books in the Internet. One artist mentioned that he only used the Internet when none of the persons he usually consulted was available or no related book was at hand, but generally referred to it as "young, genius way to bounce around" that he sometimes ended browsing around after looking for something specific.

- **Summary of prototype walkthrough and final discussion**

As far as the overall approach to use the 3D Information Landscape metaphor was concerned, the subjects mentioned positive as well as negative aspects. The art historian initially did not understand the function of the Information Landscape, calling for a more efficient explanation at the beginning of the exploration. The two artists stated that exploring the scenery was like "sitting in a car" and liked the associated cruising metaphor, one of them referred to the visualized relationships as "streets" and found the approach to fly through time to be a "nice way to do browsing", providing a perspective on the future of the Internet as a whole.

All three subjects were interested in the relationships between the featured persons, two subjects positively mentioned the approach to elevate persons based on the number of connections and the resulting mountain metaphor. The latter was not immediately appreciated by the art historian subject, although considered more positively at the end of the session, suggesting that it takes time to get acquainted with the scenery. Two subjects started to explore relationships between artists in detail and wanted to know more specific information such as whether a shown family relationship was between father and son. One artist mentioned the patronage relationships to be particularly interesting and useful, but also mentioned the simultaneous presence of many persons in the same scenery as potentially distracting. The art historian stated that it was interesting for him to learn about connections for things he already knew but only had basic knowledge about, highlighting that users needed some a priori background knowledge to make sense of the scenery. The color coded lines caused mixed reactions, one subject mentioned them to be helpful while the art historian thought of them to be distracting and suggested to make it possible to selectively highlight only one connection at once.

Concerning navigation, all three subjects managed to look around and move through the scenery using the mouse but encountered a number of issues in doing so. One subject mentioned that free traveling through the scenery was too slow, especially along connections between persons many years apart. A problem encountered by all subjects was that they got lost easily, losing orientation and the sense of location, which had already been tried to be mitigated via the provided "reset view" functionality. Although considered important and helpful in this regard, the latter function didn't come without its own problems. The most prominent issue was where to actually reset the view to, since the used approach was to "remember" the

last visited person for that purpose, triggered by passing in close vicinity to the respective node representation in the network, which also happened accidentally once in a while. Moreover, users sometimes clicked a destination on the map to teleport to and then wanted to be moved back to the point of departure.

The map itself was considered as helpful for fast traveling and getting overview on the scenery but also as a too abstracted. Subjects missed the display of person relations as lines on the map and initially did not understand the different symbols representing the different person roles. After additional explanation, however, they managed to interpret them correctly, suggesting the importance of providing a respective legend.

All three subjects quickly reached the "outskirt" of the two-hop ego network around the root person and wanted to explore further. One subject mentioned that especially the importance of famous persons with potentially many connections beyond the two steps from the current root person became deemphasized, since these connections did not influence the elevation and were not shown in the landscape as long as it was not reloaded centered on such persons. This again suggested that the mountain metaphor was in principle accepted as means to highlight importance, immediately bringing to attention famous persons who were expected to be elevated but were not due to the stated reason.

A general observation was that all three subjects were more interested in modern and contemporary artists, which, although usually present in the ULAN but with much less relationships to others, had no associated imagery in the WGA and also significantly less matching portraits in DBpedia. This revealed a much discussed aspect of digital cultural heritage, referred to as "20th century black-hole" for example in Europeana,[18] resulting from the current situation that most of the content created after about 1900 is still under copyright and thus not made available to the public without royalties. Another aspect was revealed through one of the artists unsuccessfully trying to find himself in the ULAN, prompting the question if it was possible to enter additional information into the database. As far as the digitized artworks were concerned, the art historian mentioned that their (in his eyes bad) quality was similar to a collection known in German as 10,000 "Meisterwerke" (masterworks), made available as Public Domain in 2005 under the name "Yorck project"[19], it could, however, not be verified if the WGA content indeed featured the same reproductions, although it was quite likely that this was the case.

- **Suggestions for changes / additional features**

A number of possible changes and extensions to the prototype were suggested during the interaction session. The most important feature mentioned by all subjects was

[18] https://pro.europeana.eu/post/the-missing-decades-the-20th-century-black-hole-in-europeana, retrieved Sept. 8th, 2020

[19] https://commons.wikimedia.org/wiki/Commons:10,000_paintings_from_Directmedia, retrieved Sept. 8th, 2020

a history function enabling users to see all persons visited during a session, one subject moreover mentioned the ability to create a personal collection. As far as the 3D view was concerned, subjects suggested to show an initial flyover across the scene to provide a better sense of orientation and to include time markers not only on the map but also embedded in the landscape so that users would always know about their current temporal position and direction. Moreover, they suggested to show the name of the root person as title for the current scenery and demanded to be able to center the view on the currently viewed person canvas in order to get a better focus on the image content.

Another aspect demanded by all subjects was the ability to superimpose 2D popup windows in the scenery, providing additional context information such as links to Wikipedia or other detailed sources and information about related, current museums shows. Users also suggested to be able to look at a 2D version of the network, expecting it to enable a better overview on the relationships as a whole and quicker navigation between connected persons. Two subjects also suggested the addition of a geographic map view, giving a spatial overview on the person relationships and the ability to show the locations of the institutions holding the original artworks.

- **Conclusions drawn from the user evaluation**

The user evaluation with the domain experts revealed that there was significant interest in the social network view on art history. There was higher interest in 20^{th} century and contemporary artists which, however, were found to be twofold underrepresented in the used data. On the one hand there were notably less inter-person relationships present in the ULAN records for these persons compared to artists from earlier periods, which was interesting because it could be expected that more information and sources were available for more recent persons. On the other hand, the available image material, both regarding digitized portraits and reproductions of artworks, was far more sparse for this group of persons, which was most probably due to copyright reasons.

As far as the 3D Information Landscape approach was concerned, the users acknowledged that it represented a new way to explore data about art history and found the mountain metaphor to be an adequate means for representing importance. The idea to "fly" through a landscape of art history was considered as appealing by two of the three subjects, although the actual interaction with the prototype revealed usability issues with respect to navigation and the limitations of the restricted two-hop neighborhood of one root person. As far as the interaction with the artwork canvas for switching between portraits and artworks was concerned, the users appeared to prefer more traditional 2D UI means, calling for the ability to display respective popup windows with contextual information. Especially the observation that all three subjects demanded 2D visualizations of the social network suggested to further explore this option.

3.4.2 Additional expert feedback

Additional valuable feedback was gathered from art history domain experts after presenting the prototype at a conference dedicated to using computers in the context of art history. Besides positive feedback on the approach to combine different complementary data sources into one presentation, the discussion focused on the completeness of art historical sources. Confronted with the social network based view on art history, one art historian mentioned that a significant part of the available knowledge on Italian Renaissance art history was due to Vasari whom she considered to have put more focus on Florence and its artists than on other Italian regions. She thus pointed out that such a social network view suggested a complete view on the relationships between persons relevant to art history, although this was not the case. Together with the finding that some of the portrait relationships discovered via the integration of the ULAN with the WGA data were not present in the ULAN, this feedback strongly suggested to do a more formal analysis of the Getty ULAN as source for social network data and to compare its content with other data sources featuring comparable information.

3.5 Quantitative analysis of portrait relationships

As a first step towards analyzing the ULAN's content, a closer look was given to the relationships between creators of portraits and the persons depicted therein. Identified via finding ULAN names in WGA artwork titles, such relationships were introduced as additional prototype feature in section 3.3.3. As already shown in this section, not all of the painter-subject relationships were reflected in form of ULAN relationships and it was thus of interest to get a more complete overview on their coverage there.

Figure 3.8 shows the complete identified network of 436 portraits based on creator-subject relationships connecting 471 different persons with each other, 253 creators and 245 subjects, 27 of which acting both as creators and as depicted persons. Each relationship is annotated with depictions of the respective artworks. A detail view shown in Figure 3.9 includes the titles of the artworks and focuses on a densely connected region of the network around Rubens and Titian. Thick arrows pointing from an artwork to a person node represent that the person's name is mentioned in the artwork's title, while thin arrows represent the relation between the artwork and its creator.

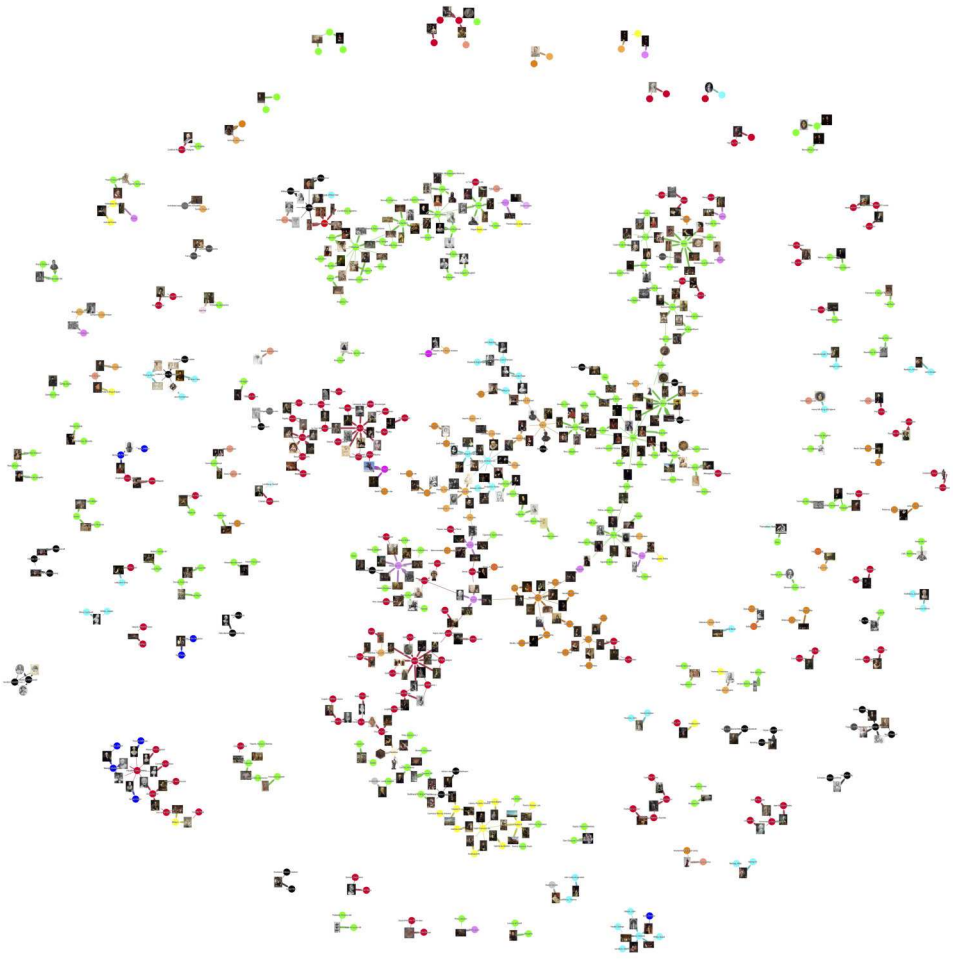


Figure 3.8: Network of creators and subjects of WGA portraits (**A2**)

Figure 3.9 shows examples for persons acting both as creator and as subject of portraits, such as the famous painters Leonardo (bottom right), Titian (right from the center) and Rubens (left from the center). Being creators of a number of portraits of historical figures and contemporaries, they were themselves again the subject of artworks made by other artists. One aspect only rarely covered, however, were self-portraits, which were found to rarely feature the artist name in their title. They were thus only included in the rare occasions when the latter was the case.

The analysis of the creator-subject-network revealed a number of interesting features, such as the presence of a large contiguous component (giant connected component, GCC) consisting of 201 (46.1%) portraits connecting 173 (36.73%) persons, 99 creators and 89 subjects, 15 of which again acting both as creator and as subject. Only three additional components were found to feature more than 10 portraits each, together these four largest

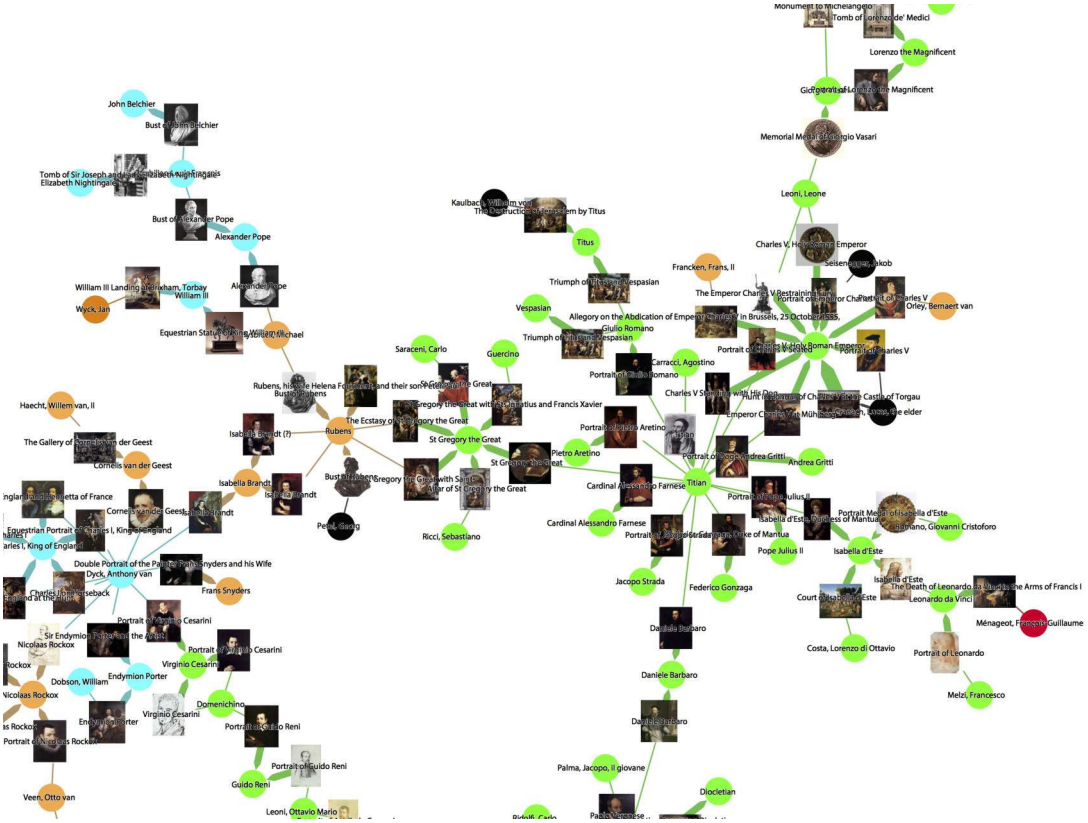


Figure 3.9: Network of creators and subjects of WGA portraits, detail view

components already included about two thirds of all portrait relationships while, as also clearly visible in Figure 3.8, the majority of components consist of one or two portraits with their respective creators and subjects.

The giant component featured many famous artists and historical figures across different art historical epochs and schools, including, amongst others, Italian (green person nodes), Dutch (orange), French (red), Spanish (yellow), British (light blue). Overall, a clear tendency towards portrait relationships involving persons of similar nationality became evident, visible as relatively clear-cut national clusters. Famous historical figures such as Greek (pink) philosophers or medieval church dignitaries, however, were often depicted by artists from different epochs and schools and thus served as constituting links between these different clusters in the contiguous components of the network. Other links between persons of different nationality reflected individual artists shifting their place of residence, such as for example Flemish born Anthony van Dyck on the left side of Figure 3.9 whose main nationality is listed as British in the ULAN. Due to the course of his career which started in Antwerp, included several years of residency in various

Italian cities and led him to become knighted court painter in London[20], Van Dyck is shown as the creator of portraits of Flemish, British and Italian persons.

Providing a more quantitative view on the birth date differences between creator and subject, Figure 3.10 shows that the majority of portrait relationships derived via the ULAN and the WGA involved contemporaries, i.e. artists depicting persons they met in person or were at least living during their own time. As visible in the chart, however, birth date differences could exceed well beyond 2,500 years, such as the 2,780 years between the birth of Ingres and the assumed birth of Homer.

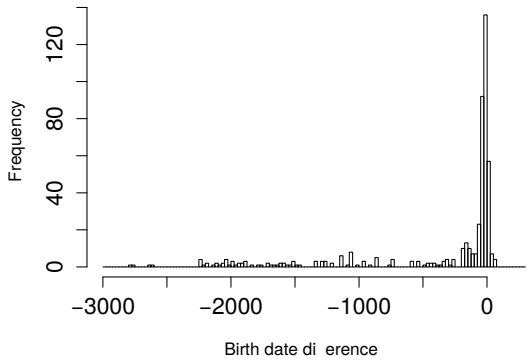


Figure 3.10: Birth date differences between creators and subjects of WGA portraits

Considering the most portrayed persons and the most prolific portraitists, Table 3.1 shows the respective rankings for subjects and creators having at least four associated portraits. The most portrayed person was Holy Roman Emperor Charles V who, according to Hackenbroch in [Hackenbroch, 1969], used depictions of himself as means to symbolize his power and omnipresence across the huge Spanish Empire. Ranks two and four were occupied by the French kings Louis XIII and XV, who most likely used portraits for similar reasons. Overall, about half of the top-ranked subjects in Table 3.1 were important rulers and other aristocratic patrons of the modern era, while the remaining frequently depicted persons there included the famous artists Michelangelo and Rembrandt and the famous writers Dante, Homer and Goethe, as well as famous historical figures from early church history and antiquity, with Daedalus appearing as frequently depicted fictional character from Greek mythology. As far as the creators are concerned, Titian appeared as the artist having created the most portraits, closely followed by Raphael and Anthony van Dyck. The frequently depicted painter Rembrandt appeared at high rank as well while, interestingly, the often portrayed Michelangelo was not found amongst the most prolific portraitists.

[20] https://www.britannica.com/biography/Anthony-Van-Dyck, retrieved Sept. 8th, 2020

Table 3.1: Ranking of subjects and creators by number of portraits

	subject		creator
4	Francis I		
4	Charlemange		
4	George Washington		
4	Urban VIII		
4	Homer		
4	Isabella d'Este		
4	Johann Wolfgang von Goethe	4	Canova, Antonio
4	Michelangelo	4	Ingres, Jean-Auguste-Dominique
4	Nicolaas Rockox	4	Mengs, Anton Raphael
4	Rembrandt	4	Pourbus, Frans, the younger
4	Scipione Borghese	4	Puget, Pierre
5	Charles I, King of England	5	Boucher, François
5	Charles III	5	Hals, Frans
5	Queen Victoria	5	Holbein, Hans, the younger
5	Daedalus	5	Rubens, Peter Paul
6	Marquise de Pompadour	6	Bronzino, Agnolo
6	Constantine	6	Veronese, Paolo
6	Saint Jerome	7	Goya, Francisco de
6	Socrates	7	Dürer, Albrecht
6	St Gregory the Great	8	Houdon, Jean-Antoine
8	Alexander the Great	9	Rembrandt van Rijn
9	Louis XV	10	Bernini, Gian Lorenzo
10	Dante	10	Dyck, Anthony van
10	Louis XIII	11	Raphael
11	Charles V, Holy Roman Emperor	13	Titian

Although the most portrayed individuals belonged to aristocracy and clergy, ranking the ULAN roles of depicted persons in Table 3.2 revealed that more than a third of all identified portraits were depictions of artists. The exceptionally frequent occurrence of the role "artist" in the ranking coincided with the observation that in the ULAN, artists were often assigned with the broad artist preferred label instead of more detailed descriptions such as painter or sculptor, while non-artists were usually assigned with more detailed role descriptions. One reason for this could be that many artists across history frequently used different media for their art.

Table 3.2: Ranking of subject roles by number of portraits

	subject role						
3	collector	5	cardinal	9	writer(s)	23	king
3	scholar	5	gentlewoman	11	poet(s)	25	pope
4	humanist	5	merchant	11	statesman	35	emperor
4	president	6	architect	17	noble	59	patron
4	public officer	6	queen	17	philosopher	156	artist

The 436 identified WGA portraits connected 377 unique pairs of persons, a number of persons were thus more than once portrayed by one and the same person. Table 3.3 provides the frequency rankings of subject-creator relationships for those occurring at least three times, with the exception of Rembrand discussed further below. The ranking especially highlighted individual artist-patron ties that had occurred throughout history, but also revealed other aspects of art historical relevance, such as Sandrart's portraits of Matthias Grünewald created for Grünewald's artist biography or Tintoretto portraying his Venetian peer Sansovino. As one rare example for artist self-portraits, Rembrandt was featured in the ranking as creator of two such works.

Table 3.3: Ranking of subject/creator pairs by number of portraits

	subject	creator
2	Rembrandt	Rembrandt van Rijn
3	Charles V, Holy Roman Emperor	Cranach, Lucas, the elder
3	Charles V, Holy Roman Emperor	Titian
3	George Washington	Houdon, Jean-Antoine
3	Jacopo Sansovino	Tintoretto, Jacopo
3	Louis XIII	Philippe de Champaigne
3	Thomas More	Holbein, Hans, the younger
3	Matthias Grünewald	Sandrart, Joachim von, I
3	Pope Benedict XIII	Bracci, Pietro, I
3	Pope Leo X	Raphael
3	Pope Urban VIII	Bernini, Gian Lorenzo
3	Scipione Borghese	Bernini, Gian Lorenzo
4	Marquise de Pompadour	Boucher, François
4	Charles I, King of England	Dyck, Anthony van

Last but not least, due to the observation described in Section 3.3.3 that the WGA portrait relationship between Isabella d'Este and Leonardo da Vinci was not explicitly present in the ULAN, it was of interest to quantitatively assess how many of the identified portrait relationships in WGA artworks were also explicitly recorded as ULAN person relationships. Since the ULAN was used to identify the person names in the portrait titles and also mapped to the WGA creators, each creator-subject pair could be looked up in the list of recorded ULAN relationships for that purpose. This way, less than

expected, only 57 (13.07%) of the 436 identified portrait relationships could also be found encoded in the ULAN, connecting 38 (10.08%) out of 377 unique pairs of persons.

The differing coverages suggested a tendency that those creator-subject pairs for whom multiple portraits could be found were more likely to be explicitly represented in form of ULAN relationships as well, which could be confirmed via a more detailed analysis. Table 3.4 shows the respective histograms of counts for the complete WGA portrait relationships and those found to be covered in the ULAN. While the two creator/subject relationships each having the most identified portraits (four) were both covered via ULAN associative relationships, the relative coverage gradually decreased with decreasing number of portraits per creator/subject pair.

Table 3.4: Recurring WGA portrait creator/subject pairs covered in ULAN relationships

Num. occ.	4	3	2	1
Complete	2	11	31	333
Matched	2	4	5	27
M/C	1.00	0.36	0.16	0.08

A look at the types of the thirty-eight ULAN relationships matched to creator/subject pairs, shown in Table 3.5, revealed that nineteen of them were directly related to patronage, eleven to professional or teaching relationships between artists and eight to family ties. It is interesting but not surprising to note that all of the six relationships matched to creator/subject pairs with more than two, eight of the eleven ties having more than one portrait were all due to patronage.

Table 3.5: Ranking of matching ULAN relationships

	ULAN relationship
1	apprentice was
1	patron was
1	employee of
1	colleague of
1	collaborated with
1	worked with
1	assistant of
1	child of
2	parent of
2	spouse of
3	court artist was
3	sibling of
5	teacher of
15	patron of

3.6 Summary

Inspired by historical examples to visualize developments in art history, the iterative design and evaluation of design artifact **A1** (Figure 3.6), the data-driven 3D Information Landscape prototype combining explicit historical social networks with digitized art collection information revealed a number of interesting observations.

On the one hand, seen from a potential user perspective, test users considered the approach to provide an immersive environment to present artworks embedded in a spatialized map of connections between persons relevant to art history to be a novel alternative to traditional Web browsing, but highlighted issues in interaction and efficient navigation, especially once they tried to get more information on specific connections they discovered while browsing. The users' interest in the inclusion of non-artists such as patrons and their call for a 2D visualization of the relationships between historical actors nevertheless underlined that the environment triggered their interest in such kind of information they were previously not aware about, especially when considering remarks that it revealed additional aspects of things already known. Seen from an exploratory search point of view, this observation suggested that the approach targeted aspects of "exploratory browsing", while neglecting those of "focused search".

On the other hand, seen from the level of the displayed content, the presentation and discussion of the prototype to and with art history domain experts challenged to consider the nature of the underlying data sources themselves. This was further underlined by the analysis of the identified portrait relationships, which revealed quantifyable gaps in the ULAN but at the same time opened up an interesting field of study showing the potential of integrating complementary data sources. It is noteworthy that the idea to analyze the creator/subject relationships evolved out of the initial design consideration to find interesting features to highlight individuals in the scene, many of them non-artists who were usually not included in traditional means to present art history in the realm of the digital. This was first approached by including DBpedia as additional data source for portraits and led to the analysis of WGA artworks for potential connections in this regard, leading to design artifact **A2** (Figure 3.8). The interesting structural properties of the creator/subject relationships as a whole motivated to further explore the full set of ULAN associative relationships (Chapter 4) and to compare them with additional related data sources, for which the successful inclusion of DBpedia suggested the Wikipedia ecosystem as potentially valuable resource (Chapter 5).

CHAPTER 4

Analysis of the Getty Union List of Artist Names

Using the ULAN network data as "fabric" for constructing the Information Landscape for design artifact **A1** (Figure 3.6) revealed that the provided interperson relationships were quite dense for certain art historical periods, such as the Renaissance, while others, especially more contemporary ones, were rather sparse in comparison. As the additionally introduced person ties derived from portrait relationships showed in design artifact **A2** (Figure 3.8), however, even the more thoroughly covered regions of the ULAN network left out certain facts that only became available via other sources. These observations suggested to perform a more thorough analysis of the ULAN network as a whole, going beyond the individual ego-networks that were considered in the interactive prototype.

A search for existing work in this regard did not return any studies dedicated to such an analysis of the network information present in the ULAN. Amongst its mentioned application scenarios described in [Schreiber et al., 2008], [Wang et al., 2009] and [Kurki and Hyvönen, 2009], only little attention had been directed at its overall network structure. Concerning these three studies, only [Kurki and Hyvönen, 2009] briefly reported the similar observation as made with artifact **A1** that the ULAN network was found to contain a strongly connected component of about 12,000 persons which featured the most important representatives of the arts, but rather neglected contemporary artists which were found to have rather sparse relationships. No further analysis, however, was reported by the authors in this regard.

A study by Clough et al. in [Clough et al., 2009] discussed some aspects of the ULAN in the context of the semantic annotation of information about artists for enhancing the search in the online database of Britain's national Tate gallery, Tate Online. Although, for the search scenario, artist relationships were found to be of high importance to users via a survey, the authors - without providing more detail - considered this type of information to be too sparsely represented in the ULAN to be studied in their work and therefore limited their focus on individual artist properties such as birth and death date and place,

nationality, role and gender. The authors found 73% of the 3,000 artists from the Tate Online database to be present in the ULAN and their analysis of agreement regarding the mentioned metadata elements suggested the presence of an editorial bias there which not always reflected "the nuances of the domain information" and also a partial incompleteness "both in terms of its coverage of artists and the properties associated with them", which the authors also attributed to some form of cultural bias of its editors.

Together with the findings made in the context of artifacts **A1** and **A2**, the overview on existing studies covering related aspects of the ULAN further motivated to perform an in-depth analysis of its content with a dedicated focus on its network data, which is described in this Chapter. Considering the ULAN network as a whole led to a visualization of its Giant Connected Component, design artifact **A3**, which provided interesting insight into the large-scale art history narrative embedded there.

4.1 Data retrieval & preparation

The ULAN data was retrieved via the Linked Data representation of the Getty Vocabularies, available via a dedicated Website matinained by the Getty Foundation[1]. The full triple dump[2] of the ULAN was downloaded and imported into a local OpenLink Virtuoso[3] triplestore.

ULAN person records feature different types of attributes, including person names, birth and death date and place, gender, nationality and role. Multiple values can be assigned to names, nationalities and roles, these fields exist as preferred and non-preferred versions. Preferred terms are unique and mandatory, while non-preferred terms can be multiply assigned. For this study, gender, birth and death date as well as the preferred attributes for the person names, roles and nationalities were considered. The respective subsets of the ULAN person records were imported into R via Virtuoso's SPARQL endpoint, resulting in a data table consisting of 193,823 rows with the attributes shown in Table 4.1

Id
Preferred name
Preferred role
Birth Date
Death Date
Gender
Preferred nationality

Table 4.1: Observed ULAN attributes

[1] `http://vocab.getty.edu/`, retrieved Sept., 8th, 2020
[2] `http://vocab.getty.edu/dataset/ulan/full.zip`,
used version: 3.1 as of 5 June 2015
current version: 3.4 (`http://vocab.getty.edu/doc/`, retrieved Sept., 8th, 2020)
[3] `https://virtuoso.openlinksw.com/`, retrieved Sept., 8th, 2020

Minor corrections had to be applied to the data with respect to different language versions, since the respective attribute values for person names, gender, role and nationality were sometimes present in multiple languages. For the latter three, the values could be straight-forwardly filtered to English representations by only considering those having the appropriate "@en" or "@en-us" language tags attached, since these attributes were appropriately tagged in all person records. Regarding person names, some records did not have labels with English language tags. In this case, the following strategy was applied: For those records having only one language version for the name, this language was used, while for those with names in multiple languages but not english, the language version with the overall next highest presence was chosen.

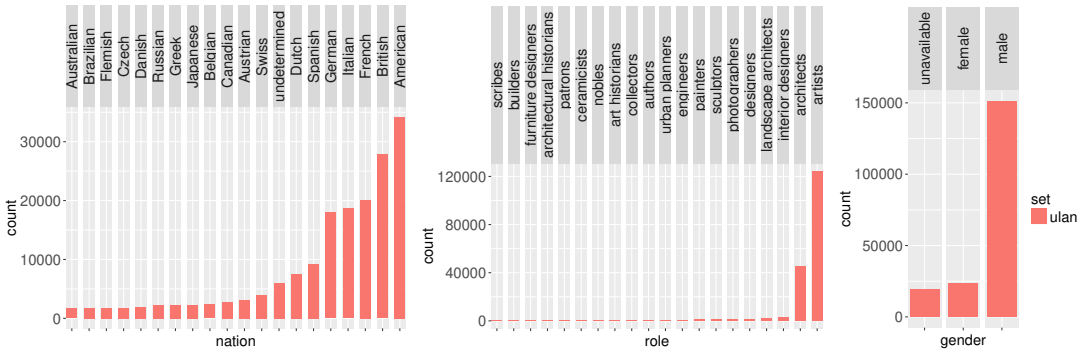
Relationships between person records were retrieved via a separate query, resulting in a second data table consisting of 51,588 rows of the form

<source, target, relationship_type>

mutually connecting 21,942 of the 193,823 person records. The 51,588 retrieved ties represented 50,076 unique relationships, 1,460 of which had 1,512 duplicates, i.e. two persons sometimes had more than one type of relationship, e.g. parenthood and teaching, each represented by a separately typed tie.

4.2 Attributes of ULAN biographies

Figure 4.1 shows the distribution of the top 20 preferred nationalities and roles and the different values for gender across all ULAN person records. As visible in subfigure (a), the top-20 of the 303 preferred nationalities featured in total were found to cover 87.4% (169,360 of 193,823) of all person records, almost a third (34,205 Americans and 27,853 Britons of 193,823) from America and Great Britain alone, the top-5 nationalities consisting of the latter two and French, Italian and German persons represented 61.31% (118,824) of all ULAN person records. While the increased presence of English speaking persons could in general also be attributed to the US origin of the data source, the overall distribution of person nationalities was clearly in line with post-colonialist statements regarding the Western-centric view of art history. A skewed distribution of attribute values became even more evident with preferred person roles shown in subfigure (b), where already the largest two out of 503 in total, artist and architect, were found to be assigned to 87.5% (169,591 of 193,823) of the person records. The top-20 roles together covered 94.99 % (184,107 of 193,823) of the records, meaning that the 483 remaining roles were used as preferred role for only about 5% of the featured persons. Subfigure (c) shows that about 12% of all ULAN records were found to describe Females and 78% Males, the remaining 10% undetermined. This in turn provided clear evidence for Feminist art historians' claims for the under-representation of Female artists in art history.



(a) (b) (c)

Figure 4.1: Distributions of attributes assigned to ULAN person records

Figure 4.2 shows a density plot of the distribution of birth dates, revealing that the majority of persons covered in the ULAN was of relatively recent birth, featuring the same gap as Priestly's chart of biography in its artist section: A few persons from Greek antiquity were followed by a long gap lasting until about 1000AD, from when on the number of covered persons increasingly grew until it decayed again during the 20$^{\text{th}}$ century. As it was also observed by Clough et al. in [Clough et al., 2009], the distributions of birth dates were found to feature many spikes mainly centered around birth years which were multiples of 10, suggesting approximations to not exactly known dates. The birth year 1850 was found to represent a very significant peak in this regard, assigned to 30,000 persons, five times more than the second most often assigned birth year. This spike aside, the year 1920 appeared to be the turning point from which the number of more recent birth dates dramatically declined.

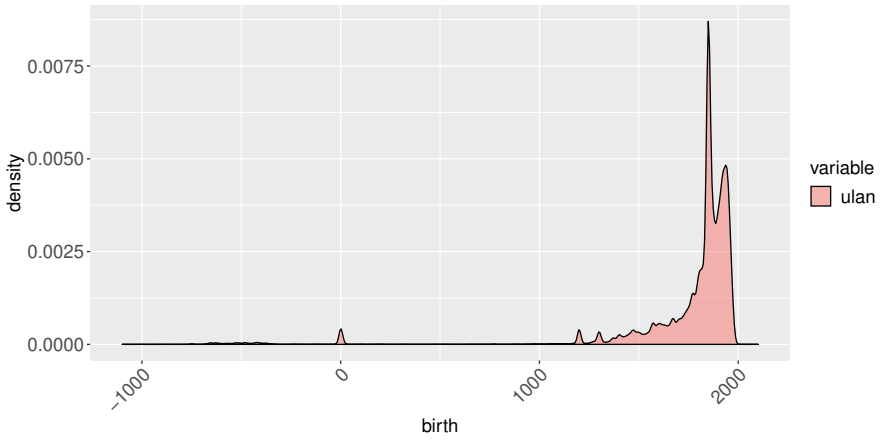


Figure 4.2: Distribution of birth years assigned to ULAN person records

Interestingly, the vast majority of the records assigned with the birth date 1850 were found to have death dates set to a future date, prompting for an explanation. As the consultation of the Getty data entry guidelines from [Getty Research Institute, 2015] (Section 4.2) revealed, these cases represented possible date ranges assigned when only unclear information was available for "contemporary" persons active during more recent times. Together with similar patterns found for persons tagged as "Active 20th century", "unidentified", etc., this provided reasonable explanation also for the other minor spikes which became visible in 4.2 besides the one around 1850. Table 4.2 provides a brief overview on the most prominent patterns encountered in this regard.

Count	Birth	Death	Display
28,888	1850	> 2015	"Contemporary"
2,145	1871	> 2015	"Active 20th century"
1,921	1	2090	"unidentified"
1,614	1771	1930	"Active 19th century"
1,539	1200	> 2015	"unidentified"

Table 4.2: Five most present patterns of unusual birth/death dates in ULAN records

Given the low overall presence of women in the ULAN, their proportion across the ages was another interesting aspect of the distribution of birth dates. Providing a more detailed look in this regard, Figure 4.3 reveals that the proportion of women amongst the ULAN persons remained very low until about the onset of the 19th century. From then on, a steep increase raised their fraction to about 32% in the year 2000. This observation suggested a fundamental shift in the arts and in art history towards a more balanced view on the domain, most likely driven by recent academic initiatives to uncover long neglected constributions by women, but also by the increasing presence of recognized and successful female artists in the art world.

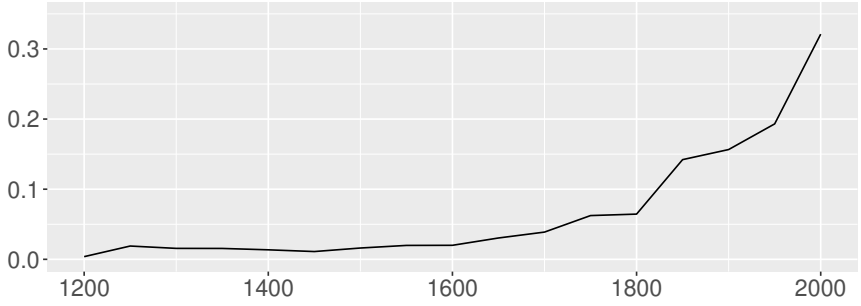


Figure 4.3: Percentage of ULAN persons having gender "female" across time

The distribution of nationalities across time revealed some interesting aspects as well, shown in Figure 4.4 for the reduced dataset with cases from Table 4.2 removed. Since

there are not enough distinguishable colors to visually tell apart all different attribute values, only a subset of values could be colored simultaneously. According to Paul Green-Armytage in [Green-Armytage, 2010], the upper limit of the number of distinguishable colors is about 26, similar to the number of letters in the alphabet. The 26 color scheme proposed by Green-Armytage was therefore used to color the top 26 counting nationalities and a light gray value added as 27th color for all the persons of "other" nationality.

As visible in the Figure upon closer inspection, the distribution of nationalities reflected major developments in art history. After the already mentioned gap between ancient Greek artists and the high middle ages, the leading role of Italian artists during the Renaissance became clearly visible, followed by the Dutch Golden Age and then by gradually increasing proportions of the other highly ranked nationalities featured in Figure 4.1 (a). Besides representing the beginning of a dramatic increase in covered persons, the time around 1800 was also followed by an increasing variety of nationalities, especially visible in the growing fraction of the "other" section, which supported Anderson's statement in [Anderson, 2005] that the social construct "nation" only started to emerge at around the end of the 18th century. As far as the distribution of roles was concerned, the large proportion of artists and architects dominated across the ages, with a similarly increased variety of assigned roles for persons born after 1800.

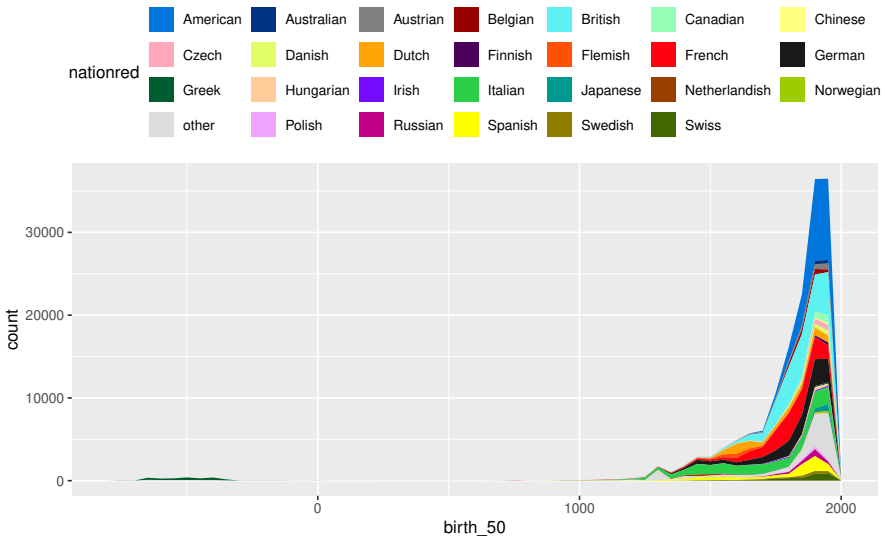


Figure 4.4: Distribution of ULAN nationalites across time

The gap between antiquity and the early middle ages visible in Figure 4.4 revealed a strong similarity between the distribution of ULAN records over time and the artist's portion of Priestley's Chart of Biography presented in Section 2.1.2. Figure 4.5 shows a juxtaposition of a jittered plot of all the ULAN person records assigned with the preferred role "artist" by birth date with this part of Priestley's timeline. Both displays featured an almost similar gap stretching across centuries and the dramatic increase of notable

artists observed by Priestly for his own time in [Priestley, 1765] was continued towards the present in the temporal distribution of ULAN artists.

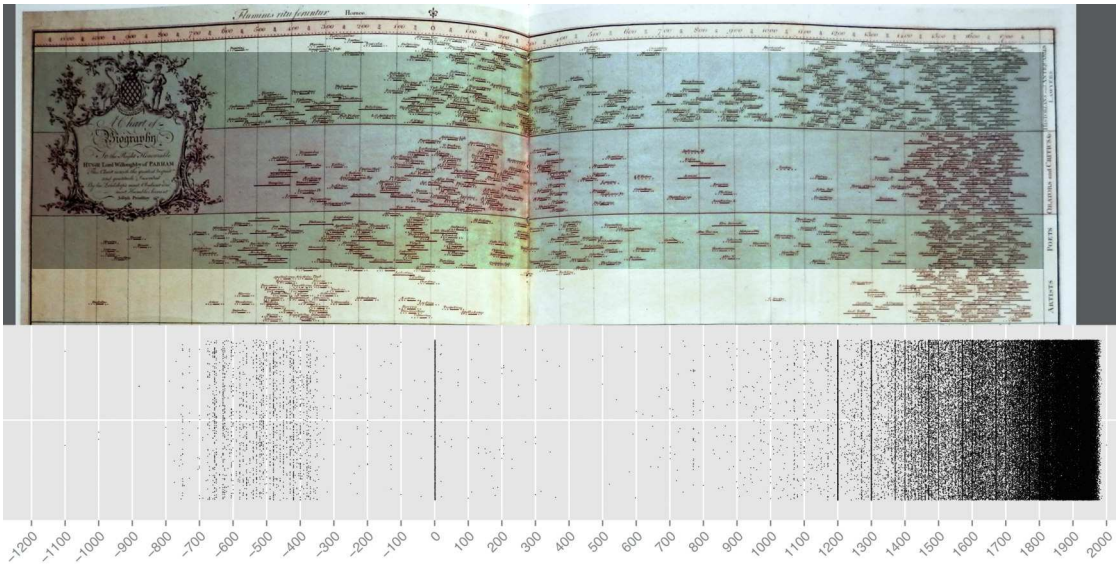


Figure 4.5: ULAN artist birth dates against time, compared to Priestley's chart

4.3 Analysis of the ULAN network

The distributions of ULAN person attributes provided interesting insight into the overall content of the data collection. As far as the network of mutually interlinked ULAN person records was concerned, however, only a subset of the ULAN records were connected via associative links. About 11.3% (21,942 of the 193,823) person records were found to feature at least one of 51,588 associative links to another person, which raised the question if the distribution of person attributes was different in this much smaller subset. This question was based on the assumption that persons for which associative links were recorded might represent more important representatives of art history.

Comparison of linked and unlinked ULAN person records

The ULAN person records were thus split in two groups, one consisting of the 21,942 linked person records (linked), the other one of the 171,881 records without such links (unlinked). Figure 4.6 shows the distributions of the observed person attributes in the two groups side by side, suggesting the attribute distributions of the linked and the unlinked subsets to be relatively similar in some aspects, but also different in others.

The same top nations appeared to dominate both nationality distributions, although with a clearly shifted count rank order. As far as roles were concerned, the two main groups artist and architect also remained the most present ones in the linked subset, but it appeared to feature an even higher fraction of artists compared to the unlinked subset. For both nationality and role, the linked subset was found to feature clearly less distinct values than the unlinked one. Of the overall 303 different values for nationality and 503 for role, only 143 and 208 were found to be present in the linked subset and in both cases these "remaining" values also had a higher count rank in the overall set: While the mean/median global count rank of nationality values in the linked subset was 90/75, the mean/median global count rank of nationality values not present there was 207 and 216, respectively. The same was found for roles, 161/117 vs 316/326.

The overall distribution of gender still strongly favored males in the linked subsets, but it appeared to contain much fewer persons with unavailable gender information (1.3% vs. 11.2%), which was also the case for undetermined nationality (0.08% vs 3.5%). This also became visible in the distribution of birth dates, where the linked dataset was found to contain only a fraction of persons assigned with the broad lifespan information such as shown in Table 4.2, linked persons thus featured much more accurate birth/death data. Another interesting difference became evident regarding the distribution of birth dates in the two different sets, shown in their overlaid density plots in Figure 4.6 (d). The linked set was found to have a visibly increased birth date density for the period from the 15th until the 18th century, suggesting that it featured a higher proportion of especially Renaissance and Baroque artists and other persons. This also became evident in the mean and median birth years, 1728AD vs 1768AD and 1809AD vs 1850AD, respectively.

In order to test the differences between the subsets in a more formal way, Chi-square tests of homogeneity[4] were applied to the categorical variables nationality, role,

[4] https://stattrek.com/chi-square-test/homogeneity.aspx, retrieved Sept., 8th, 2020

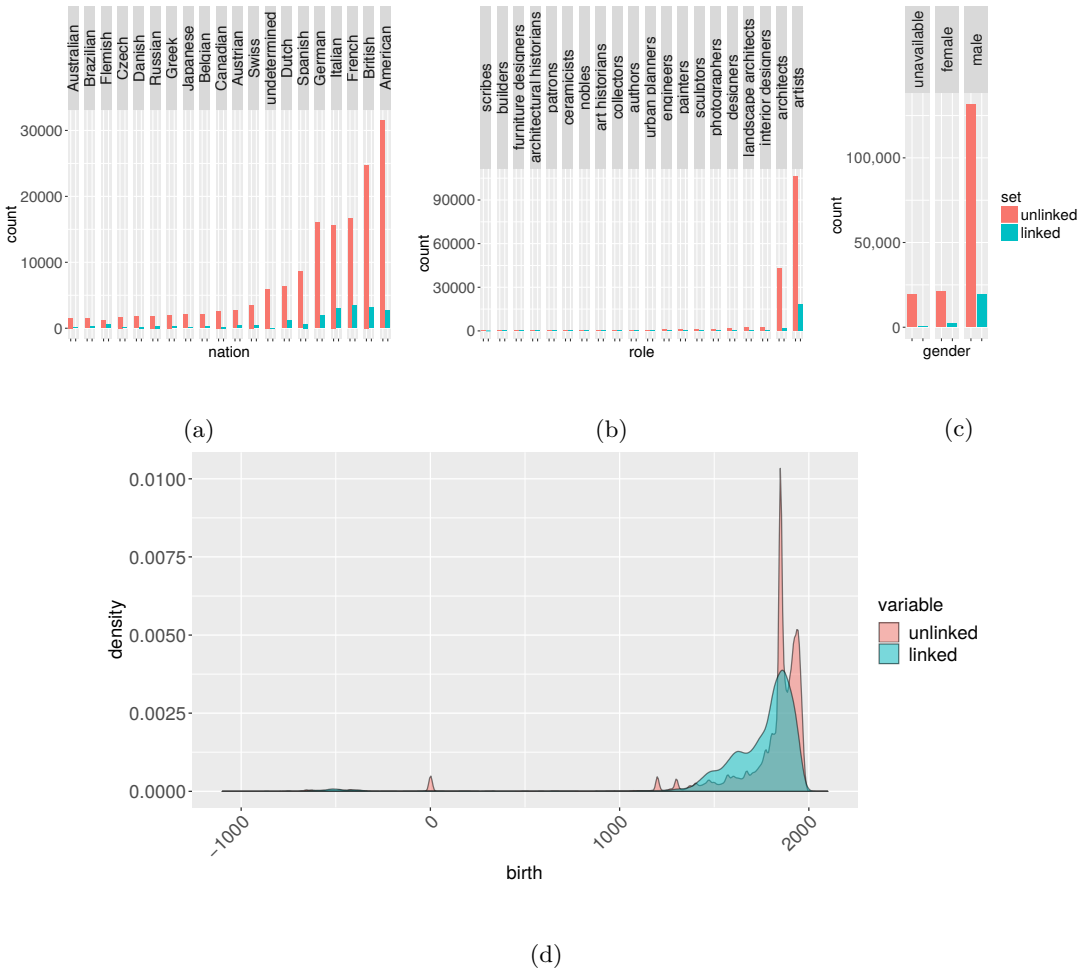


Figure 4.6: Distributions of attributes in linked and unlinked ULAN person records

gender and birth dates binned to 25 year periods, while a Kolmogorov-Smirnov test was additionaly used to compare the distributions of "raw" birth dates. Chi-square tests of homogeneity are determined to measure if two different and non-overlapping count distributions for the same categorical variable, placed column wise in a contingency table, tend to be similar or significantly different from each other, based on the deviations — the residuals — of the individual value counts from an expected value which is derived from the marginal distributions of the contingency table. The two sample version of the Kolmogorov-Smirnov test in turn provides means for comparing two continuous distributions for their similarity by comparing their empirical cumulative distribution functions (ecdf) and deriving the result from the largest difference between the two.

One general prerequisite of Chi-Square tests is the requirement that small counts in the contingency table should be avoided, a general rule of thumb being that no cell (=attribute value) should have an expected count less than five. Otherwise, small counting

attribute values could result in very high residual values, suggesting high importance not supported by the underlying data and leading to unreliable results. In order to meet these requirements, attribute values having expected counts less than five in at least one subgroup were thus combined into an "other" category. This reduction affected only about 1-2% of all records, as shown in Table 4.3. The three available values for gender did not have to be collapsed.

	303 Nationality values		**503 Role values**		**129 birth date values rounded to 25**	
	83 original	220 combined into "other"	84 original	419 combined into "other"	52 original	77 combined into "other"
Unlinked	170,214 (99%)	1,664 (1%)	168,761 (98%)	3117 (2%)	171,100 (99.5%)	778 (0.5%)
Linked	21,720 (99%)	225 (1%)	21,615 (98.5%)	330 (1.5%)	21,839 (99.5%)	106 (0.5%)

Table 4.3: Number of person records affected by attribute value combination

For the three categorical attributes nationality, role and gender, as well as the binned birth dates, the tests showed significant differences between the distributions of the counts of their respective values (p-value for observed distributions under null hypothesis "same distribution" « 0.01). The same was found to be the case for the KS-test comparison of the birth date distributions in linked and unlinked ULAN person records (p-value for observed distributions under null hypothesis "same distribution" « 0.01). Representing an overall measure for (dis-)similarity, the obtained results of course did not provide insights on the individual contributions to the differences between the distributions, i.e. which individual value counts for the observed person attributes were rather similarly represented in the linked and the unlinked datasets and which were not.

In order to get a better overview in this regard, it was desirable to find a way to observe the individual contributions to the differing distributions in more detail and, as far as the Chi-Square test based comparisons were concerned, to highlight the relationship between the observed values and the expected values in this regard. One approach was to visually compare the quantitative rank orderings of the individual values for nationality, role, gender and birth date between the subsets using 2D scatterplots, where each attribute value was represented as a point in 2D space positioned on the two axes based on its count in each subset. If all attribute values were ranked similarly in both subsets, their positions in the plot would show a monotonously increasing relationship and if not, they would appear to be more erratically placed across the canvas.

The Figures 4.7 show such visualizations for each person attribute, where individual attribute values were placed along the x and y axes by their frequency of occurrence in the unlinked (x-axis) and linked (y-axis) subsets. Due to the skewed distributions of value counts, logarithmic scales were used for both axes. The straight red lines represent the hypothetical positioning of the attribute values if their observed counts in both subsets would have matched their expected counts calculated from the contingency table set up for the Chi-Square test as described by Agresti in [Agresti, 2007, p. 37]. This way, attribute values positioned further away from the diagonal indicate that their observed counts in one or both of the subsets deviated from what would have been expected based

on their overall fractions in the two subsets. In order to particularly highlight strong deviations, the disc symbols representing the individual attribute values were scaled and colored by the value of their standardized Pearson residuals (sometimes also referred to as adjusted residual) as again derived from the contingency table for the Chi-square test, calculated as described by Agresti in [Agresti, 2007, p. 38]. The latter enabled to assess the deviation of observed counts from their expectation via units of standard deviations independent of the underlying absolute counts and thus, as outlined by Sharpe in [Sharpe, 2015], allowed to identify specific attribute values which more prominently contributed to the overall difference between the compared subsets for each person attribute. The size of each disc reflects the absolute residual value, while its color (green: linked subset, blue: unlinked subset) highlights the subset where the attribute value was found to be more prominently present, i.e. whose residual had positive sign. Light colors show absolute standardized residual values below two while those above that value, reflecting more significant deviations from expectation, are colored boldly.

Subfigure 4.7 (a) shows the rank scatterplot for nationalities. The standardized residuals from the Chi-Square test corresponded well with the rank differences derived from simple counting, nationality values with clearly higher rank in the interlinked dataset tended to have significantly positive residuals, while the same was the case for values ranking higher in the unlinked dataset. With growing rank, however, the differences in ranking decreased but the residuals still gave an indication about which dataset featured a nationality more prominently. The plot shows that the linked dataset seemed to especially put more emphasis on European nationalities considered "classical" players in the arts, such as French, Italian, Dutch and Flemish, but interestingly also featured specific groups usually considered to be underrepresented, such as Native American persons, in a higher proportion than in the unlinked dataset.

The scatterplot for the roles, shown in Subfigure 4.7 (b), revealed that linked person records especially featured a higher proportion of patrons of the arts, such as important persons from nobility and clergy. While persons of role "patron" still had higher absolute count in the unlinked dataset, related persons with more detailed roles such as popes, emperors, queens and kings were even found to have higher absolute counts in the linked dataset, i.e. even though the linked dataset contained only about one tenth of the amount of persons present in the unlinked dataset, it nevertheless featured more than 50% of the popes, kings, queens and emperors present in the ULAN overall. This suggested that such persons played important roles in historical social networks of art history.

Although the ranks for gender, visible in Subfigure 4.7 (c), were found to correspond between both datasets, the residual suggested a significant focus on male persons in the linked dataset. Regarding birth dates rounded to quarter-centuries, Subfigure 4.7 (d) confirmed what's already visible in Figure 4.6 (d): Greek antiquity and the time between early Renaissance and early 19th century — with the exception of the period around 1850 — were found to be a strong focus of the linked dataset, while persons from the late middle ages and the 20th century were less prominent there.

The comparison of the person attribute distributions in the subset of ULAN records having associative links and those in the one without suggested that the linked subset

tended to contain more person records associated with "classical" periods in art history, born in either Greek antiquity or in the modern age between the Renaissance and the onset of the Industrial Revolution. Persons having French, Italian, Dutch and ancient Greek nationalities, but interestingly also unexpected and usually "neglected" groups such as Latin and Native American people, therefore had higher relative presence in the linked dataset than for example Americans, British and, interestingly, also Spanish persons, who rather were expected to be more significantly represented. Moreover, the fraction of artists in the linked subset was higher than the one of architects and as far non-artist roles were concerned, it contained higher proportions of members of aristocracy or the clergy as well as other person groups usually being patrons or collectors of the arts. The even lower presence of Females observed in the linked subset could be explained by the lower presence of 20^{th} century persons there, who featured a clearly better gender balance compared to earlier days, as already shown in Figure 4.3.

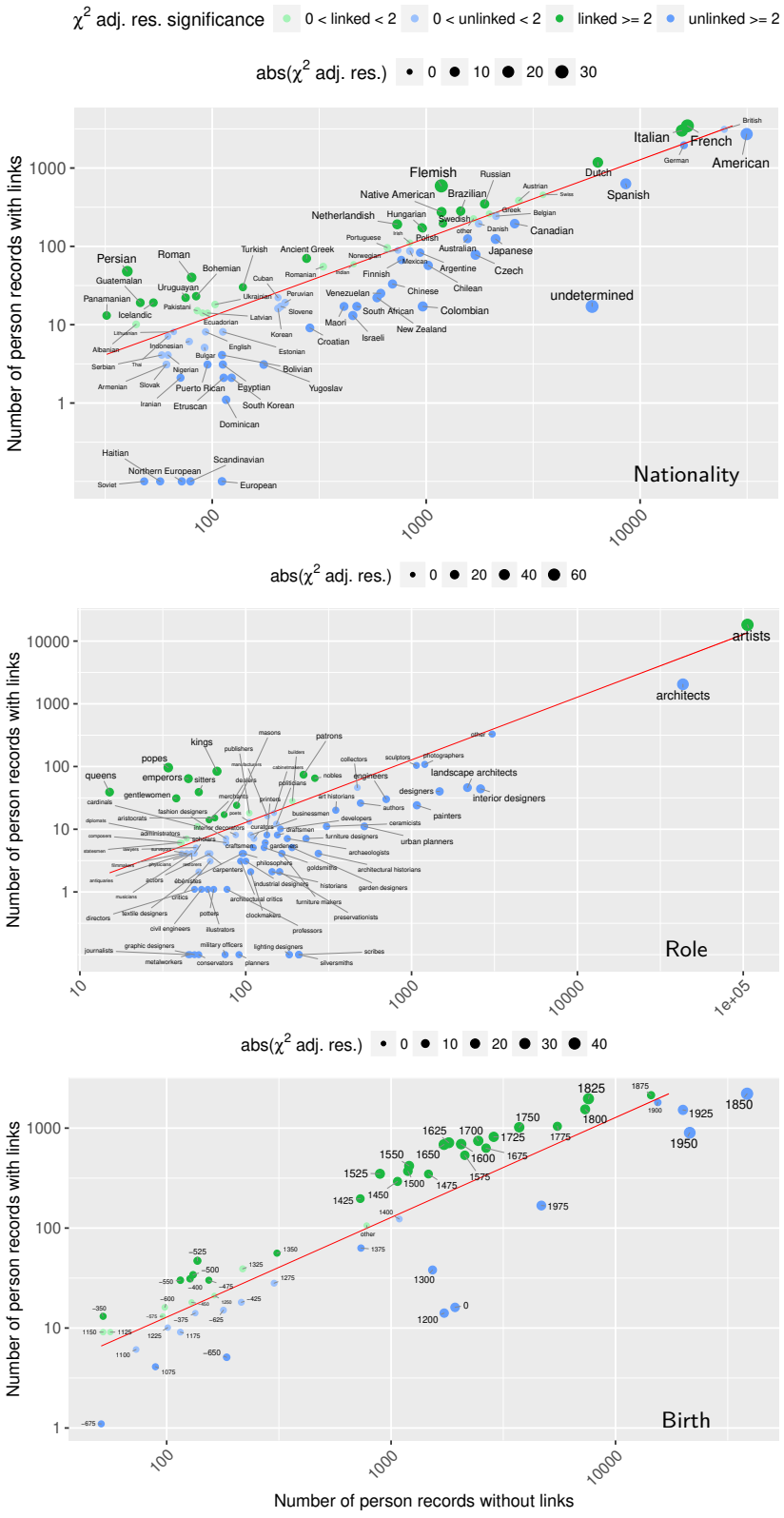


Figure 4.7: Attribute value frequencies in linked and unlinked ULAN record subsets

4.3.1 Network structure

Having gained initial insight on the differences between the subsets of person records with and without links to others, it was of particular interest to put more focus on the existing person connections. One main question was if the set of interlinked persons formed a contiguous network in which all members were connected via paths of varying length or if it rather contained multiple, island like groups of interlinked persons. A straight-forward way to get a quick overview in this regard was to create a visual representation of the set of the 21,942 interlinked person records and their 51,588 mutual links, shown in Figure 4.8, colored by nationality using the same color-code as used for Figure 4.4.

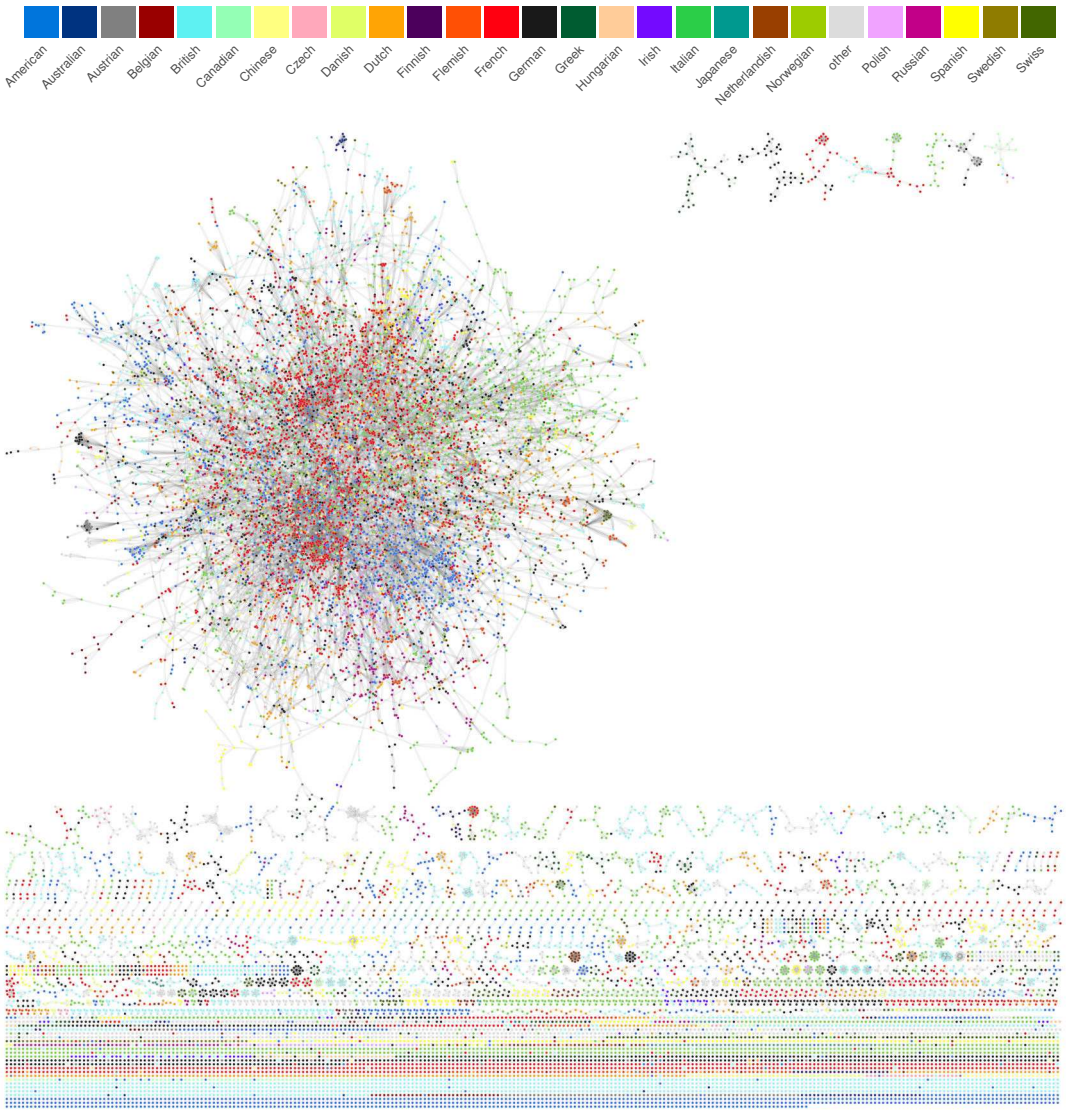


Figure 4.8: Different-sized components of the ULAN person-to-person network

As visible in Figure 4.8, the network was found not to be coherent but rather separated into 4,083 connected components, ordered by size in the Figure. One of the components, however, was very large compared to the rest, consisting of 10,444 (47.60%) of all linked persons. As stated by Easley and Kleinberg in [Easley and Kleinberg, 2010], such a so-called giant connected component (GCC) can often be found in complex networks. The relation of the size of the GCC to the sizes of the other components becomes visible in Table 4.4, showing the distribution of component sizes ordered by size from left to right. The second largest component, having only 39 members (0.18%), was already more than two orders of magnitude smaller than the GCC and the rest of the network was scattered across even smaller components.

size	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20	21	23	25	27	37	39	10444
count	2780	674	255	130	60	56	32	22	21	13	8	6	4	2	4	3	1	3	2	1	1	1	2	1	1

Table 4.4: Counts of connected components of size 2-10444

Homogeneous Network Components

Featuring nodes colored by nationality, Figure 4.8 shows that many of the smaller components of the network were quite homogeneous in this regard, which also turned out to be the case for the role (r) and gender (g) attributes: 87.84% (r: 85.83% g: 76.62%) of the components of size two, 76.8% (r: 81.01% g: 78.78%) of size three and 72,55% (r: 76.08% g: 75.69%) of size four where homogeneous for these attributes. Overall, 3,386 (82.93%) components composed of 8,839 persons (40.28%) were found to be homogeneous with respect to nationality, while a similar number but not exactly the same set of components, composed of 9,053 persons (41.26%), was homogeneous with respect to role, 3,091 (75.70%) components composed of 8,343 persons (38.02%) were homogeneous for gender. Notable observations regarding the nationality distribution were that the largest number of homogeneous components was composed of British persons and that Native Americans were especially often found in homogeneous components. In fact, of the 275 Native American person records with associative links, 210 were part of components consisting solely of such person records, which on the one hand highlighted the effort to include these usually marginalized groups of persons into the ULAN, but at the same time underscored their "isolation" from the remaining actors there. As far as gender was concerned, it was interesting to see that only the small fraction of about 2.62% of the gender homogenous components was all female, while 96.05% were all male. While this finding was to be expected given the overall low proportion of women in the ULAN and especially in records having associative links, it was even more interesting to see that while 68.51% of the all male components were of size two, this was the case for 91.36% of the all female components, suggesting that the proportion of recorded female professional networks in art history was even lower than the overall low fraction of women.

Assortativity

As it already became visible via the encountered high numbers of homogeneous components, there was a general tendency of ULAN links to connect persons having similar attributes, i.e. also within inhomogeneous components such as the GCC, there was a clearly higher proportion of links between persons having similar features. Easley and Kleinberg in [Easley and Kleinberg, 2010] called this network trait "homophily", also known as "assortative mixing" or simply "assortativity", the tendency of similar nodes to connect with each other. The notion of similarity between nodes can on the one hand be based on structural attributes such as node degree which provides a measure for the tendency of nodes having the same number of links to connect with each other, or on "external" attributes such as nationality, role and gender. A straight forward approach for assortativity is to measure the fraction of links between persons having the same attribute value. In case of the full ULAN network, the overall fraction of links between persons of same nationality was found to be 0.82, for the giant component it was 0.78. For roles, it was 0.88 and 0.90, for gender, 0.86 and 0.87: In all cases the fraction of links between records having the same attribute value was found to be much higher than for those between differing ones.

A similar observation also emerged for the birth attribute when considering the difference between the birth dates of interlinked persons. Figure 4.9 shows a density plot comparing the birth date differences for the full network and for the giant component. While there was a clear overall tendency towards links between contemporaries than between persons many generations apart, it was interesting to see that the full network featured a peak for links between persons of same age while the giant component instead featured two "shoulders" at about -22.5 and 22.5 years, reflecting a stronger presence of links between two succeeding generations which could be based on teacher-student or parent relationships. Of the 1,256 same age links in the full network, however, 684 were based on birth dates ending at full decades, suggesting that many of the identified "same-age" links were due to such rounded dates. Overall, Figure 4.9 suggested that the majority of links between persons in the ULAN were within a range of about three generations, which was confirmed by looking at the underlying distribution shown in Table 4.5, listing a selection of the respective quantile values. For both the full network and the giant component, about 70% of all links were in the range between -34 and 34 years, while 98% of all links were between -100 and 100 years in the full network and between -90 and 90 years in the giant component.

Interestingly, the observations regarding birth date differences correlated with the ULAN editorial guidelines about the recording of associative relationships from [Getty Research Institute, 2015, Section 3.5], stating that only direct relations between persons should be stored. This distinction between direct and indirect relationships suggested that it could be helpful to define a "rule of thumb" cutoff for quantitatively classifying inter-person links accordingly by using a simplified measure for birth date difference. Since direct ties were expected to only happen between living persons, it seemed feasible to define a suitable maximum birth date difference for such ties based on the typical lifespan of persons in the ULAN. The subset of interlinked ULAN records was thus used

to calculate the median lifespan between person birth and death dates, which was found to be 74 years. It was thus assumed that the typical birth date difference of directly interlinked persons was in the range 0-75 years, which was found to be the case for 95.55% of all associative relationships in the ULAN.

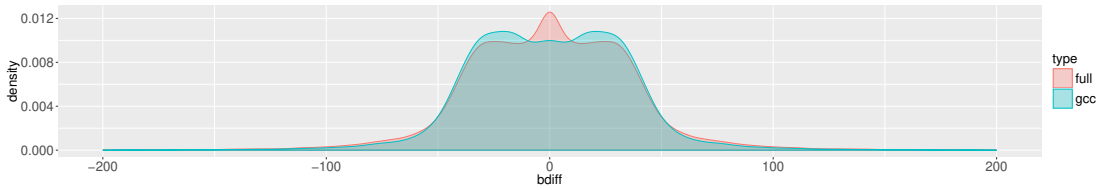


Figure 4.9: Density plot of birth date differences between linked ULAN person records

	1%	2%	5%	15%	50%	85%	95%	98%	99%
full	-99	-79	-54	-34	0	34	54	79	99
gcc	-89	-70	-49	-34	0	34	49	70	89

Table 4.5: Quantiles for birth date differences in ULAN links, full network and GCC

The observations that major fractions of links connected ULAN persons having similar attribute values, however, did not take the general distribution of attribute values into account. Comparing the observed counts of same-attribute links to expected counts based on a random model could provide more insight on the significance of the observations. A simple random model would be to assume links between person records to happen by chance, determined only by the fractions of attribute values in the population. 3,456 records of all 21,942 linked records for example had French nationality, a fraction of 0.1575062. Based on the assumption of independence, the expected fraction of links between French persons would be 0.1575062 * 0.1575062 = 0.0248082, or 2.48082%. The observed fraction of French-French links, however, was 0.1587187, or 15.87187%, about six times higher than expected and thus suggesting a strong focus on relationships between French persons. One measure that follows this concept is called the assortativity coefficient, defined by Mark Newman in [Newman, 2003b], consisting of a single numerical value between -1 and 1 expressing the assortativity of the network with respect to one specific attribute compared to the expected assortativity in a random network. A perfectly assortative network, i.e. only persons with similar attribute values are connected, has assortativity 1, a network with completely random connections has value 0 and a disassortative network values between 0 to -1. As described in [Newman, 2003b], there are different calculation methods for the assortativity coefficient, based on whether the observed attribute is discrete or continuous.

In the general ULAN network consisting of all components, interconnected nationalities were found to have the assortativity coefficient 0.80, while the giant component alone had the coefficient 0.75, clear indicators that persons of similar nationality were more likely to be connected to each other in the ULAN even when taking the distribution of attribute values there into account. The tendency of persons of similar role to connect with each

other was less pronounced with respect to the distribution of roles, having assortativity coefficients of 0.53 for the whole network and 0.37 for the giant component. This was due to the very high proportion of persons assigned with the role artist, which was even higher in the giant component, reducing the significance of the many links between persons tagged with that role. For gender, the coefficient of 0.16 for the full network and 0.07 for the GCC was even lower. Contrary to the first impression due to the high fraction of same gender links in the network, these values suggested that persons of same gender were not significantly more connected to each other when the underlying gender distribution, heavily leaning towards males, was taken into account. The assortativity coefficient for birth dates in contrast yielded almost perfectly assortative networks with respect to the tendency for contemporaries to be connected with each other, 0.97 for the full network and 0.96 for the giant component.

Link types

The 51,588 ULAN links constituting the different network components were assigned with 93 different types of relationships, 70 of them representing 35 pairs of mutual relationships such as teacher and student, parent and child, etc. All links were found to be usually assigned reciprocally: Whenever there was a link from a person to another, there was also a mirrored link in the other direction. Figure 4.10 compares two different count distributions of link types, one for those assigned to all 51,588 links, the other for the de-duplicated 50,076 unique links, 1,460 of which having multiple person relationships collapsed to an artificial type "multi". The comparison shows that the de-duplication had only minor effect on the general distribution and that teacher/student, sibling and parent/child relationships had the highest presence there. Especially for teacher/student and parent/child relationships, the similar counts for each link directions underscored the almost fully reciprocal presence of such mirrored links.

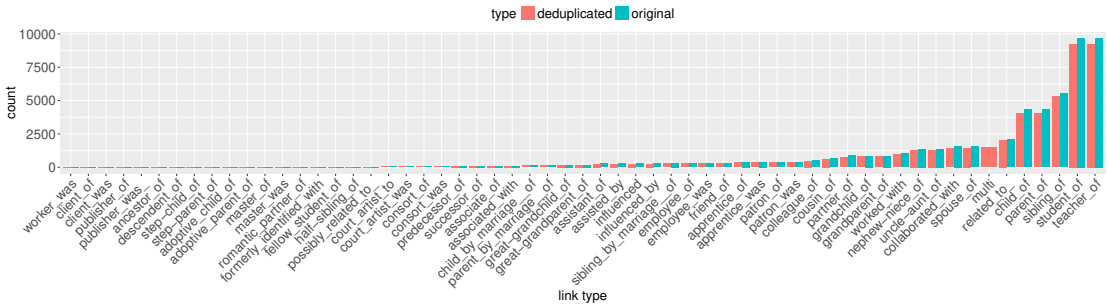


Figure 4.10: Types of ULAN links appearing more than five times

Since the 93 different link types were hierarchically ordered into 9 upper level categories in the ULAN, it was possible to group them accordingly and to look at the resulting distribution. Table 4.6 shows the counts for the higher levels with the four lowest counting ones combined into an "other" category and "multi" representing links with duplicates. Family and teaching relationships were found to have equal importance, together accounting for 84.84% of all links.

personal	other	patronage	multi	professional	teaching	family
274	368	1372	1460	4118	19734	22750

Table 4.6: Counts for upper level link categories

A brief look into duplicated links revealed that many teacher-student or professional relationships also reflected family ties. Table 4.7 lists the top-20 co-occurring link types, showing that many parents were also teachers of their children, that siblings or spouses used to work together and that spouses also used to teach/learn from each other. Other simultaneous relationships seemed more redundant, such as students collaborating with or assisting their teachers.

parent of	teacher of	238	sibling of	worked with	26
child of	student of	238	collaborated with	teacher of	23
sibling of	collaborated with	82	collaborated with	student of	23
uncle-aunt of	teacher of	47	assisted by	teacher of	22
nephew-niece of	student of	47	assistant of	student of	22
spouse of	partner of	44	spouse of	teacher of	17
sibling of	partner of	42	spouse of	student of	17
sibling of	teacher of	36	sibling by marriage (in-law) of	collaborated with	16
sibling of	student of	36	child of	apprentice of	14
spouse of	collaborated with	30	parent of	apprentice was	14

Table 4.7: Top 20 co-occurring relationship types

Interesting differences could be identified between the distributions of birth date differences between linked persons for the different link categories. Figure 4.11 shows separate density plots for each link category for the full network. The distribution of birth date differences for family relationships featured three peaks, the highest one at around same aged persons and two mirrored ones at birth date difference +/- 32 years, suggesting the former to arise from the relatively large number of siblings and the latter from parent/child relationships. The distribution of links with multiple type assignments featured a similar shape, although the differences between the peaks were less pronounced, which could be due to the high presence of parent/child ties there. The distribution for patronage featured the highest proportion of birth date differences to be +/- 10 years, while personal and professional relationships were clearly centered at same-aged persons. Interestingly, the highest proportion of birth date differences in teaching relationships was found to be at about +/- 22 years, more than for patronage but less than for family relationships and featuring a clear "dent" for same aged persons.

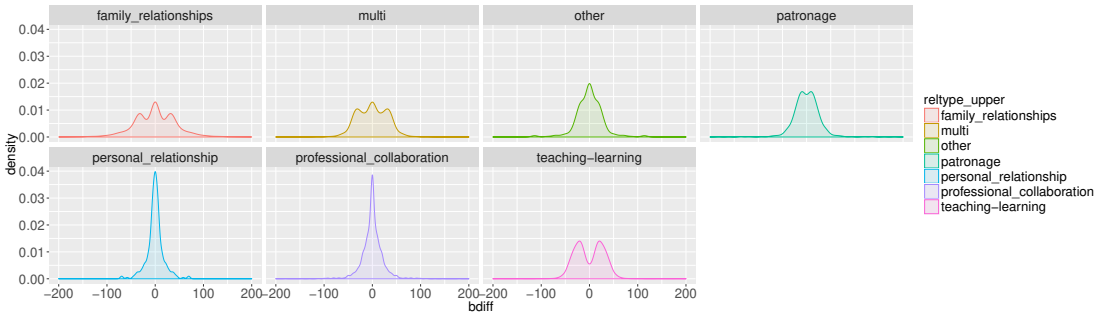


Figure 4.11: Distribution of birth date differences for different link categories

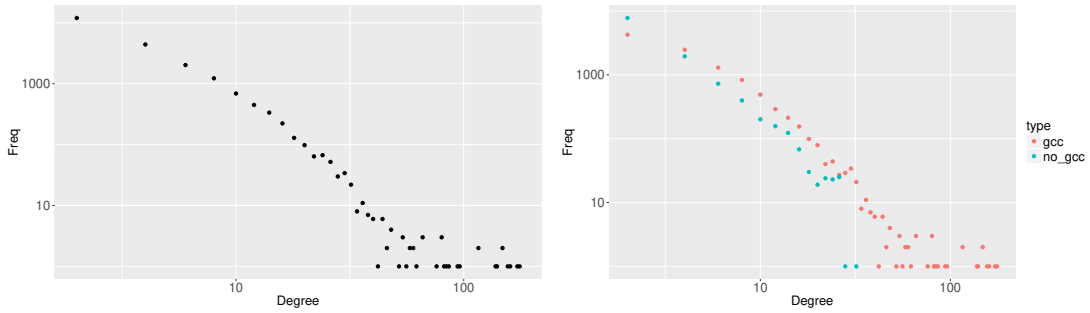
Node degrees

The number of persons n in each identified network component resulted in an upper bound for the maximum number of possible connections for each member of the component, which was, assuming no self-loops, $n-1$ for undirected and $2*(n-1)$ for directed networks. As visible in the network visualization of the individual components in Figure 4.8, however, the majority of persons appeared to have far less connections than that, although some of them appeared to have more than others. It was thus of interest to study these so-called node degrees in more detail. A straight forward way to study node degrees was to consider their distribution in the overall network and in selected subsets of it.

Figure 4.12 provides plots of the node degree distribution in log-log form, with the x-axis representing the observed node degrees and the y-axis the respective counts of person records having the corresponding number of links. Subfigure (a) shows the plot for the full ULAN network, while Subfigure (b) shows two separate distributions, one for members of the GCC and the other one for the members of all the other components. As expected, the plots revealed a tailed degree distribution in the ULAN, with many persons having very low, few persons having very high numbers of links, which became visible via the relatively linear relationship between the log of the node degree and the log of the respective person counts.

While the degree distribution for the GCC (shown in red in Figure 4.12 (b)) appeared to be similar to the overall distribution (Subfigure (a)), the distribution of the combined "non-giant" components (shown in blue in Figure 4.12 (b)) in turn featured a clearly higher number of nodes of degree two while having lower counts for all the other degrees and moreover no occurrences of node degrees larger than 32. The former was due to the high number of independent two-person components, already shown previously in Table 4.4, while the latter clearly reflected the upper bounds of possible links in the smaller components. The highest degree in the GCC was 176 (a person mutually interlinked with 88 persons), which was quite low given the 10,443 possible persons each member of the GCC could be connected to, but explainable by the strong tendency in the ULAN network to only connect contemporaries. It was in turn remarkable to find a node of degree 32 and other relatively large degrees in the set of smaller components, the largest

of which having only 39 members, which suggested that there were specific subgroups of persons which were very tightly interconnected with each other.



(a) (b)

Figure 4.12: Degree distribution of nodes in the ULAN network

The observed linear relationship between the logarithms of the node degrees and those of their frequencies reflected a very common observation made in many social and other type of networks, which were often found to have so-called "heavy-tailed" degree distributions, i.e. the probability for nodes of very high degree was higher than it would have been if the degree distribution was exponential. One of the most prominent conceptions in this regard referred to such networks as "scale-free", having degree distributions following a power-law, a generative model for such networks was proposed by Barabási in [Barabási, 2002]. As discussed by Clauset et al. in [Clauset et al., 2009], however, this property did not always apply to networks encountered in the real world and the authors therefore suggested to also consider alternative classes of distributions to better fit the observed distribution. While the discussion of finding the best fitting theoretical distribution for the observed ULAN network was beyond the scope of this work, it was nevertheless of interest to look at the present degree distribution in more detail.

The linearity observed in the ULAN became more visible when plotting the distribution of node degrees by their complementary cumulative distribution function (CCDF), representing the probability (y-axis) for nodes in the network to have a degree greater than or equal to a given degree (x-axis). Figures 4.13 (a) and (b) show the CCDF for the full ULAN network in form of black dots, featuring two prominent "bends", the first one at about degree 32 where the initially slightly curved decay turns into an approximately straight line and the other one at about degree 130, where the straight line quickly falls off. While "cut-offs" such as the latter were reported — although without discussing possible reasons for these observations — in [Clauset et al., 2009] for various real-world networks, a degree distribution with a shift from a curved to a linear log-log decay was already observed in an analysis of a huge online social network by Ahn et al. in [Ahn et al., 2007], who referred to this as multi-scaling behavior and suggested that different types of networks underlying the full ensemble were responsible for this appearance. It was therefore of interest to see if any of the observable characteristics of the ULAN network could be found as distinguishing feature in this regard.

The separation of the ULAN network into the GCC and the group of all the remaining components, shown in Figure 4.12 (a), did not yield a reasonable explanation, since the GCC featured a degree distribution almost similar to the one of the full network. Separating the network by the individual types of relationships in turn resulted in the degree distributions shown in Figure 4.12 (b), suggesting that the "multi-scaled" degree distribution of the full network was mainly caused by family and teaching relationships, the two most represented link categories encountered in the network and its GCC. Testing for best fitting types of distributions using a state-of-the-art procedure described in [Clauset et al., 2009], its implementation created and described by Gillespie in [Gillespie, 2014], revealed that the degree distribution of the family network appeared to follow a log-normal distribution (Figure 4.12 (c)), while the one of the teaching network followed that of a Power-Law (Figure 4.12 (d)). This observation suggested that the teaching network featured a form of "rich-get-richer" effect where a small number of artist-teachers appeared to accumulate a large number of students, which were far more modest for the remaining majority of teachers. The log-normal distribution found for the family network suggested it to have a more even distribution of connectivity in turn.

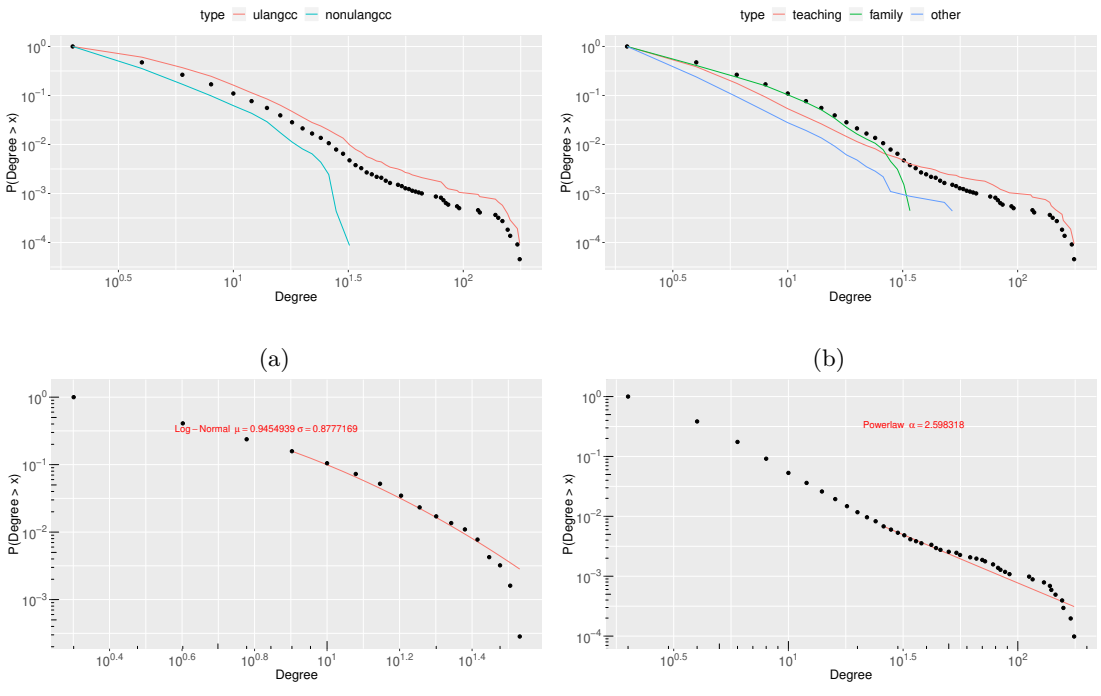


Figure 4.13: Complementary cumulative distributions of ULAN node degrees

One aspect contributing to the more equal distribution of degrees in the family network was the limited number of children that persons had in the ULAN. Table 4.8 shows the distribution of the number of children for ULAN persons divided by parent gender.

Besides the general gender disparity present in the ULAN and coinciding with assumed biological limits for parenthood applying for women, the available "parent of" links showed that women tended to have less children then some men represented there, but also the latter had a maximum of eight children. It has to be noted, however, that at least one well known woman, Maria Theresia, empress of Austria, who is known to have had 16 children, was featured in the ULAN but only two of her children were found to be present there as well. The same was nevertheless the case for Louis XIV, King of France, for whom only one child was found to be recorded in the ULAN while he was believed to be father of even more children than Maria Theresia was mother to.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mothers	147	20	6	2	2			
fathers	2146	451	153	44	14	5	2	1

Table 4.8: Number of mothers/fathers having given number of children in the ULAN

In contrast to parents and the numbers of their children, some teachers were found to have many more students in the ULAN. As Table 4.9 shows, the gender disparity present there again became clearly visible in that there were much fewer female teachers than mothers, while there were many more teachers than fathers in the ULAN. Moreover, some male teachers were linked to as many as 86 students, whereas female teachers were found to have eight students at most.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21-86**
female	64	9	2		1	1		1													
male	2726	686	279	157	78	61	45	28	17	10	14	10	5	3	6	5	1	1	3	2	29

Table 4.9: Number of female/male teachers having given number of students in the ULAN

The main difference found between teaching and family networks in the ULAN, however, was that in all components including the ULAN GCC, the family network tended to feature many small but highly interconnected clusters, none of which significantly larger than the rest, while the teaching network was rather tree like and sparse but still with a giant component. Table 4.10 shows the distributions of component sizes after dividing the full network (top) and the ULAN GCC (bottom) into teaching and family subnetworks. The largest remaining "family component" consisted of 75 persons, while the teaching network still had its "own" giant component of size 5,518, including more than 50% of the members of the unfiltered ULAN GCC.

Since all nodes within a component had to be connected with each other at least such that there was a path between any pair of nodes in the component, it was possible to provide bounds for the minimum and maximum possible numbers of links per component. In case of strictly reciprocal networks such as the ULAN, the minimum number of links interconnecting all nodes in a component of size n is $2*n-2$. Assuming no multiple links between any pair of nodes in the network, each component of a specific size n also has an

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	28	29	35	42	71	75	93	5518
full																															
teach.	993	273	109	51	38	12	10	13	6	8	4	3	1			1	1		2		1	1		1	1		1	1		1	1
fam.	2551	627	246	133	72	45	35	21	14	7	6	5	2	4	2	1	1	1		2			2			1			1		
gcc																															
teach.	188	76	37	25	15	8	7	11	2	5	3	2	1			1			2		1	1		1	1		1	1		1	1
fam.	680	221	94	63	34	20	20	10	9	5	3	3		4	2	1	1	1		1			1			1			1		

Table 4.10: Component sizes and counts for the teaching and the family network

upper limit for the number of its links, $n * (n - 1)$ for the case of strictly reciprocal links. Dividing the number of observed links in each component by this maximum resulted in a measure for the component's link density, shown, averaged across all components of each size, in Table 4.11 for the teaching and the family networks. Besides components of size two, which, naturally, always had full density, the components of the family network on average clearly had higher density then those in the teaching network, while the link counts in the latter were very close to the minimum number of links.

The made observations went in line with general assumptions about the different types of subnetworks encountered. With 34 types of family relationships, it was likely that family members were more densely connected by being (grand-)parents and (grand-)children, brothers and sisters, aunts and nephews, etc. It was interesting to see, however, that the parent-child relationships recorded in the ULAN did not result in genealogical tree-like structures spanning across multiple generations, which was in turn much more the case for the teaching network featuring a contiguous tree-like structure involving thousands of persons. Since, as visible in Table 4.11, the number of links in the latter was slightly above the minimum number of links, however, it was not a "perfect" tree but had cycles in it. This could be imagined to happen for example if two students of a teacher later both became teachers of another student. The finding that family subnetworks represented small and isolated, but internally highly interconnected groups of persons, while about half (5,518) of the persons with teacher/student links were organized in a giant component suggested that if any chronological continuity could be found within the ULAN network, it would be realized through teaching relationships acting as "genealogical structure" of art history through the ages.

4.3.2 Artist rankings

The distribution of node degrees suggested that it was possible to quantitatively rank the linked persons in the ULAN, following the assumption that the number of links to and from a person could be used as a measure for his or her importance in the canon of art history. It was therefore of interest to create such rankings and to compare them with existing quantitative approaches, which could be found in a number of existing art history studies whose authors had mainly ranked artists by counting their appearances in art history literature. These approaches were based on the assumptions such as stated by James Elkins in [Elkins, 2007] that "Canons [...] can be considered as texts about objects or artists or movements [which] [...] can be enumerated", suggesting that the amount of available literature correlates with an artist's historical and or current importance. While

	Teaching				Family			
Compsize	Min. Links	Obs. Links	Max. Links	Density	Min. Links	Obs. Links	Max. Links	Density
2	1986	**1986**	1986	**1**	5102	**5102**	5102	**1**
3	1092	**1098**	1638	**0.67**	2508	**3138**	3762	**0.83**
4	654	**660**	1308	**0.5**	1476	**2346**	2952	**0.79**
5	408	**410**	1020	**0.4**	1064	**2130**	2660	**0.8**
6	380	**382**	1140	**0.34**	720	**1626**	2160	**0.75**
7	144	**148**	504	**0.29**	540	**1388**	1890	**0.73**
8	140	**142**	560	**0.25**	490	**1250**	1960	**0.64**
9	208	**214**	936	**0.23**	336	**1124**	1512	**0.74**
10	108	**118**	540	**0.22**	252	**820**	1260	**0.65**
11	160	**182**	880	**0.21**	140	**396**	770	**0.51**
12	88	**88**	528	**0.17**	132	**462**	792	**0.58**
13	72	**76**	468	**0.16**	120	**524**	780	**0.67**
14	26	**26**	182	**0.14**	52	**358**	364	**0.98**
15					112	**526**	840	**0.63**
16					60	**290**	480	**0.6**
17	32	**34**	272	**0.12**	32	**34**	272	**0.12**
18	34	**34**	306	**0.11**	34	**280**	306	**0.92**
19					36	**58**	342	**0.17**
20	76	**82**	760	**0.11**				
21					80	**294**	840	**0.35**
22	42	**42**	462	**0.09**				
23	44	**44**	506	**0.09**				
25					96	**262**	1200	**0.22**
28	54	**58**	756	**0.08**				
29	56	**56**	812	**0.07**				
35					68	**120**	1190	**0.1**
42	82	**82**	1722	**0.05**				
71	140	**142**	4970	**0.03**				
75					148	**222**	5550	**0.04**
93	184	**194**	8556	**0.02**				
5518	11034	**13436**	30442806	**0.0004**				

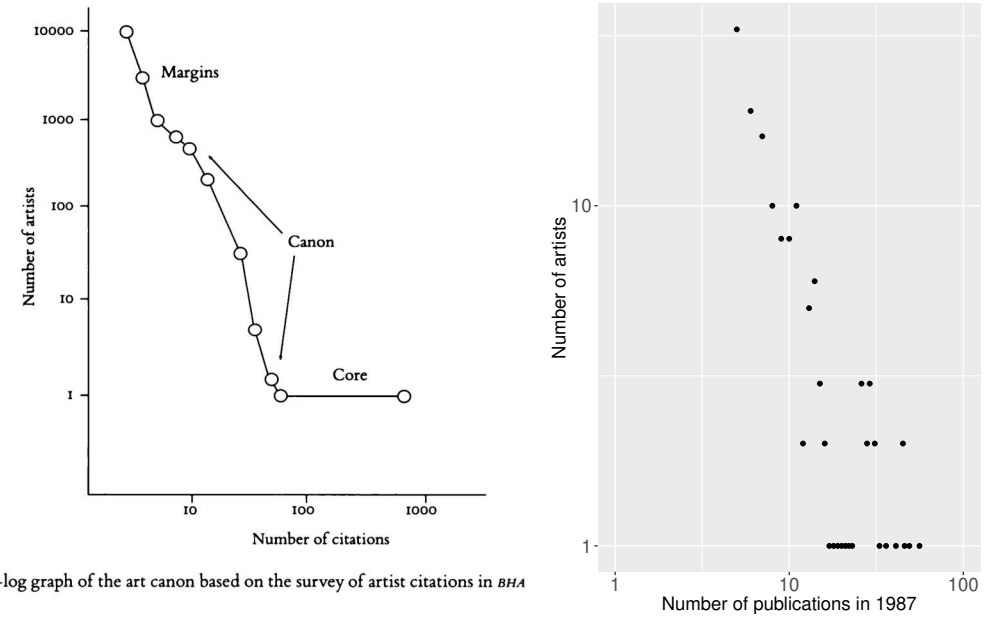
Table 4.11: Link densities in the teaching and family subnetworks, aggregated by component size

Elkin's work presented a ranking of the 50 artists most cited in the Bibliography of the History of Art (BHA)[5], other examples included Robert Jensen's studies of 19th century artist rankings in [Jensen, 2007], based on the appearance of illustrations of their works in European and American textbooks as well as on the total number of citations of their names found in subject headings of the library catalog of the Getty Research Institute[6], and Heinrich Dilly's ranked list of 147 artists each of which featured in at least five art

[5] http://www.getty.edu/research/tools/bha/, retrieved Sept., 8th, 2020
[6] http://primo.getty.edu/, retrieved Sept., 8th, 2020

publications published in 1987, as derived from the Répertoire d'art et d'archéologie (RAA)[7], described in [Dilly, 1992].

Although, according to Dilly in [Dilly, 1992], such rankings were not necessarily stable over time, Elkins in [Elkins, 2007] found that they generally divided the featured artists into three major groups which he described as the "core", the "canon" and the "margins": The core constituted the few "superstars" of art history who received the lion's share of references, while the canon included of artists referenced more then once and the margin represented the majority of artists who only had one or two references. Figure 4.14 shows two log-log plots taken from the identified studies, Subfigure (a) from [Elkins, 2007], representing the visual argument for Elkin's division in to core, canon and margin, Subfigure (b) derived from [Dilly, 1992], Dilly's ranking of 147 artists by the volume of art history literature published about them in the year 1987. The latter contained only artists having at least five publications and thus represented the canon and the core, according to Elkin's terminology, and it was interesting to find a clear visual similarity of both distributions with the one of the ULAN node degrees shown in Figure 4.12.



(a) (b)

Figure 4.14: Log-Log plots of artist rankings by Elkins (a) and Dilly (b)

Comparing the names behind the degree rankings of the ULAN network with those in the existing studies' ranking results — all of the persons featured there were present in the ULAN — showed that there was only moderate accordance between them, at least as far as the direct comparison of top counting persons were concerned. Table 4.12

[7] https://www.inha.fr/fr/ressources/outils-documentaires/repertoire-d-art-et-d-archeologie-raa.html, retrieved Sept., 8th, 2020

lists the top-54 ULAN person records ranked 1-49 by node degree. The distribution of nationalities was clearly dominated by French (30 of 54) persons, Americans came second (10 of 54) and Italians third (5 of 54), while Dutch, Flemish and German persons were only featured once each in the top 54 list. Not only artists and architects were present in the table, five French and Spanish kings, queens and emperors were listed there as well, the two of the royals were the only Spanish representatives in the top-54. The column named "ranked" shows that only few entries (14 of 54) of the ULAN top-54 were present in the scholarly rankings at all.

With one exception, the top-10 names on the ULAN list were French 18/19th century academist/neo-classicist artists, while of the "big four" Italian Renaissance artists Leonardo, Michelangelo, Raphael and Titian for example, which would be expected to occupy high ranks in the list if node degree was directly representative for art historical importance, only Titian and Michelangelo were present in lower ranks (Rank 37 and 47), which was also the case for the famous Baroque artists Rembrandt and Rubens (Rank 29 and 37) — Raphael and Leonardo only occupied ranks 73 and 468, respectively. In contrast, Elkin's list (See Table A.1 in the Appendix) featured Michelangelo, Leonardo, Raphael and Titian on ranks four to six and eight and Rubens and Rembrandt on ranks three and seven, while Dilly (See Table A.2 in the Appendix) found Raphael, Michelangelo and Leonardo to be on ranks four, six and 16, and Rembrandt and Rubens on ranks seven and nine. The top positions in these scholarly rankings in general featured a more mixed selection of artists from various epochs whose names appeared quite reasonable to be high ranked from a subjective point of view based on general art history knowledge.

Jensen's rankings from [Jensen, 2007] (See Table A.3 in the Appendix) were different to those from Elkins and Dilly in that they explicitly focused on arbitrarily selected 19th century artists. The two lists mainly included French artists and in contrast to the ULAN degree ranking, their top positions featured many representatives from early modernist movements such as Impressionism in higher rankings than French proponents of academic art.

A quantitative way of comparing the rankings was to calculate the Spearman rank correlations between them. This measure of correlation checks for monotonic agreement between the rankings without assuming a linear relationship between the underlying counts, which differed due to the different methods of data collection. This revealed weak positive correlation between the ULAN degree ranking and the 50 entries of Elkin's list (0.3368675), as well as to the 147 entries found by Dilly (0.3842861), while the results for correlating the ranks with Jensen's lists each of length 50 yielded rather uncorrelated, negative results (-0.02 and -0.1).

Since the correlations were derived from comparing the ULAN ranking with the individual scholarly examples featuring different numbers of persons, the values couldn't be directly compared with each other. Setting aside Jensen's arbitrarily selected lists and limiting the comparison to the 46 persons featured in both Dilly's and Elkins's rankings yielded a moderate correlation between Dilly's list and the ULAN degree (0.5959780) which exceeded the moderate accordance between Dilly and Elkins (0.5084754), while

#	name	deg	nation	oth	#	name	deg	nation	ranked
1	Laurens, Jean-Paul	176	French	0	28	Carolus-Duran	56	French	0
2	Bonnat, Léon	172	French	0	29	Rembrandt van Rijn	54	Dutch	1
3	Chase, William Merritt	160	American	0	29	Picot, François-Edouard	54	French	0
4	Gérôme, Jean-Léon	156	French	1	29	Bourdelle, Emile-Antoine	54	French	0
5	Cabanel, Alexandre	148	French	0	32	Boulanger, Gustave	52	French	0
5	David, Jacques-Louis	148	French	1	32	West, Benjamin	52	American	0
7	Cogniet, Léon	142	French	0	34	Whistler, James McNeill	48	American	1
8	Lefebvre, Jules	138	French	0	34	Pyle, Howard	48	American	0
9	Bouguereau, William	118	French	1	34	Maratti, Carlo	48	Italian	0
10	Cormon, Fernand	116	French	0	37	Rubens, Peter Paul	46	Flemish	1
11	Lhote, André	96	French	0	37	Fontana, Carlo	46	Italian	0
12	Delaroche, Paul	94	French	1	37	Titian	46	Italian	1
13	Robert-Fleury, Tony	86	French	0	37	Eakins, Thomas	46	American	0
14	Constant, Benjamin	84	French	0	41	Merson, Luc Olivier	44	French	0
15	Ingres, Jean-Auguste-Dominique	82	French	1	41	Vanloo, Carle	44	French	0
15	Couture, Thomas	82	French	1	41	Pils, Isidore	44	French	0
17	Henri, Robert	80	American	0	41	Vernet, Horace	44	French	0
17	Gros, Antoine-Jean, Baron	80	French	0	41	Piloty, Karl Theodor von	44	German	0
19	Hofmann, Hans.	76	American	0	41	Cortona, Pietro da	44	Italian	1
20	Gleyre, Charles	66	Swiss	0	47	Michelangelo Buonarroti	42	Italian	1
20	Louis XIV, King of France	66	French	0	47	Louis XV, King of France	42	French	0
20	Napoleon I, Emperor of the French	66	French	0	49	Duveneck, Frank	40	American	0
23	Le Corbusier	64	Swiss	1	49	Percier, Charles	40	French	0
24	Regnault, Jean-Baptiste, Baron	62	French	0	49	Elizabeth Farnese, Queen consort	40	Spanish	0
25	Léger, Fernand	60	French	0	49	Fuseli, Henry	40	Swiss	0
26	DuMond, Frank Vincent	58	American	0	49	Malfatti, Anita	40	Brazilian	0
26	Wright, Frank Lloyd	58	American	1	49	Philip V, King of Spain	40	Spanish	0

Table 4.12: Top 54 persons in the ULAN ranked 1-49 by node degree

the rank correlation between the ULAN degree ranking and Elkin's list remained weak (0.3688702).

The previously made observation that subnetworks based on different link types in the ULAN network had different degree distributions there suggested to repeat these quantitative measurements for the subnetworks. Removing all family relationships from the network yielded a clearly higher Spearman correlation between Elkin's list of 50 persons and the ULAN degrees (0.4162503 vs. 0.3368675) but a slightly less pronounced correlation between the ULAN degrees and Dilly's ranking of 147 persons (0.3655326 vs. 0.3842861), while correlations with Jensen's top-50 rankings remained very low and negative (-0.03763825 and -0.12582955 vs -0.02 and -0.1). Again looking at the 46 persons covered by both Elkins and Dilly similarly revealed higher rank correlations than achieved when including the family relationships (ULAN/Elkins: 0.4388122, ULAN/Dilly : 0.6161863 vs. ULAN/Elkins: 0.3688702, ULAN/Dilly : 0.5959780). Considering teaching relationships alone, however, resulted in worse correlations for all comparisons.

Additional measures of node centrality

Amongst the many existing network centrality measures, node degree is the simplest, quantifying only the immediate neighborhood of each node in a network. More sophisticated centrality measures also take the global network structure into account. Freeman in [Freeman, 1978] stated the so-called betweenness and closeness centrality measures as more global alternatives to degree centrality. Betweenness centrality for a given node is based on the number of shortest paths, also called geodesics, connecting any other two distinguished nodes in the network and passing through the currently observed node. In

a social network, a person with high betweenness centrality thus more often serves as a "(wo)man in the middle" in shortest paths between other persons than a person with low betweenness centrality. Since it is assumed that shorter paths are preferred to longer ones, such persons thus possess more "power" in a social network since they can control the flow of information between its members. Also taking shortest paths into account, closeness centrality in turn rewards those nodes in the network from which all other nodes can be reached most efficiently, i.e. with the least amount of links. Such nodes can also be seen as those nodes having to rely on the least nodes with high betweenness [Freeman, 1978].

Additional measures of centrality are approaches that quantify the importance of a node in the network in a recursive manner, based on the importances of its neighbors, which are again derived from those of their neighbors, and so on. Koschützki et al. in [Koschützki et al., 2005] referred to this type as feedback centrality measures. They stated one well known example for such a measure to be the so-called Eigenvector centrality, where the centrality of one node in the network is based on the sum of all centralities of its neighbors. It can be shown that this recursive relationship can be solved by calculating the Eigenvector belonging to the largest Eigenvalue for the adjacency matrix of the network, its elements then represent the respective centralities. One variant of this centrality index, the PageRank algorithm, became especially famous in the context of the World Wide Web, serving as the foundation of the Google Web engine. Conceived by Page et al. and described in [Page et al., 1999], this algorithm considers the popularity of a node as a function of the popularity of nodes linking to it and is thus based on in-degree. According to Clauset in [Clauset, 2013], it overcomes some problems associated with the classical Eigenvector centrality which enables it to operate on large, heterogeneous networks such as the WWW.

The outlined different centrality concepts beyond immediate node-degree suggested to look at the respective additional rankings for the ULAN network and compare them to the one obtained through degree centrality. Since closeness centrality only applies to fully connected networks, only the giant component of the ULAN network was considered here. Table 4.13 shows the top-20 ranked persons for degree, PageRank, closeness and betweenness. It became visible that degree and PageRank overlapped in 18 of the top-20 persons, six of which even having the same rank. The overlap was much less expressed between the other rankings and it came to attention that the top 20 persons ranked by closeness and betweenness contained many representatives from the monarchy/aristocracy. Recalling that these two measures rank persons by their position in the structure of the network, it appeared obvious that especially patrons were expected to maintain such positions throughout the history of art, serving as important bridges between many otherwise disconnected persons.

Considering all the persons in the giant component, it was again possible to compare their different rankings through the Spearman rank-correlation. As shown in Table 4.14, closeness centrality appeared to be the most distinguished measure amongst the four compared ones, since the other three appeared to be quite correlated with each other, which was especially the case for degree and PageRank. The high agreement between

Degree	PageRank	Closeness	Betweenness
Laurens, Jean-Paul	Laurens, Jean-Paul	Bonnat, Léon	Louis XIV, King of France
Bonnat, Léon	Bonnat, Léon	David, Jacques-Louis	Bonnat, Léon
Chase, William Merritt	Cabanel, Alexandre	Madrazo y Agudo, José de	Rubens, Peter Paul
Gérôme, Jean-Léon	Gérôme, Jean-Léon	Cogniet, Léon	Charles IV, King of Spain
Cabanel, Alexandre	Cogniet, Léon	Picot, François-Edouard	David, Jacques-Louis
David, Jacques-Louis	David, Jacques-Louis	Gros, Antoine-Jean, Baron	Louis XV, King of France
Cogniet, Léon	Chase, William Merritt	Robert-Fleury, Joseph Nicolas	Madrazo y Agudo, José de
Lefebvre, Jules	Cormon, Fernand	Charles X, King of France	Elizabeth Farnese, Queen consort of Philip V
Bouguereau, William	Lefebvre, Jules	Charles IV, King of Spain	Philip V, King of Spain
Cormon, Fernand	Bouguereau, William	Ingres, Jean-Auguste-Dominique	Vincenzo I Gonzaga, Duke of Mantua
Lhote, André	Lhote, André	Louis XVIII, roi de France	Cogniet, Léon
Delaroche, Paul	Delaroche, Paul	Louis XVI, King of France	Vanloo, Carle
Robert-Fleury, Tony	Hofmann, Hans.	Madrazo y Kuntz, Federico de	Gros, Antoine-Jean, Baron
Constant, Benjamin	Couture, Thomas	Louis XV, King of France	Federico II Gonzaga, Duke of Mantua
Couture, Thomas	Ingres, Jean-Auguste-Dominique	Cabanel, Alexandre	Charles X, King of France
Henri, Robert	Gros, Antoine-Jean, Baron	Boucher, François	Laurens, Jean-Paul
Gros, Antoine-Jean, Baron	Wright, Frank Lloyd	Robert-Fleury, Tony	Lhote, André
Ingres, Jean-Auguste-Dominique	Constant, Benjamin	Girodet, Anne-Louis	Gérôme, Jean-Léon
Hofmann, Hans.	Robert-Fleury, Tony	Philip V, King of Spain	Robert-Fleury, Joseph Nicolas
Napoleon I, Emperor of the French	Le Corbusier	Friesz, Othon	Bouguereau, William

Table 4.13: Top 20 ranked persons for different centrality measures in the ULAN networks

the latter, however, was due to the reciprocal nature of the ULAN network which made it equivalent to an undirected graph, for which PageRank and degree are, according to Perra and Fortunato in [Perra and Fortunato, 2008], highly correlated.

	Degree	Closeness	Betweenness
Closeness	0.15		
Betweenness	0.78	0.25	
PageRank	0.94	0.01	0.8

Table 4.14: Spearman correlations between centrality rankings of the ULAN GCC

4.3.3 The giant component of the ULAN network

A very high ranked artist in both Elkin's (rank two) and Dilly's (rank one) lists was the famous German Albrecht Dürer. In the ULAN degree ranking, however, he was only found on rank 300 (Node degree 22). The same was also true for the German Hans Holbein, having rank 44 in Dilly's list and only rank 862 in the ULAN (Node degree 14). A search for them in the ULAN network revealed that both artists were not part of the ULAN GCC but of a separate component of size 37, almost exclusively composed of Germans. The observation that a very famous artist such as Dürer — who, as stated by Fry in [Fry, 1913], was known to have traveled to Italy at least once and to have made contact with a number of well known Italian Renaissance artists such as Giovanni Bellini — was not part of the GCC therefore raised the question how many of the other persons featured in the scholarly rankings were present there at all, which could be used as an indicator for the overall representativeness of the ULAN network and especially its GCC.

In order to get an estimate on its coverage, all the individual artists present in the scholarly rankings by Elkins, Dilly and Jensen were combined into a list of 178 unique persons, which was assumed to be a representative selection from the group of the most

important figures in art history. Checking the presence of persons from this list in the GCC revealed that 136 of the 178 (76.4%) were present there, while 32 were spread across the remaining components and 10 were not linked to other persons in the ULAN at all. The GCC thus contained many, but not all notable persons present in the scholarly rankings. Table A.4, provided in the appendix, shows the component membership of the 178 persons in the ULAN network as contiguous blocks of the same background color, dividing them by whether they were part of the GCC (gcc), present in other components (oth) or not linked at all (NA). As shown in the Table, also a number of other very popular artists besides Dürer and Holbein were not part of the giant component and out of the 32 artists located in others, only Dürer and Holbein were members of the same one. Interestingly, the Germans were the largest group of persons who were present in smaller components, while the majority of the persons from the list not linked to any other ULAN person were French artists.

Considering the correlations of the rankings from the previously introduced additional centrality measures with the scholarly rankings (limited to persons from the ULAN GCC due to the inapplicability of the additional measures to fragmented networks) revealed that the highest correlation could be obtained through the degree and the PageRank measures. The obtained coefficients for these two did not strongly diverge from the results achieved by correlation with the full network. For Dilly's ranking, closeness centrality yielded the lowest Spearman coefficient of 0.22, betweenness was 0.35, degree 0.43 and PageRank 0.44, while the values for Elkins' ranking were -0.06, 0.27, 0.36 and again 0.36, respectively. Jensen's ranking was equally uncorrelated with all of the four measures. Both Elkins' and Dilly's scholarly rankings were based on keyword counts in bibliographic databases, reflecting the amount of available art history literature available or currently published about the featured artists at a specific point in time. Considering the ULAN to be a product of the available art history information about its featured artists, it therefore appeared quite logical that rankings based on the extent of art history literature were more related to the amount of ties documented for the persons than to their "strategic" position in the topological structure of the resulting network, which was previously shown to particularly reward important "non-artist" actors in this regard, who were not featured in the scholarly rankings at all. It was therefore no surprise that degree centrality and the strongly correlated PageRank performed better than the two structural measures. The overall differences between the ULAN ranking and its scholarly counterparts were nevertheless quite evident, especially concerning the very modest presence of very important Italian Renaissance artists amongst the highest ranks of the former, although they were nevertheless part of the giant component of the ULAN network.

Attribute distributions in the ULAN giant component

Comparing the differences between characteristics of linked persons in and outside the giant component on a more general level, notable differences could be found for their nationalities and birth years. Figure 4.15 shows the absolute counts for the top-20

nationalities in the ULAN subset with associative (red) links and the smaller subset of persons present in the ULAN GCC (cyan). While most of the top-20 nation counts in the two subsets appeared to correlate well with each other, a number of notable differences could be identified. A remarkable observation was the reduced proportion of British persons in the GCC, which disturbed the otherwise unchanged order of the six highest counting nationalities. This suggested that although an obvious focus was put on recording British persons and their mutual links, many of these links tended to be present in separate components disconnected from the "main body" of the GCC. This also appeared to be the case for Spanish persons sharing a similarly decreased proportion there, while the opposite was observed for Russian persons whose links were also present in the GCC in an unusually high proportion. Greek persons in turn appeared to be interlinked with each other almost exclusively outside of the GCC, suggesting that its temporal range of included persons was limited to certain periods.

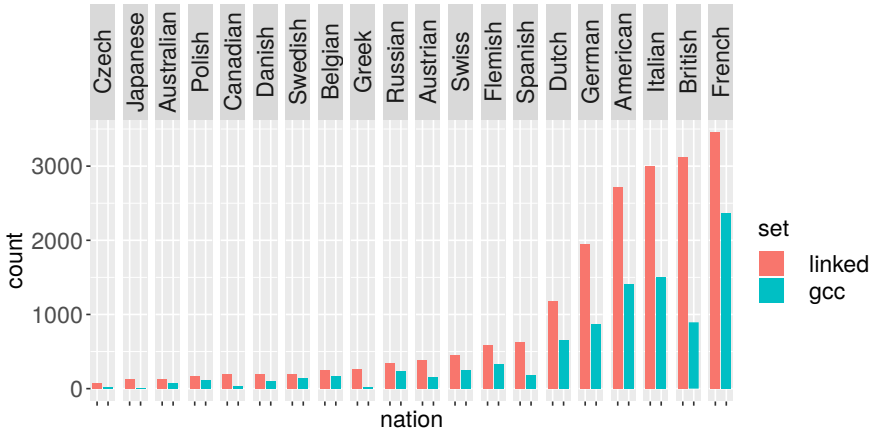


Figure 4.15: Nationality counts in the ULAN network and its GCC

Taking the skewed count distributions of person attributes into account, the additional attribute comparison provided in Figures 4.16 and 4.17 followed the same strategy which was already employed for comparing unlinked and linked person records in Figure 4.7, this time comparing nationality and birth counts for persons with associative links separated between those present in the GCC and those not being part of it. It was again based on plotting the counts on a 2D scatterplot featuring logarithmic axes for each of the two subsets, revealing notable "outliers" by placing them far away from the center line which was visually augmented by including the residuals of an associated Chi-Square test of homogeneity. As far as nationalities were concerned, Figure 4.16 added some notable details to what had already become visible in the absolute counts presented in Figure 4.15. It on the one hand underscored that French persons were especially well represented in the GCC, but also higlighted an unexpectedly high presence of Brazilian persons there. As far as other "non-Western" person groups were concerned, however, it also revealed that their GCC presence was in turn almost zero, which highlighted that

although obvious effort was put into recording information for these groups in the ULAN, it could not make up for the historical divide between them and the main "players" in art history. Moreover, the Figure showed that, besides Greek persons, the GCC also missed many other protagonists from antiquity (Persian and Roman).

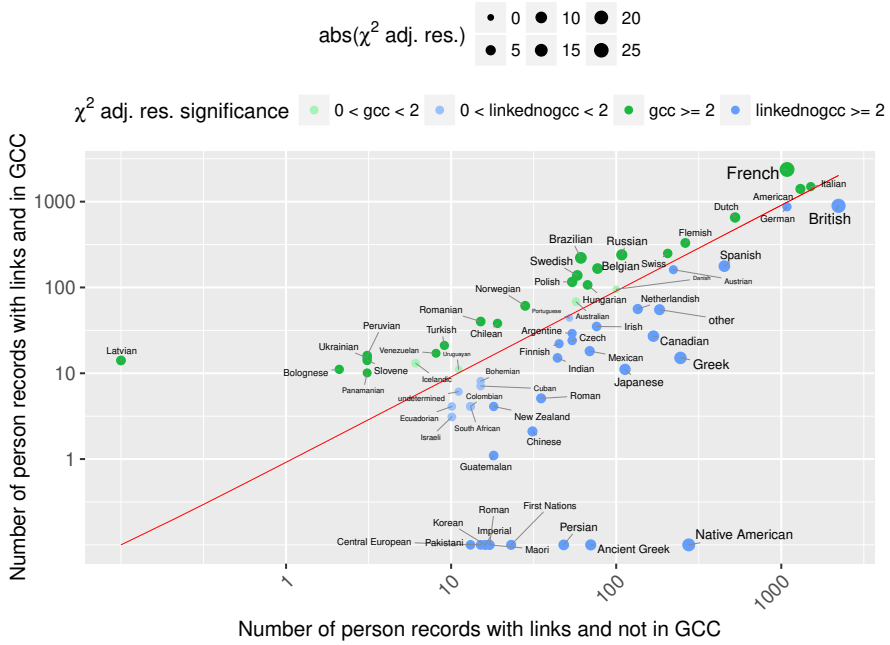


Figure 4.16: Nationality rankings in the ULAN network and its GCC

Missing antiquity also became visible in Figure 4.17 which revealed that persons born before 1500 were less likely to be present in the GCC, while those born before Christ were not featured there at all. Given that the ULAN associative links almost exclusively connected contemporaries born within 75 years of each other, this again highlighted the identified gap in ULAN biographies about persons born between antiquity and the high middle ages, which made an uninterrupted succession of interlinked persons impossible there. Besides that, 20th century persons were less present in the GCC as well, which also resulted in a low presence of women. As far as roles were concerned, the GCC tended to feature persons having the role artist, while others appared to be less present there.

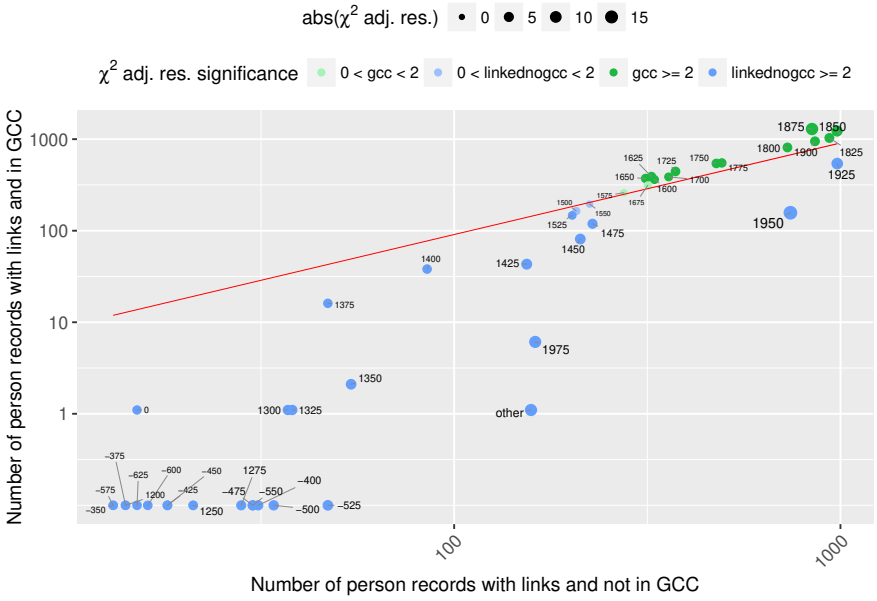


Figure 4.17: Birth date rankings in the ULAN network and its GCC

4.4 Visualizing the ULAN giant component

Figure 4.18 shows a network visualization of the GCC which was created using the software package Gephi[8] 0.8.2. The placement of the nodes was calculated using Gephi's Force Atlas 2 layout algorithm, described by Mathieu Jacomy et. al. in [Jacomy et al., 2014] and belonging to the class of force-directed layout algorithms. The positions of the person nodes in the visualization were therefore determined by the structure of the underlying network, none of their attributes' values was immediately mapped to a spatial position in this regard. This provided visual expressions for a number of features in the topology of the ULAN network which were identified previously.

The high assortativity with respect to birth date differences between linked persons resulted in a stretched visual representation which approximately followed a chronological flow from left to right. One section of the network, shown on the leftmost side of the Figure, mainly contained persons from about the time of the Renaissance, while the section shown on the right hand side mainly represented more recent persons. In between, there was a succession of interlinked persons roughly based on their birth date. The progress of time, however, was not always represented linearly and there were cases where the birth dates of persons did not follow a strict succession. This was on the one hand due to the temporal range of birth date differences between linked persons which included "distances" up to +/- 100 years, leading to individual persons potentially connecting two persons born 200 years apart. Moreover, the authors of [Jacomy et al., 2014] stated that

[8] http://gephi.org/ ,retrieved Sept., 8th, 2020

Force Atlas was specifically designed to bring poorly connected nodes closer to highly connected ones, the latter thus serving as attractors in this regard. Another aspect was the high assortativity of the network regarding similar nationalities and to some extent also roles of linked persons, resulting in tight aggregations of persons of same nationality and role, the latter became visible in the rightmost part of the Figure mainly representing architects of mixed nationality.

The network section shown on the left side of Figure 4.18 was dominated by groups of Italian (green), Flemish (dark orange) and Dutch (orange) persons, whose birth dates spanned the period from about 1400 to 1700. The tight group of Italians in this section represented famous Renaissance artists, while the loose clusters of persons, mainly shown on the upper left, were from the Dutch Golden Age period. Relatively loose groups of mainly Italian and Flemish persons from rather mixed, (Northern) Renaissance, Mannerist and Baroque periods were connected to the Renaissance Italians, shown further to the right of them. The network was clearly dominated by a contiguous cluster of French (red) persons shown in center of the Figure, starting at around 1600 and progressing towards about 1850-1900, including many famous representatives of French classicist/academist art. As far as connections to more recent persons were concerned, this cluster was attached two quite distinct groups, one representing 19^{th} and many well known 20^{th} century Americans (blue), shown in the upper right, the other one composed of many different nationalities, including many important modern artists and also many representatives from the "periphery", such as Brazilian artists (brown). Another important group in this regard was even more detached from the main network, composed of 19^{th} and early 20^{th} Russian (purple) artists shown in the lower right of the Figure.

Seen as a whole, the visualization of the ULAN GCC suggested a chronological process which at first sight appeared to be based on the interactions between three main groups of protagonists: The group of Italians gradually blended into a tightly interconnected group of French persons which clearly dominated the scenery and later faded into a group of American persons who dominated more recent periods, implying a succession of Italian-French-American cultures to be dominating the field across the centuries. The succession of the three main nationalities in the GCC became clearly visible when plotting nationality counts against birth date, shown in Figure 4.19. This graph clearly showed that other main groups such as Flemish and Dutch, but also German (black) and British (light blue) also had high overall presence in the network, which was less visible in the network visualization because they did not feature comparably large contiguous groups there but instead rather tended to appear in small, internally interconnected clusters at different points in time, loosely connected with similar groups from the same nationality but also with other persons in the network.

The juxtaposition of Figures 4.18 and 4.19 suggested that the topology based network visualization was to some extent indeed able to convey large scale chronological features of the network without requiring the direct mapping of time related attributes to visual features. Especially for smaller clusters in the network, however, there were apparent limitations in this regard, such as the placement of many small groups of Dutch Golden

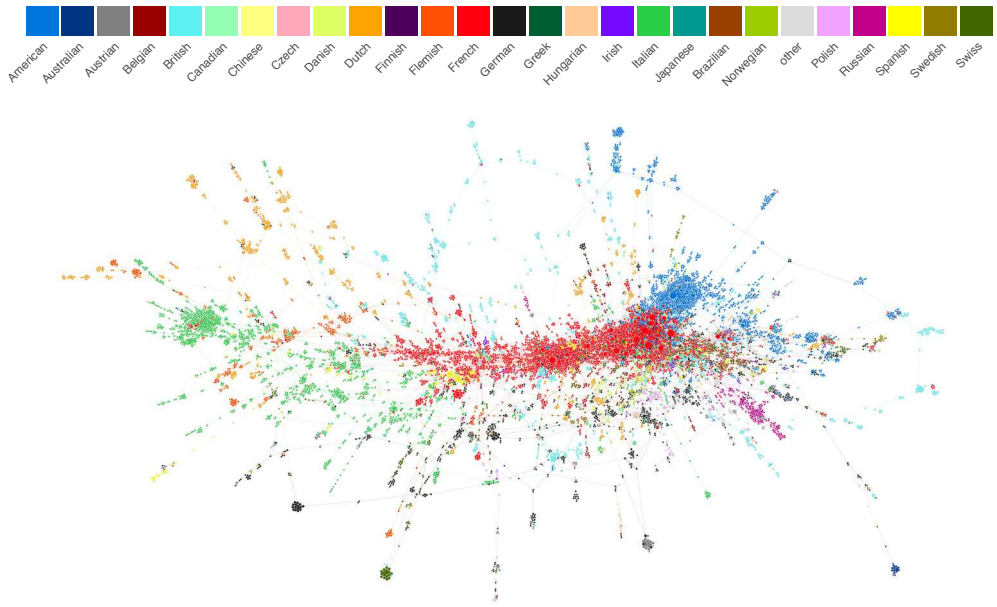


Figure 4.18: ULAN giant component, colored by nationality

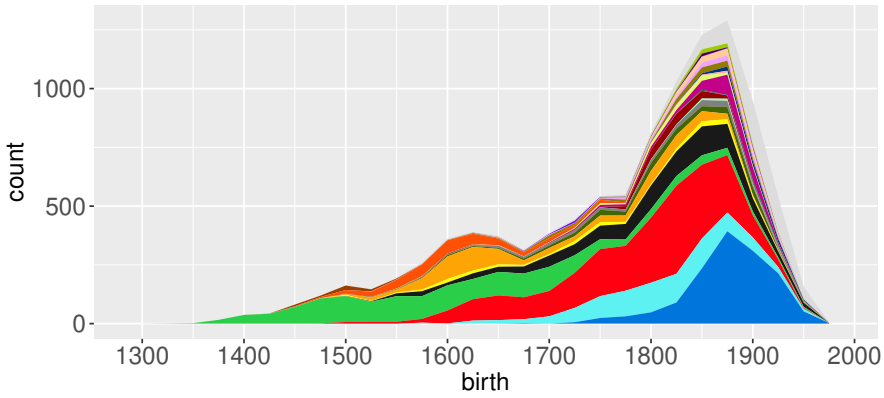


Figure 4.19: Nationalities in the ULAN giant component over time

Age artists close to or even before the cluster of Italian Renaissance artists. It was thus of interest to analyze the chronological adherence of the topology based layout in more detail, which could be achieved by changing the coloring of the existing plot: Figure 4.20 shows the GCC using the same layout method as in Figure 4.18, but colored by birth dates rounded to the nearest multiple of 50, making the temporal organization of the network's layout much better visible. This approach revealed that there was indeed a coarse chronological sequence of birth dates from left to right in the network layout, more accurate in the main body of the network but with shifting overlaps and discontinuities especially in its periphery. Smaller groups of persons which were highly connected within

each other but not with the remaining network were often attached like satellites, for which the layout often ignored any temporal context.

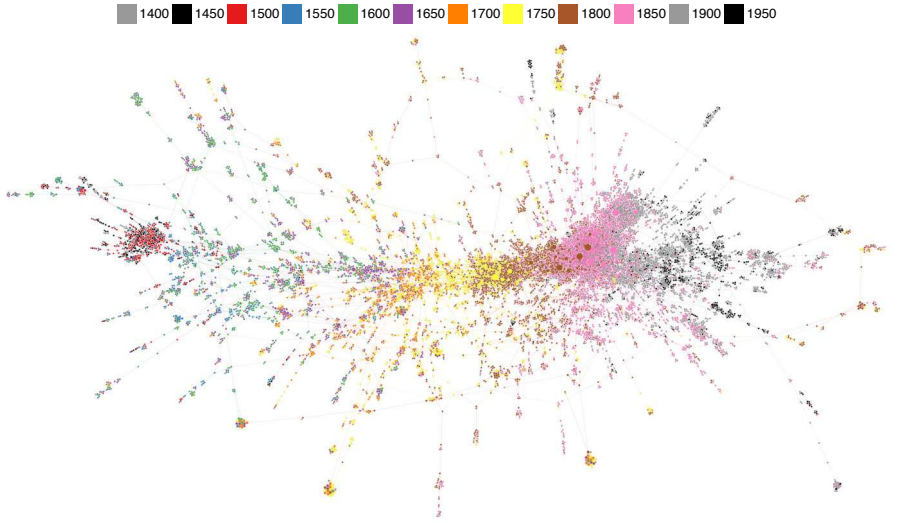


Figure 4.20: ULAN giant component, colored by birth dates rounded to nearest 50

In line with the previously made observations regarding the varying structures of subnetworks of different link types, the highly connected "satellites" seemed to represent a different network structure than present in the main body of the GCC. It was therefore of interest to also visualize the network from such a link centered point of view. Figure 4.21 shows the same layout as the previous displays of the GCC, but with neutrally colored nodes and links colored by their type instead. It revealed that many of the identified "satellite" groups were based on family relationships, while the central part of the network was to a large extent constituted of teaching, patronage and professional relationships. This correlated with the link densities provided in Table 4.11, showing that the teaching relationships of the GCC tended to form a tree like giant component, while the removal of all but family ties broke the ULAN GCC down into many small, internally highly connected clusters. In line with this findings, Figure 4.21 revealed that non-family ties often seemed to serve as bridges between the highly connected family clusters, linking them across the centuries. It moreover suggested that the distribution of link types differed across time, since the cluster of mainly Italian Renaissance persons on the left hand side for example appeared to be constituted to a large extent by links of type patronage and professional collaboration, while the more recent groups of mainly French and American persons rather seemed to be established through teaching relationships.

The observation that some link types appeared to be more present in specific periods than in others suggested a more detailed analysis in this regard. Since almost all links between persons in the ULAN spanned at most three generations, the majority of them even not more than one, the mean date of the birth dates of each pair of interlinked persons was considered to be a suitable measure for counting link types across time.

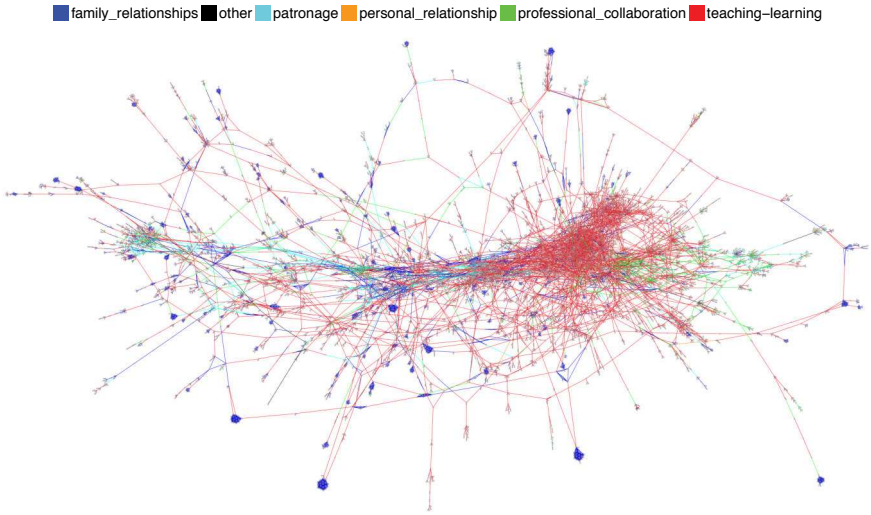


Figure 4.21: ULAN giant component, links colored by type

Figure 4.22 shows such aggregated counts of link types, chronologically binned by the links' mean date rounded to decades. The plot revealed that the high share of family ties decreased with time, also their absolute count clearly shrank during the 19th century. Teaching relationships on the other hand gradually increased until a peak at around 1870, from about 1800 on they clearly outnumbered all other present link types well into the 20th century. A strong increase in professional relationships started shortly before 1900, after a relatively low but stable presence by raw numbers throughout the centuries, similarly to patronage ties. Seen by their proportion, professional ties represented a relatively high fraction of all links of the period between around 1450 and 1500, which later greatly decreased in favor of teaching relationships. For the most recent periods, however, it appeared as if especially the group of modern artists and the cluster of architects were mainly constituted by professional ties again, suggesting a shift away from academy based teacher/student relationships towards more dynamic means of interaction.

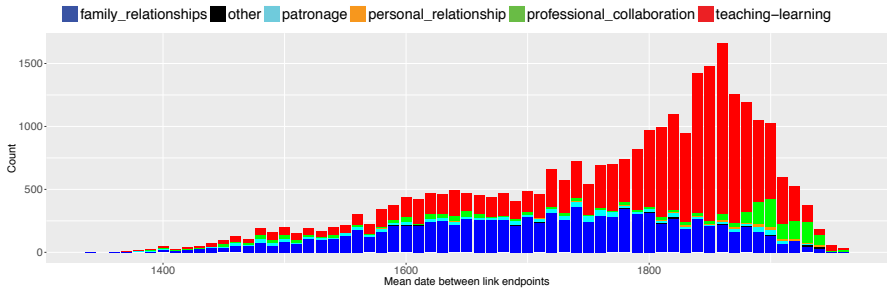


Figure 4.22: Link types in the ULAN giant component, aggregated by mean birth date

The mean birth date of linked persons could also be used to characterize relationships between nationalities over time. Figure 4.23 shows the number of links between persons of the top12-counting nationalities, aggregated by nationality and by mean birth date rounded to decades, in form of a matrix of plots. Each plot represents the number of links from persons of one nationality to persons of the same or another nationality across time. The lower triangle of the matrix shows the values for the full ULAN network, while the upper triangle shows the data for the giant component. The total number of links between the pair of nationalities is shown in red. In general, the shapes of the chronological distributions of nation-nation links did not differ between the full ULAN network and the giant component, although there were interesting differences in the overall counts of links. Some nationality combinations showed only little differences in links counts between the full network and the giant component, i.e. most of the available links between them were also present in the giant component, while others, such as same-nationality links between British or German persons, were more than halved there.

It immediately became visible that for all nationalities, the proportion of links between persons of the same nationality was much higher than those of links to different nationalities, which went in line with the observed high assortativity in this regard. The highest number of links between two different nationalities happened to be found between French and American persons, with a mean birth date of about 1850. The close connection between Americans and French persons could also be seen in the network visualization in Figure 4.18, where the blue cluster of Americans immediately attached to the red French cluster. Less visible in the network visualization were the second most counting international links between Germans and Austrians. The plots generally showed the changing proportions of links between nationalities over time, such as for Italian, Dutch and Flemish persons having strong presences around the Renaissance and the Baroque which decreased in later periods, while links between French or German persons had their peak after 1800, which was also the case for Americans, Russians and Brazilians, with peaks closer to 1900.

Figure 4.24 shows the giant component again colored by nationality but including a selection of labeled person nodes. They were selected from the combined scholarly artist rankings used previously and augmented by additional, arbitrarily chosen persons of high art historical importance. In this visualization, the positions of some of the labeled nodes were adjusted in order to improve readability, which distorted the visualized network structure determined by the layout algorithm in the unlabeled Figures. Provided with names, however, the visualization lost its anonymity and the succession of important persons from art history became visible, coarsely following broad art historical epochs such as Renaissance, Baroque, Neoclassicism and Modernism but with chronological discontinuities on the individual level. One example for such a discontinuity was the German painter Max Beckmann, who was linked to the Italian Renaissance painter Luca Signorelli via one of the few links exceeding contemporary relationships. Although being a 20^{th} century artist, his node was thus placed on the very left of the plot, right next to his Renaissance influence.

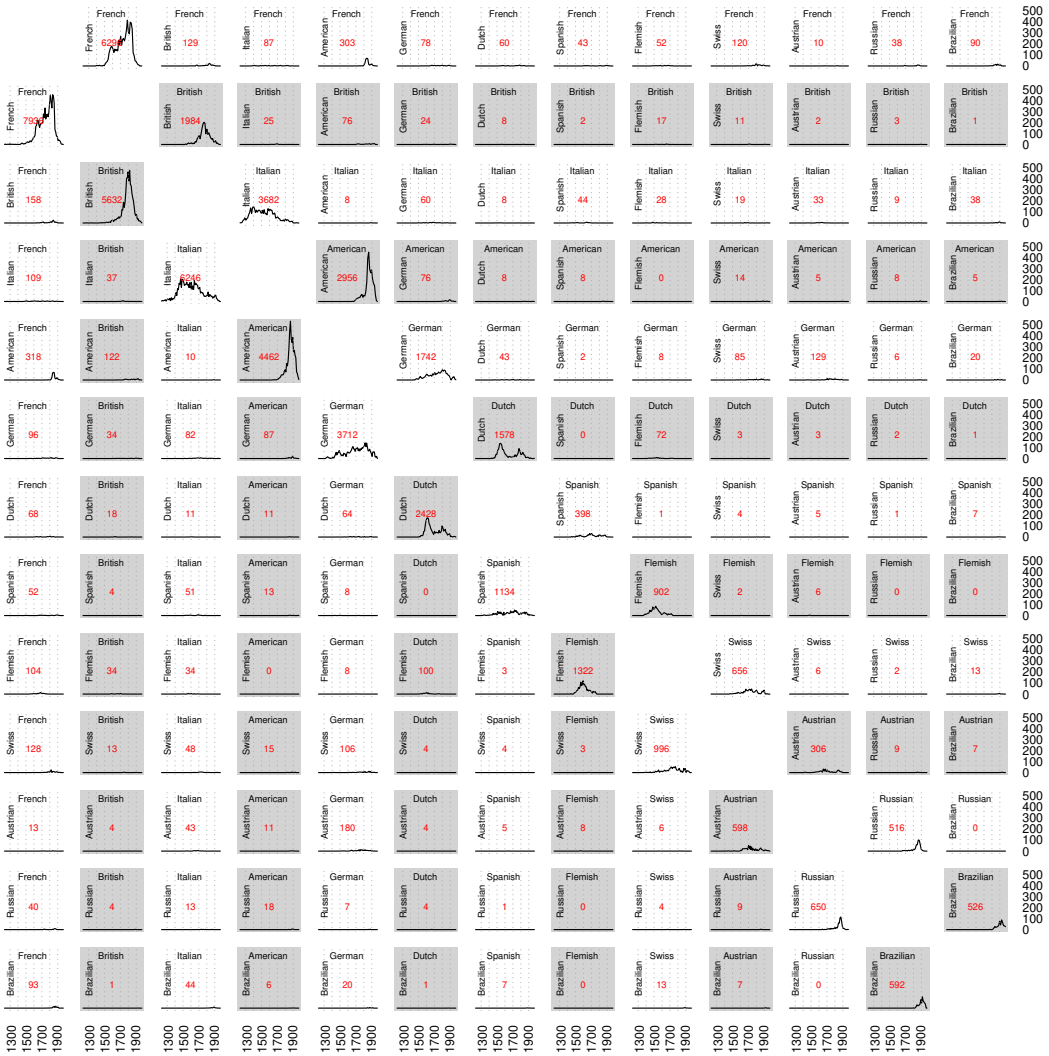


Figure 4.23: Aggregated links between nationalities across time

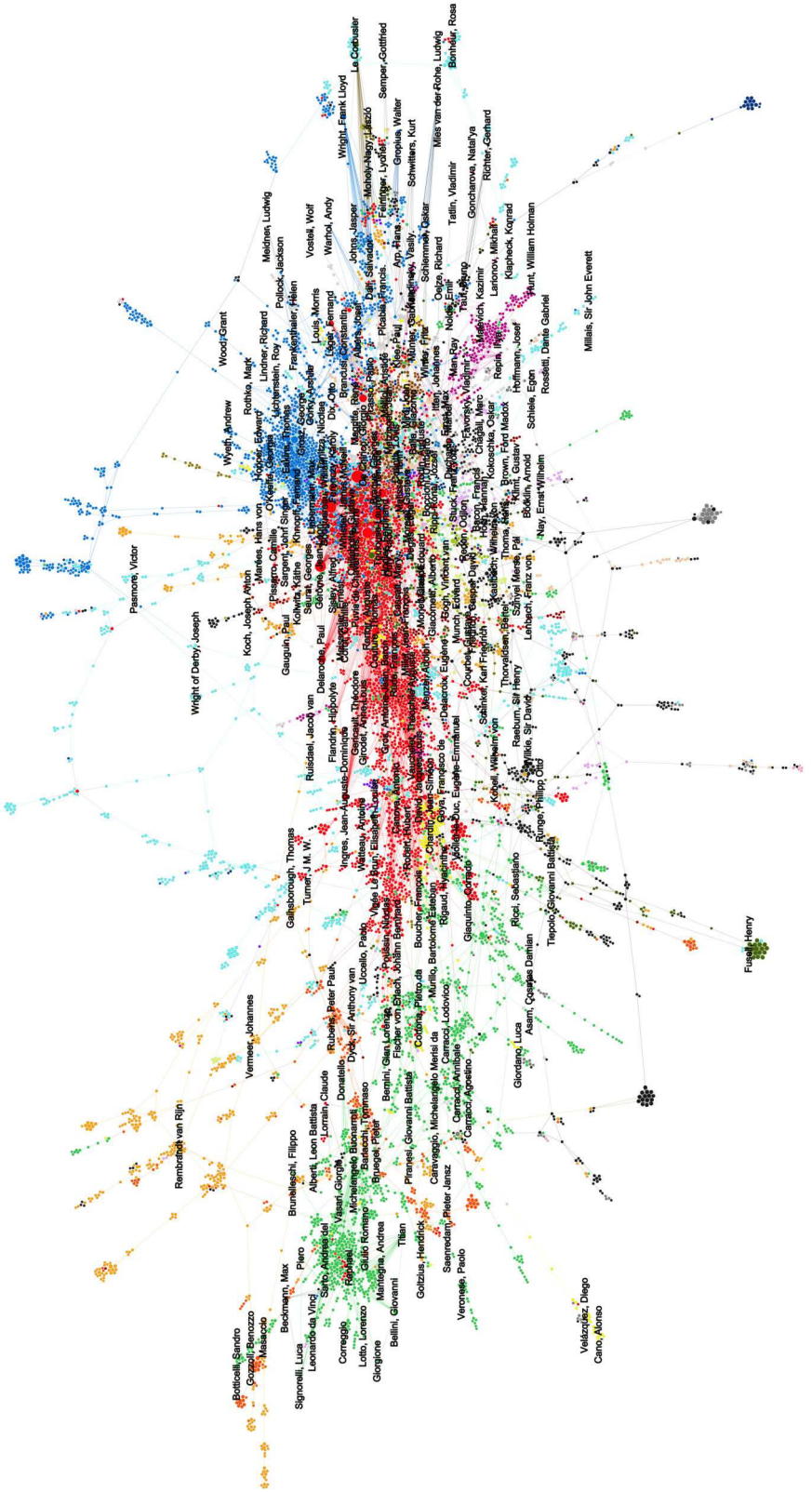


Figure 4.24: Visualization of the labeled ULAN GCC, colored by nationality (**A3**)

Regardless of the identified inconsistencies and -accuracies, the roughly chronological display of the sequence of national clusters, their mutual interactions and temporal parallelisms in Figure 4.18 indeed appeared to show a specific view on art history. Interestingly, the result to some extent resembled an historical example identified in the literature, Newton's chart of art history from [Newton, 1941], for convenience shown again in Figure 4.25. Although Newton's detailed separation of regional Italian schools was not represented in the data and he moreover assigned the vertical positions of the persons to dates, his central top-to-bottom chronological order of Italian "superstars" followed by famous French artists, with Flemish artists such as Rubens "in-between" and Dutch and Spanish artists "in parallel" also became visible in the left-to-right display of the ULAN giant connected component. In Newton's view, British and German artists were also rather at the periphery than in the center of the Figure, the only Germans present there were Northern Renaissance artists Dürer, Grünewald and Holbein — Which were in turn not part of the giant component of the ULAN. In Newton's diagram, German art even appeared like an artistic strand that died out at some point in time, while the other nations' contributions at least lived forth through the visualized influences Newton granted them to have on later generations and cultures.

4.5 Summary

Motivated by the application of the ULAN data for the creation of the 3D Information Landscape for design artifact **A1** (Figure 3.6) and the analysis of its overlap with the portrait relationships derived from the WGA data representing design artifact **A2** (Figure 3.8), the analysis of the ULAN data as a whole resulted in valuable insight into its overall content and the structure of the network of person relationships present for a subset of it. Concerning the compositon of person records, the analysis of the assigned nationality, role and gender attributes revealed a clear dominance of male artists from European and Western nations with a strong emphasis on those coming from English speaking countries. Looking at the subset of person records with person relationships, however, revealed that the "classic" nations of art history, France, Italy, Germany and the Dutch/Flemish countries, played a much more important role there.

The giant connected component which was found in the topology of the ULAN network contained about 50% of all the persons for which associative links were recorded and showed an even stronger focus on persons from continental European nations, although North American persons were strongly represented there as well. This observation was likely related to the predominantly US based origin of the ULAN dataset. Taking the available information about link types into account revealed that the GCC featured the majority of teaching relationships, while smaller components tended to be composed of family ties. Further analysis in this regard also revealed that while family relationships tended to form well-connected subgroups which were only sparsely interconnected with each other, it were the teaching, professional and patronage relationships which formed a contiguous structure resembling the notion of a genealogy spanning across multiple generations.

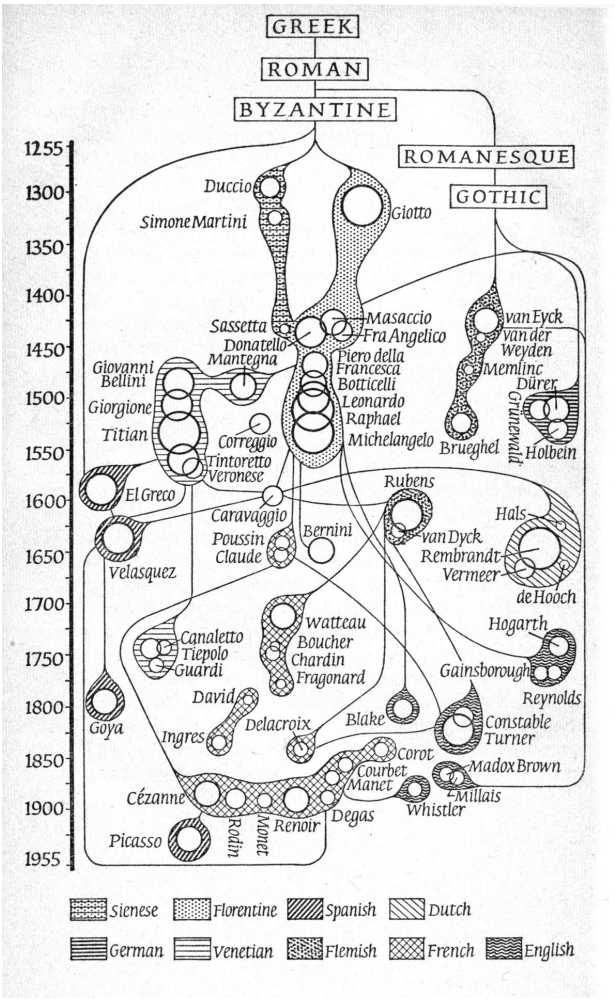


Figure 4.25: Chart of art history, Newton, 1941

Looking at the degree distribution of the network revealed its heavy-tailed nature, featuring a small number of persons having many ties with others, while the majority had only a few connections in turn. The shape of the degree distribution of the ULAN links was found to coincide with observations made by art historians interested in quantifying the popularity of artists via counting references to them found in art history literature. Detailed comparisons of the resulting rankings with the link-degree ranking of ULAN persons showed positive correlation in two of three cases, but generally suggested that counting links in the historical social network found in the ULAN did only partially reflect the general popularity of individuals featured there, rather showing the strong focus on relationships between representatives of French academic art in the dataset. The consideration of additional network centrality measures nevertheless revealed that they could successfully be used to highlight different structural aspects in the network, such as the "bridiging" role that important patrons had there.

The comparison of the persons present in the ULAN GCC with the top ranked persons from the scholarly rankings revealed a number of well-known artists that were not part of the GCC, Albrecht Dürer being an important example in this regard. This provided further evidence for the incompleteness of the network information present in the ULAN, as it was already suggested by the comparison with the portrait relationships derived from the WGA. Overall, however, the ULAN GCC contained more than three thirds of the persons featured in the scholarly rankings, suggesting that it tended to contain a significant share of the most important figures from art history.

Another relevant observation was the tendency of the ULAN network to be assortative, connecting persons having similar traits which especially was the case for nationality and birth. As far as the latter was concerned, birth date differences between interlinked persons rarely exceeded 100 years, which was found to be in line with respective ULAN guidelines for data acquisition and, at the expense of the separation between developments in antiquity from those in the modern age, resulted in the quite well-expressed chronological structure of the network topology which could be made visible by creating the network layout of the ULAN GCC representing design artifact **A3** (Figure 4.24). This artifact could be considered as proof that comprehensive amounts of recorded person relationships from art history formed a network topology which, under the given temporal contraints, unfolded into a chronological map featuring major periods in art history. The quite well-expressed accordance between design artifact **A3** and Newton's hand-drawn map of European art history provided further support for the validity of the approach. The network's assortativity with respect to nationality was reflected in the graph visualization via quite homogeneous and tightly interconnected national clusters, which underscored the deep embedding of the concept of national schools in art history and provided the impression of a certain narrative of succeeding nationalities, superseding each other in being the global center of artistic activity. This reflected the "flagship" role of national artists during the formation of nation states especially in the 19th century which co-incided with and thus also influenced the establishment of art history as academic discipline, an influence which was especially pointed out by a number of critical art history scholars of the late 20th century.

The results obtained from the analysis of the ULAN network suggested further exploration, which especially prompted for the analysis of additional network data sources for comparison. Key questions were if other network datasets featured a similar chronological structure as it was discovered in the ULAN network topology, if they contained similar clustering with respect to nationality and how the composition of featured nationalities was compared to the ULAN network, if and what kind of "national narratives" could be observed there. Other potential objectives were to consider more complex network constructs featuring additional entities than just persons and to explore networks from other domains than art history. With the ULAN being a curated, professional resource, it was moreover desireable to seek for more crowd-sourced alternatives in this regard.

CHAPTER 5

Art History on Wikipedia through the Lens of the ULAN

This chapter discusses the extension of the analysis of large-scale art history person-to-person networks to additional data sources, using data from the free encyclopedia Wikipedia in this regard. It was motivated by the results of the analysis of the ULAN network and the insight from design artifact **A3** (Figure 4.24). Main outcomes of the extension were the methodological framework **M2** for comparing the ULAN network with its Wikipedia counterpart, method **M3** to filter the network of multilingual Wikipedia biographies in order to uncover chronological features of its topology, which was visualized as design artifact **A4** (Figure 5.19), the juxtaposition of visualizations of equally filtered biographical networks in individual Wikipedia language versions representing design artifact **A5** (Figure 5.25) and, last but not least, method **M4** as approach to aggregate individual persons in the biographical network by specific attribute values and filter the resulting aggregate network for significant relationships, visualized in design artifact **A6** (Figure 5.27).

Following a short motivation for using Wikipedia as data source in Section 5.1, an overview on existing Wikipedia research is provided in Section 5.2. The data retrieval procedure and the infrastructure for aligning interlinked Wikipedia biographies with the ULAN data, the latter representing the foundation for the framework **M2**, are described in Section 5.3. The coverage of ULAN biographies in Wikipedia is analyzed in Section 5.4, while the network of ULAN-mapped Wikipedia biographies is analyzed in Section 5.5, featuring the discussion of method **M3** and design artifacts **A4** and **A5** in Sub-section 5.6, the presentation of method **M4** and artifact **A6** in Sub-section 5.7. Section 5.8 summarizes the results of analyzing ULAN-mapped biographies and their networks on Wikipedia.

5.1 Wikipedia as alternative data source to the ULAN

As to be expected from a single dataset, the network content of the ULAN only represented one specific snapshot of art history information, which could for example be shown via the overlap analysis with the WGA portrait relationships (Artifact **A2**, Figure 3.8). Being a collection of art history data created by professionals to meet specific institutional needs, the ULAN content was mainly based on the collections held at the participating institutions. It was therefore of interest to visit other data sources that featured comparable networked structures in order to study potential differences and similarities to the ULAN data.

One particularly interesting type of data source in this regard was user generated content, a concept which had emerged mainly alongside the so-called Web 2.0 paradigm coined and outlined by O'Reilly in [O'reilly, 2010] as a new Internet paradigm focusing on the user as potential content provider. It sought to keep traditional editorial "top-down" approaches to a minimum by providing interactive Web platforms for users to directly upload or create multi-medial content and share it with others there, who in turn were enabled to freely describe, categorize, comment, and sometimes even extend it. As described by Trant in [Trant, 2009b], user based categorization proved to be an interesting approach to alternatively describe museum content via so called user tags, which enabled users to annotate and find objects of interest via their own terminology and language.

While approaches such as outlined by Trant focused on the description of individual art works via custom keywords, other forms of user generated content could be identified as source for biographical information about persons relevant to art history. A particularly useful source in this regard were public online encyclopedias which usually consist of openly available, clearly delineated and mutually cross-referencing articles on distinct topics, including abstract concepts but also concrete entities such as things, places or persons. The most extensive result of a community based effort to create such a public knowledge repository is Wikipedia. Founded by Sanger and Wales in 2001, the free encyclopedia exceeded 42.5 Million articles in 278 different language versions by November 2016 — its English version alone at that time consisting of about 5.2 Million articles — and included distinct 1,432,907 biographies about living and historical persons, each of which present in at least one language version as of January 2017[1].

The Wikipedia coverage of biographies relevant to art history had already been shown by existing studies: Already by 2009, when the English Wikipedia contained about 3 Million articles, Clough et al. in [Clough et al., 2009] identified a significant overlap of about 15,000 biographies between the ULAN and the English version of Wikipedia's structured derivate DBpedia, described by Auer et al. in [Auer et al., 2007]. The latter also provided language version specific datasets with the extracted hyperlinks between all the underlying Wikipedia articles and therefore offered itself as a viable dataset for comparison with the network structure of the ULAN.

[1] `https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Biography&oldid=762549501`, retrieved Sept. 14th, 2020

5.2 Wikipedia research

In parallel to its growing popularity since its inception in 2001, Wikipedia increasingly became subject to dedicated scientific research. Related studies were located along a variety of dimensions. One aspect was the open availability of the fully attributed and time-stamped version history of each article, allowing the scientific analysis of the processes of Wikipedia content creation, often including disputes also known as edit-wars[2], such as described by Sumi and Yasseri in [Sumi et al., 2011]. This lead Pentzold in [Pentzold, 2009] to conceive Wikipedia as a global memory place in cyberspace, in which the traceable negotiation of opposing views provided the unique opportunity to follow the formation of cultural memory. While Pentzold's study was focused on the evolution of an article about a recent event, the 2005 London train bombing, this process could also be seen in a more general context, such as the discussion what to include in Wikipedia and what not. In this regard, Wikipedia maintains so-called notability criteria[3], providing a general outline on what is deemed important enough to be included, further refined into specific guidelines for different aspects such as biographies[4]. Besides criteria for inclusion, other initiatives sought to overcome existing inequalities. In the context of art history, dedicated projects sought to increase the coverage of female artists[5] who traditionally had been underrepresented in scholarly sources.

The coverage of Wikipedia was thus another focus of related research, seeking to assess the nature of the covered content there and how it differed from traditional sources. Dedicated studies compared the coverage in both qualitative and quantitative settings. Qualitative studies focused on the content of the observed articles and sought to find out what Wikipedians deemed to be important and what not, how accurate their assessment of the topic in question was in comparison to the "general body of knowledge". Quantitative studies went beyond such treatment in that they sought to find countable parameters for comparison, such as the overall topics of the observed articles.

5.2.1 Assessment of quality and coverage

One of the first attempts to compare the quality of Wikipedia to existing, commercial encyclopedias was presented by Giles in [Giles, 2005], comparing 42 Wikipedia articles about science topics with related content in the Encyclopedia Britannica via blind peer review. Four articles in each encyclopedia were found to contain serious errors, but the reviewers also identified minor inaccuracies, about four per Wikipedia article and three per Britannica entry. In general, Wikipedia was considered to be relatively on par with its commercial counterpart regarding accuracy but to clearly lack behind in writing style. It is interesting to note that the results of this study caused quite a stir, with Britannica

[2] https://en.wikipedia.org/wiki/Wikipedia:Edit_warring, retrieved Sept. 14th, 2020
[3] https://en.wikipedia.org/wiki/Wikipedia:Notability, retrieved Sept. 14th, 2020
[4] https://en.wikipedia.org/wiki/Wikipedia:Notability_(people), retr. Sept. 14th, 2020
[5] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_artists, retrieved Sept. 14th, 2020

issuing a rebuttal against being put on par with Wikipedia, summarized in an editorial entry in [Nature, 2006].

Rosenzweig in [Rosenzweig, 2006] compared the English Wikipedia and the commercial encyclopedia "Microsoft Encarta" with the professional reference work "American National Biography Online". While the two encyclopedias did not match the overall number of biographies in the reference work, Wikipedia exceeded Encarta in this regard. A close comparison of a set of 25 selected English Wikipedia biographies on figures from American history revealed that the Wikipedia articles were of very arbitrary length, although usually four times as long as their Encarta counterparts. This variable length, however, did not reflect the historical importance of the described persons as it would be judged by historians. In this regard, Rosenzweig considered Wikipedia to put special focus on "curious details", especially regarding personal information, outweighing historically significant information. For him, the Wikipedia view was thus more similar to an antiquarian's or collector's than to a historian's view on the topic, leading him to describe it as "factualist", with an "obsession for list making". The free encyclopedia was nevertheless found to be very accurate regarding names, dates and facts, and the few errors identified in four biographies were rather minor and in part also present in Encarta, one even in the reference work. For the latter, he too identified the quality of writing, but also the quality of the scholarly references as its main advantage, but not without mentioning some editorial bias there.

Following a more quantitative approach by taking the complete English Wikipedia into account, Holloway, Božičević and Börner in [Holloway et al., 2007] presented a visualization of the topical coverage of the more than 1,500,000 unique articles there based on a weighted network of the co-occurrences of the approximately 78,000 categories assigned to the articles by Wikipedia authors. The resulting map showed the variety of Wikipedia's content ranging from mathematics and sciences over history and culture to society and people, broken down into contiguous clusters of closely related categories. Using different node colorings, the authors managed to highlight additional aspects behind the content structure of the encyclopedia, such as the specific topical interests of its most prolific authors or the timestamps of the last edits, providing evidence that specific sections such as those about geographical places were partially automatically populated from other sources.

Halavais and Lackaff described two experiments in [Halavais and Lackaff, 2008], dedicated to mapping the diversity of content in Wikipedia. The first one compared distribution of topics in a random sample of 3,000 English Wikipedia articles, subject indexed by two project related coders, with the distribution of topics present in the bibliographic database "Bowkers Books in Print". The authors found significant differences between the two, with Wikipedia having a strong focus in topics such as music, military, geography, history and the sciences, while the topics having higher presence in the literature database included the social sciences, technology, philosophy, literature, law and medicine. Interestingly, the proportions of entries classified as fine arts were equal amongst the two repositories. Many of the Wikipedia articles for the highly represented topics, however, were found to be brief descriptions of individual entities such as US

and British navy ships or general lists of military equipment, or geographic locations automatically imported from sources such as the US census. Especially articles about music and the fine arts nevertheless had the highest number of edits, which Halavais accredited to some form of "fan effect". Overall, law and medicine were found to have the worst coverage in Wikipedia. The second study was dedicated to check for Wikipedia's coverage of the content of three scholarly encyclopedias from the fields of poetry, linguistics and physics, performed by searching Wikipedia for all the names of their entries. This yielded a coverage ranging between 62.5% and 81.75%, the highest for physics.

In [Samoilenko and Yasseri, 2014], Samoilenko and Yasseri examined 400 randomly selected English Wikipedia biographies on living scholars from biology, physics, computer science and psychology/psychiatry by correlating their article length, in-degree, number of views, editors and edits with common citation indexes such as number of publications, citation count and h-index derived from Scopus. The found 91% of the selected persons to have a researcher profile in Scopus, while the remaining 9% still had verifiable academic background. Only relatively few persons from the Wikipedia selection, however, could be considered as notable by their Scopus h-index (22%-36%) and the correlations between the Wikipedia measures and the citation indexes yielded only very modest, insignificant correlations. Moreover, they checked Wikipedias coverage of a random sample of 219 scientists marked as influential in the Thomson Reuters list of most influential researchers, revealing that only between 22% and 48% of these researchers were present in Wikipedia. Treating the h-index based notability measure as precision and the Thomson Reuters coverage as recall, they concluded that a scientist's presence in Wikipedia did not signify academic notability and that many notable scientists were in turn not represented there, suggesting that Wikipedia is not a reliable resource for finding important scientists and that decisions to include a scientist in Wikipedia were motivated by other reasons than scientific importance, such as popularity in the media.

5.2.2 Analyses of the Wikipedia hyperlink network

While some of the studies dedicated to assess the quality or the coverage of Wikipedia, such as [Holloway et al., 2007] and [Samoilenko and Yasseri, 2014], already used means of network analysis in their methods, they did not apply them in order to directly study the structure of the Wikipedia hyperlink network itself. Studies dedicated to the latter can be distinguished between those focusing on the structure of the complete hyperlink network and those studying topically specific subsets.

An attempt to get insight into the overall structure of the hyperlink network of articles in the English Wikipedia as of 2005, at that time encompassing about 600,000 articles, was described by Bellomi and Bonato in [Bellomi and Bonato, 2005]. The authors crawled the full set of articles, although discarding so-called "lists" only containing links to topically related articles, and collected the hyperlinks between them. The articles in the network were subsequently ranked by two network centrality measures, HITS authority and PageRank. The 300 top articles in the HITS based ranking were mostly about geographical entities such as countries and about historical events or time periods such as WWII or the 1980s, while the PageRank results had a strong focus on religious

topics, revealing that the different centrality measures highlighted different aspects of the present network. Overall, the HITS based ranking appeared to favor more recent topics related to the Western world while PageRank was found to be more global, the authors nevertheless also noted that none of the top-300 rankings contained articles about women.

Zlatić et al. in [Zlatić et al., 2006] analyzed the hyperlink network topologies of several Wikipedia language versions and found that with regard to network structure, different language versions shared remarkably similar topological features, suggesting that the encountered differences were based on different stages of development. Overall, the node degrees in the hyperlink networks of the different language versions appeared to follow heavy-tailed, power-law like distributions featuring many articles with a few and a few articles with many connections. As far as the tendency of pairs of articles to be mutually interlinked with each other was concerned, Wikipedia was found to contain more bi-directional links than association and dictionary networks, but less than the WWW in general. Articles in each language version had high connectedness, i.e. single Giant Strongly Connected Components, taking link direction into account, included more than 85% of all articles in each language version, while single Giant Weakly Connected Components contained at least 98% of all articles in each language. Average path-lengths between any pair of articles were found to be between three and five depending on whether link directions were taken into account, suggesting a small-world effect to be present in Wikipedia as it was generally observed in various forms of networks such as described by Watts and Strogatz in [Watts and Strogatz, 1998]. Wikipedia was moreover found to show higher article clustering than expected, based on the tendency of two connected articles to have mutual connections to other articles.

Focusing on a network consisting only of Wikipedia biographies, Athenikos and Lin in [Athenikos and Lin, 2009] presented a study dedicated to analyzing and visualizing the connections between 330 philosophers as derived from the English Wikipedia's hyperlink structure extracted from their respective articles there. The authors used both hyperlinks between the biographical articles present in their full-text, as well as dedicated influenced/influenced-by links extracted from their info-boxes[6]. The extracted links were weighted based on the number of their appearances, i.e. the link weight represented if one article linked to another article more than once in its full-text. Ranking the philosophers by their degree yielded a focus on figures from Western philosophy, including persons from Greek antiquity but also from the enlightenment and later periods. Similar to the results for the complete Wikipedia networks reported by Zlatić et al., the degree distribution was found to follow a power-law like pattern, and the authors moreover identified the network to follow the small-world-phenomenon in that all interlinked persons were connected with each other through paths of maximum length six. The authors moreover used different means of network visualization to further analyze the network on the level of individual nodes. Visualizations of the links to and from single individuals enabled the exploration of their immediate neighborhood and the strengths of the different relationships, while representations showing the connections of two individuals with

[6] `https://en.wikipedia.org/wiki/Help:Infobox`, retrieved Sept. 14th, 2020

others highlighted the influences they mutually received from or exerted onto their peers. Tree like visualizations were additionally used to explore the small-world structure of the network by highlighting link paths between pairs of interconnected individuals, revealing the succession of philosophers across the ages. Since visualizations of the full network were very dense and did not show meaningful patterns, the authors subsequently employed a reduction technique they called "strongest link paths", retaining only the highest weighted link for each individual. This way, the network disintegrated into different components, revealing clusters of related philosophers.

5.2.3 Comparison of Wikipedia language versions

Another interesting feature of Wikipedia is its multi-linguality stemming from the co-existence of distinct language versions, partially covering similar topics with varying detail but also featuring unique content specifically relevant to the respective culture. This invites to compare different language versions for such culturally specific differences.

In [Hecht and Gergle, 2009], Hecht and Gergle introduced the concept of self-focus bias as an extension to measures of topical coverage such as presented in [Holloway et al., 2007] or [Halavais and Lackaff, 2008]. While the latter concentrated on measuring whether a given set of entities or topics were present in Wikipedia and how they were distributed, the proposed approach also took the inter-article network structure of Wikipedia into account, thus additionally considering the prominence of the observed entities in the network expressed through links pointing at them from other articles. The authors presented a large-scale experiment measuring the importance of Wikipedia articles about geographic locations in 15 different Wikipedia languages, taking all the identified locations and their in-links from articles in each language version into account, and found that the majority of language versions featured significantly more links to regions culturally related to the respective language.

A qualitative comparison of biographies about 15 famous American and 15 Polish persons in the English and the Polish Wikipedia was presented by Callahan and Herring in [Callahan and Herring, 2011]. The authors compared the 30 persons across the two language versions according to different criteria. This included a structural content analysis based on article length, structure of article outlines, number of photographs and notes/references and a thematic content analysis based on close reading, such as tone of coverage (positive/negative), type of included information (personal life, career, education, nationality, political ideology, etc.). Based on this data, the authors concluded that Americans received more positive coverage and were more associated with personal relationships and controversy (romance, sexual lifestyle, etc.), which they related to the overall media interest in the private lives of famous persons in American mass media, while Poles got more balanced coverage. Differences between the language versions were that English articles had a more positive tone, were longer and contained more diverse information, while Polish articles were found to focus on professional achievements and personal life only. Interestingly, the authors only found partial support for the claim for self-focus bias as outlined in [Hecht and Gergle, 2009], but rather found that the different cultures behind English-speaking and Polish-speaking editors resulted in quite

different representations of persons from the own culture, driven by different cultural values. These findings led the authors to conclude that the variety of information in different language versions contradicted the claims of Wikipedia's Neutral Point of View policy which suggests that encyclopedic entries should be consistent across language versions. They appreciated the encountered diversity and recommended that the different language versions should thus be grown organically and not by inter-seeding them with automatic translations.

Aragon et al. in [Aragon et al., 2012] presented a study dedicated to quantifying similarities and differences in biographical networks across 15 Wikipedia language versions, using the set of biographies present in the English Wikipedia as sample, also for identifying biographies in other languages. The authors found the different networks to be well-connected, although of significantly different size. Ranking the figures in the different language versions by various centrality measures, they identified a number of internationally agreed upon personalities considered to be globally important, but also found cultural differences regarding the importance of these figures for the individual languages. A similarity measure between Wikipedia language versions was introduced based on the Jaccard index of common hyperlinks connecting the same notable individuals in each pair of the observed language versions, yielding for example that languages such as Dutch appeared to serve as bridges between the Western and the Eastern cultural spheres. A visualization of the links between famous figures present in most of the observed language versions highlighted groups of globally agreed upon figures important to history.

In [Eom et al., 2015], Eom et al presented a study comparing the top 100 ranked historical figures in biographical networks present in 24 different Wikipedia languages. For the top-ranked persons, the authors analyzed the global and language-version specific distributions of their birth places, birth dates and gender, comparing the coverage in the respective Wikipedia language versions in this regard. The global distributions showed that across the different languages, most of the highest-ranked persons were from Western countries, born after the 17^{th} century, and were male. Moreover, a global top-100-ranking of persons derived from averaging the rankings from all language versions overlapped with a scholarly top-100 ranking by up to 43%. The local distributions of birth places of the top-100 persons in each language on the other hand revealed that the majority of the featured persons were born in countries related to the respective language version. The rankings nevertheless also contained historical figures also present in those of other languages, enabling the authors to construct a "network of cultures", where the notion of "culture" was related to the language of each Wikipedia version. The ties between two cultures were weighted based on the appearance of persons of one culture (mapped based on the main language spoken in the region of their birth place) in the top-ranked figures of another culture, yielding a bi-directional network where the weight of each in-link represented the influence of one culture on others (The presence of persons of this culture in top-rankings of other cultures) and the weight of out-links the other cultures' influence on the culture itself (the presence of persons of other cultures in the top-ranking of the given culture), respectively. While the cultures with highest importance were mainly of

Western origin, such as English, German and French, the juxtaposition of the in- and out-ranks of each culture revealed that specific cultures such as Arabic and Greek appeared to be those with the most balanced importance for and openenness towards other cultures.

5.2.4 Wikipedia and art history

As far as the relationship of art history with Wikipedia is concerned, the encyclopedia has already quite early been recognized as potential tool for art education, e.g. by letting students edit Wikipedia articles about artists, as suggested by Buffington in [Buffington, 2008]. Regarding the consideration of Wikipedia as source of information, however, comments have been more cautious. Hilles in [Hilles, 2014] commented on not further specified English Wikipedia articles about modern and contemporary art history which were found to be missing important content and lacking in writing quality. Moreover, the author experienced a specific bias towards easily accessible content available in freely available online sources, such as the New York Times, which she explained by the Wikipedia requirement to only provide content that had already appeared in existing publications. Hilles nevertheless argued that regardless of any potential shortcomings, it should be taken as a fact that the free encyclopedia is widely used and the focus thus be put on teaching the society, especially students, for its responsible use.

Describing her conversion from "strictly against" towards recognizing Wikipedia as valuable educational tool and important public information resource in general, art historian Hamlin in [Hamlin, 2017] outlined an educational approach she developed which combined art education with critical reading of online sources and at the same time contributed new knowledge about female artists to Wikipedia. The author explicitly mentioned the value of large-scale initiatives such as the Art + Feminism Wikipedia-edit-a-thon[7], with its aim to raise the percentage of female Wikipedia contributors and the coverage of female artists in the free encyclopedia, for highlighting marginalized artists and thus contributing to a more critical reading of art and art history.

Besides special efforts such as Art + Feminism, there are other dedicated curatorial initiatives by Wikipedia editors to improve the quality and coverage of art history related Wikipedia content. The "WikiProject Visual Arts"[8] for example serves as a general forum for to discuss organizational and style related issues for articles concerning visual arts content. Projects such as the "WikiProject Biography/Arts and entertainment"[9] in turn deal with specialized types of articles, such as artist biographies.

Quantitative Wikipedia research on art history content

Only few studies were found which were dedicated to quantitative research on art history content on Wikipedia. Besides the previously mentioned work by Clough et al.

[7] http://www.artandfeminism.org/, retrieved Sept. 14th, 2020

[8] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Visual_arts, retrieved Sept. 14th, 2020

[9] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Biography/Arts_and_entertainment, retrieved Sept. 14th, 2020

in [Clough et al., 2009] who studied the coverage of ULAN artist records in DBpedia without considering the link structure itself, two more works could be identified, which, however, only appeared after the work presented here was initiated.

A recent study by El Zant et al. in [Zant et al., 2018] was based on 223 painters featured in the top-100 PageRank ranking in seven different Wikipedia versions based on 3,334 person names identified via the Wikipedia "List of painters by name" article. The authors analyzed a derivative of their mutual network based on a transformation of its original adjacency matrix which they referred to as the "reduced Google Matrix", which included direct links as well as indirect links via other persons. By assigning artists to broad categories "Cubism", "Fauvism", "Impressionism", "Great masters" and "Modern 20-21", the authors showed that persons of the same category tended to be closely interlinked with each other. The authors also analyzed a bi-partite network between artists and countries constructed via the same approach, which they used to highlight the importance of each artist for various world countries.

An earlier study by Kitromilidis and Evans presented in [Kitromilidis and Evans, 2018] used the Web Gallery of Art to identify painters on Wikipedia and constructed a network of 2,474 artist biographies connected via 9,568 links. The authors applied the Louvain detection method to the network which revealed that some of the obtained clusters were based on a mixture of common artistic style as well as nationality. Motivated to find a more fine-grained way to detect communities which were more homogeneous with respect to the style and nationality, the authors proposed an entropy based measure for describing the homogeneity of a specific Louvain partition, which were used as a cutoff to decide whether each of the identified clusters should be further subdivided by repeated application of the Louvain detection on only that cluster. In order to highlight influential painters, modified betweenness and closeness centrality measures were furthermore proposed which took the refined community structure into account by operating only on shortest paths spanning across different communities. This way, painters with low "traditional" centrality values could be highlighted as neverthless being influential on larger scale.

5.2.5 Summary and motivation

Existing work dedicated to studying the content and structure of Wikipedia revealed a number of important characteristics of the free encyclopedia which motivated its use as alternative data source for the large-scale observation of interlinked biographies from art history. First of all, studies which addressed the immediate critique of the concept of Wikipedia as user-generated resource with respect to accuracy of its content found that the public encyclopedia could in principle compete with its professional counterparts in this regard, although it was also found to be lacking in writing quality and from time to time having more focus on information addressing rather general than expert interest. Regarding topical coverage, Wikipedia was moreover found to cover many different topics from a variety of different domains, but appeared to favor some of them at the cost of others. Many articles about places and specific objects were found to be sometimes automatically sourced from third parties, but content dealing with information about

culture, however, such as music and fine arts, was found to be actively edited and amongst the topics having neither too high nor too low coverage. As far as the coverage of specific topics which had also often been neglected in other encyclopedias or comparable sources of information was concerned, such as biographies about female artists for example, dedicated working groups had moreover been formed in order to increase their presence. Given that Clough et al. in [Clough et al., 2009] had already identified an overlap of about 15,000 English Wikipedia biographies with the ULAN as of 2009, these different aspects suggested that Wikipedia provided actively curated, extensive information about art history related topics providing accurate enough information to be considered as "serious" source for comparing it with the professionally curated institutional resource.

As far as the network structure of Wikipedia hyperlinks was concerned, dedicated studies had shown that Wikipedias in different languages were highly connected and featured much more extensive giant components than for example found in the ULAN, which suggested a much higher link density between articles in the encyclopedia. It was therefore of strong interest to analyze hyperlinks betweeen Wikipedia biographies about persons which were also featured in the ULAN and to compare them with the links found in the professional resource. One key question in this regard was if the topology of the network of Wikipedia hyperlinks between biographies featured similar temporal characteristics as found in the ULAN. Another relevant aspect for such a comparison was that different language versions of Wikipedia had been found to put specific focus on content which was culturally related to the respective language. Given the strong clustering of persons having similar nationality in the ULAN and the overall nationality distribution encountered there, an analysis of the overlap beween ULAN person records and corresponding biographies in different Wikipedia language editions was of great interest as well in order to explore their potential self-focus bias concerning biographies from art history.

The analysis of existing studies regarding Wikipedia and art history showed that besides Clough's study of the coverage of ULAN persons and their key attributes in English DBpedia from [Clough et al., 2009], no analysis had been dedicated to comparing the information from the ULAN with its Wikipedia counterpart, especially multilingual aspects and the comparison of interperson links left many open questions in this regard. Existing work which focused on qualitative issues of art history content on Wikipedia was found to discuss educational and "ethical" aspects of using the encyclopedia as a resource and at the time of the inception of the work presented here, no additional existing studies had been dedicated to quantitative aspects. The two recent studies which appeared a couple of years after the work of this thesis has been initiated focused only on relatively small numbers of Wikipedia biographies and did not take any person attributes into account besides considering the art historical classification of their work. The concentration on small groups of persons was also the case for other studies dealing with biographies on Wikipedia which were not immediately related to art history, such as [Athenikos and Lin, 2009] and [Eom et al., 2015]. The large-scale multi-lingual network analysis of all biographies identified via the English Wikipedia presented in

[Aragon et al., 2012] studied the topology of the network again without considering specific measurable person attributes such as birth, nationality or role.

Considering the bird's eye view on art history gained through the analysis and visualization of the structure of the ULAN network described in Chapter 4 and its basic agreement with existing scholarly points of view, it was thus of interest to analyze the coverage of ULAN persons and their networks in multiple Wikipedia versions from a similar network perspective. This would on the one hand provide a crowd-sourced dataset that could be compared with the professionally curated ULAN network for differences and commonalities regarding the network structure and the high-level view on art history potentially embedded there. On the other hand, it would contribute to the growing body of Wikipedia research by describing the structure of a domain specific subset of Wikipedia articles in multiple languages.

5.3 Data retrieval

The data retrieval process involved different sources requiring separate steps. First, Wikipedia biographies about persons featured in the ULAN had to be identified in multiple language editions of the free encyclopedia. Second, the mutual Wikipedia hyperlinks between the identified articles had to be retrieved for each language. This section describes the process in more detail.

5.3.1 Mappings between ULAN and Wikipedia

In recent years, many Wikipedia articles, especially biographies in the German Wikipedia, have been assigned with identifiers from external authority files such as the "Gemeinsame Normdatei"[10] (GND, "Integrated authority file") maintained by the German National Library. The aim of this process was to increase the quality of Wikipedia by providing means of disambiguation between potentially mistaken entities, such as persons having the same name. A dedicated Wikipedia project page "Authority Control"[11] was established to provide Wikipedia authors with guidelines on how to include external authorities in Wikipedia articles by using a code template[12] for that purpose.

A significant number of Wikipedia biographies therefore already featured such authority information, which had either been added manually by Wikipedia authors, or via automatic mapping. The existing mappings, however, did not contain references to the ULAN, which could be established via another means instead: Presenting the project "VIAFBot" which was dedicated to automatically enrich Wikipedia with authority information, Klein and Kyrios in [Klein and Kyrios, 2013] described a service called the Virtual International Authority File[13] (VIAF) which had been created by the international library community together with a global library service provider with the aim to

[10] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html, retrieved Sept. 14th, 2020

[11] https://en.wikipedia.org/wiki/Wikipedia:Authority_control, retr. Sept. 14th, 2020

[12] https://en.wikipedia.org/wiki/Template:Authority_control, retrieved Sept. 14th, 2020

[13] http://viaf.org/, retrieved Sept. 14th, 2020

unify authority files from different institutions under common identifiers by automatic means. The aggregated authority files came from many institutions around the world and the ULAN was amongst them. The main aim of the VIAF was to match individual records from different authorities against each other in order to unify corresponding ones under a dedicated unique VIAF identifier.

One feature of the VIAF was that some of its records also contained references to corresponding articles in the English Wikipedia which had been established via an automatic mapping procedure. This information could be used to derive 1:1 mappings between 18,000 ULAN records and corresponding English Wikipedia articles for those VIAF records including references to both of them. The obtained mapping could already be used for preliminary proof of concept experiments. Since one aim of this work was to analyze multilingual Wikipedia content, however, this single language mapping was not sufficient, since it only supported the identification of ULAN persons in other Wikipedia language versions that also had a corresponding article in the English Wikipedia. It was therefore necessary to find independent mappings between the ULAN and non-English Wikipedia versions. Unfortunately, this could only be achieved for the German Wikipedia, whose previously mentioned manually added GND references could be used to independently look-up VIAF records containing both GND as well as ULAN references.

As mentioned in [Klein and Kyrios, 2013], another relevant service emerged at about the same time as the authors performed their Wikipedia enrichment experiments. The fact database Wikidata, conceived in 2012 and described by Vrandečić and Krötzsch in [Vrandečić and Krötzsch, 2014], is based on the idea to provide structured descriptions of uniquely identified central entities in order to serve as central fact database for multilingual Wikipedia content, providing unambiguous and language-independent information about, for example, birth dates, country sizes and occupations. In order to fulfil this purpose, each Wikidata record contains references to all its corresponding Wikipedia articles across the existing Wikipedia language versions, which, based on the subject, does not always include the largest English version. Some Wikidata records indeed only feature a single reference to an article from a less prominent Wikipedia language version and thus represent content of highly local relevance. It could thus be expected that mappings between the ULAN and Wikidata would provide a more independent coverage of ULAN records across different Wikipedia language versions.

Throughout the course of the VIAFbot project, Klein and Kyrios had already started to annotate Wikidata records with authority information from VIAF and since then, Wikidata has continuously been updated with new information from various authority files, also including the ULAN. Dedicated handbooks such as provided by Voß et al. in [Voß et al., 2014] were created to support potential authors in extending Wikidata with authority mappings, e.g. by listing the 250+ different Wikidata properties for storing the respective identifiers. Using Wikidata therefore appeared as a feasible approach for obtaining multlingual mappings between the ULAN and Wikipedia.

ULAN mappings in Wikidata

Wikidata records were found to be linked to ULAN identifiers via the distinct property *P245*[14]. All records containing such identifiers could be queried via the Wikidata (WDA) SPARQL endpoint[15], yielding 56,277 WDA-ULAN mappings as of March 2016. The removal of ambiguous mappings (103 WDA and one ULAN identifier occurred more than once) left 56,070 1:1 assignments of Wikidata records to ULAN identifiers. The WDA records were subsequently integrated with the ULAN dataset obtained as RDF from the dedicated Getty Vocabularies Website[16] in July 2015 and loaded into a local triplestore. Checking for ULAN record type revealed that 55,122 of the aligned WDA records were about persons, the remaining 948 about corporate bodies. A search for links from the mapped Wikidata records to 274 different Wikipedia language versions[17] resulted in 48,373 records with at least one connected Wikipedia article, 47,463 about persons and 910 about corporate bodies. 13.89% (7,659/55,122) of the WDA person records with ULAN identifiers thus had no corresponding Wikipedia article, while this was only the case for 4.01% of the corporate bodies (38/948).

Of the 47,463 person records with existing Wikipedia articles, 18,602 were uniquely present in only one language version, while 28,861 existed in at least two versions. In accordance with the distribution of overall sizes of different Wikipedia versions, distributions of unique and complete ULAN biography counts across the languages were highly skewed. Figure 5.1 shows this on a logarithmic scale for the 51 Wikipedia language versions having at least one unique ULAN biography, with blue dots representing the general counts of Wikipedia articles present in each of the language versions, red dots showing the counts of biographies mapped to ULAN records in that language and green dots for unique ULAN biographies only present in that language, respectively. While the overall count of ULAN biographies identified in each of these versions was very strongly correlated with the overall edition size (Spearman ρ 0.955), this was slightly less the case when looking at the correlation between unique ULAN counts and edition size (ρ 0.661). Looking at the unique counts in more detail revealed higher counts especially for those languages whose respective nationalities also had a higher presence in the ULAN, while some non-European language versions such as Farsi or Hebrew in turn contained only very few unique ULAN biographies although they had normal general ULAN coverage with respect to their size. Being the non-Western language with the most unique biographies, the Japanese Wikipedia appeared to be an outlier in this regard.

5.3.2 Completeness of the ULAN-Wikipedia mapping via Wikidata

In order to assess the coverage of the ULAN-Wikidata-Wikipedia mapping, a random sample of 400 ULAN records was taken from the set of 138,506 ULAN records not referenced in Wikidata and checked for the presence of unmapped Wikidata entries and

[14] `https://www.wikidata.org/wiki/Property:P245`, retrieved Sept. 14th, 2020
[15] `https://query.wikidata.org/`, retrieved Sept. 14th, 2020
[16] `http://vocab.getty.edu/`, retrieved Sept. 14th, 2020
[17] Specialized Wikipedia versions such as Wikisource, -quote, species- or commonswiki were ignored

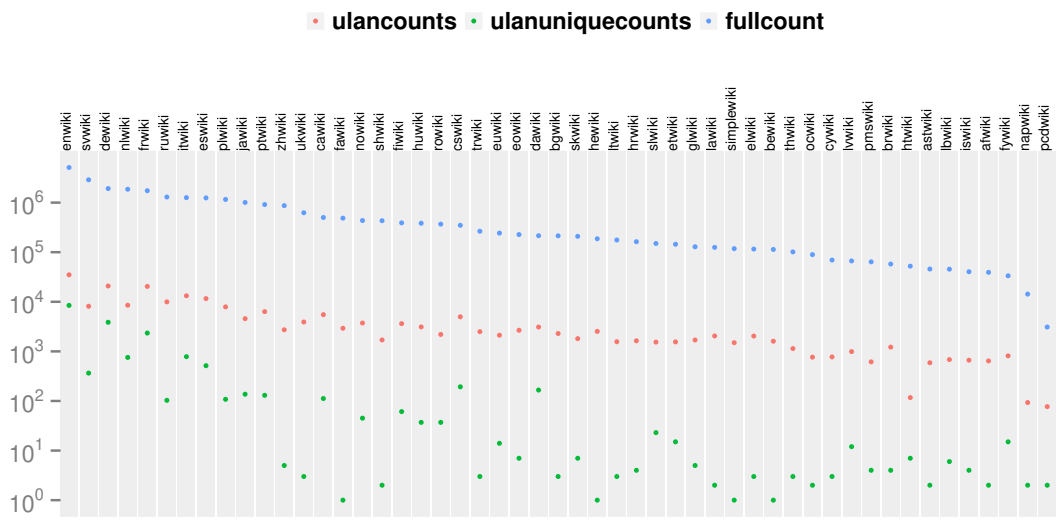


Figure 5.1: ULAN biographies in 51 Wikipedia language versions

respective Wikipedia articles. The number of samples was chosen in order to provide a maximum margin of error of +/-5% at a minimum 95% confidence level, which could already be achieved using only 370 samples, based on the formula for simple random sampling under a known population from [Lippe, 2011].

Of the 400 ULAN records, 351 (87.75%) were "true negatives", neither having a corresponding Wikidata record, based on a search for their name in Wikidata's internal search engine, nor an article in Wikipedia, based on an analysis of Google search results for their name. 39 of these 351 persons, however, were mentioned in secondary Wikipedia articles about their families or companies but since there were no dedicated articles about them in Wikipedia, they were nevertheless counted as "true negatives". 14 additional ULAN persons were represented in Wikidata records without an ULAN reference and without linked Wikipedia records, they too were counted as "true negatives", resulting in 365 (91.25%) true negatives out of 400 overall. All of the remaining 35 ULAN records (8.75%) were "false negatives", having corresponding Wikidata records linked to at least one Wikipedia article, but without any ULAN identifiers present there. Taking the 400 samples based 4.9% margin of error into account, it was therefore assumed at a 95% confidence that between 3.85% and 13.65% of the apparently uncovered ULAN records were nevertheless represented by Wikipedia articles.

The coverage of ULAN-Wikidata-Wikipedia links also raised the question about the correctness of the existing mapping. Regarding Wikidata and VIAF, Klein and Kyrios in [Klein and Kyrios, 2013] reported a 2% error rate for a random sample of 100 VIAF - Wikidata mappings. Assuming a similar error rate to be present in the obtained ULAN mapping reduced the necessary sample size for the full "population" of the 47,463 ULAN-

Wikidata-Wikipedia mappings to as little as 31, considering an error rate of 2% with a 5% margin of error at a 95% confidence level. 100 random ULAN-Wikidata-Wikipedia mappings were therefore analyzed for their correctness, which translated to a margin of error slightly above 2.74% at a 95% confidence level, given an expected error rate of 2%. As it turned out, only one mapping in this sample was wrong, yielding an even smaller error estimate of 1%, up to 3.74% including the margin of error.

5.3.3 Extraction of Wikipedia page links

Since Wikidata only contained structured metadata about its featured entities, an additional source was required to retrieve the hyperlinks between their respective Wikipedia articles in different language versions. This was provided by datasets from the DBpedia project which was conceived by Auer et al. in [Auer et al., 2007] as repository of structured data about Wikipedia content, extracted from semi-structured parts of each article's HTML code which included all hyperlinks to other Wikipedia articles. The latter were available in aggregated form as language specific datasets ("Page Links") which could be integrated with Wikidata records via additional mappings ("Interlanguage Links") of DBpedia and Wikidata identifiers. Using the datasets[18] from early 2016, 46,882 of the 47,463 ULAN-WDA person records (98.78%) could be linked to corresponding multilingual DBpedia records. Figure 5.2 shows the resulting coverage of ULAN-Wikidata entities in 50 DBpedia language versions, the red line represents the counts for ULAN records mapped to Wikidata identifiers in each language, while the green and the black lines show the absolute and relative numbers of DBpedia IDs matching them. The largest editions, English, Swedish, German and French, were covered to about 98% and only few of the other versions had a coverage below 90%.

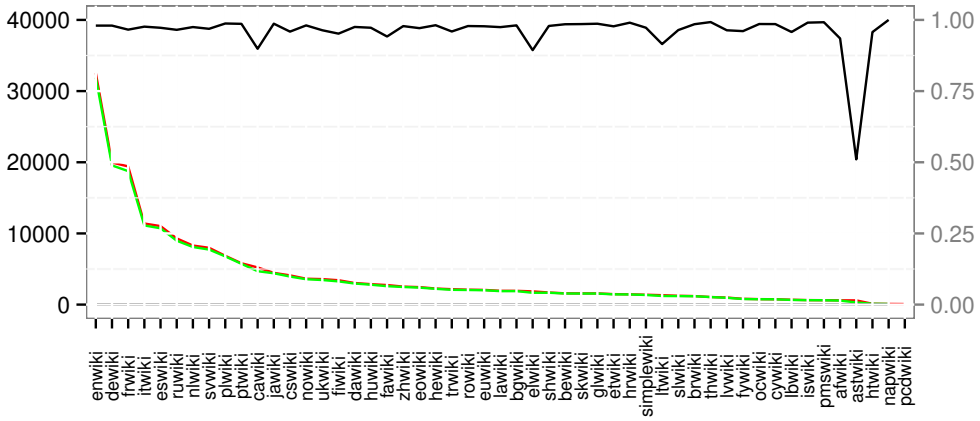


Figure 5.2: Coverage of ULAN-Wikidata records in 50 DBpedia language versions

Like some ULAN records had corresponding Wikipedia articles in only one, others in multiple language versions, the same was the case for the extracted hyperlinks between

[18] http://downloads.dbpedia.org/2015-10/, retrieved Sept. 14th, 2020

ULAN mapped DBpedia records. Figure 5.3 shows the total de-duplicated hyperlink counts (red) retrieved for each language version overlaid by the fraction of unique hyperlinks encountered there (blue). In general, the number of retrieved links corresponded well with the number of ULAN biographies covered in each version, but with some notable exceptions such as the French Wikipedia which featured more links than the German version although it had less covered ULAN biographies. The leftmost column labeled "all" represents the de-duplicated count of all the hyperlinks retrieved in all the observed languages, overlaid by the fraction of those of them only occurring in one language. The fraction of unique hyperlinks (179,254 of 283,368, 63.26%) was clearly higher compared to the 39.2% fraction of uniquely covered biographies.

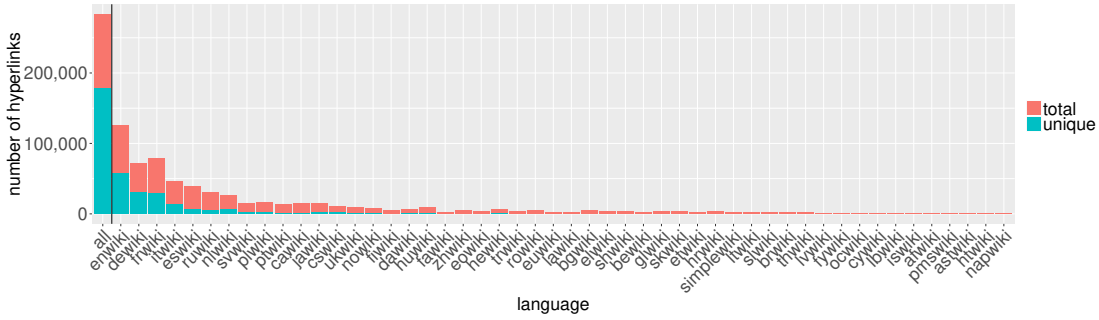


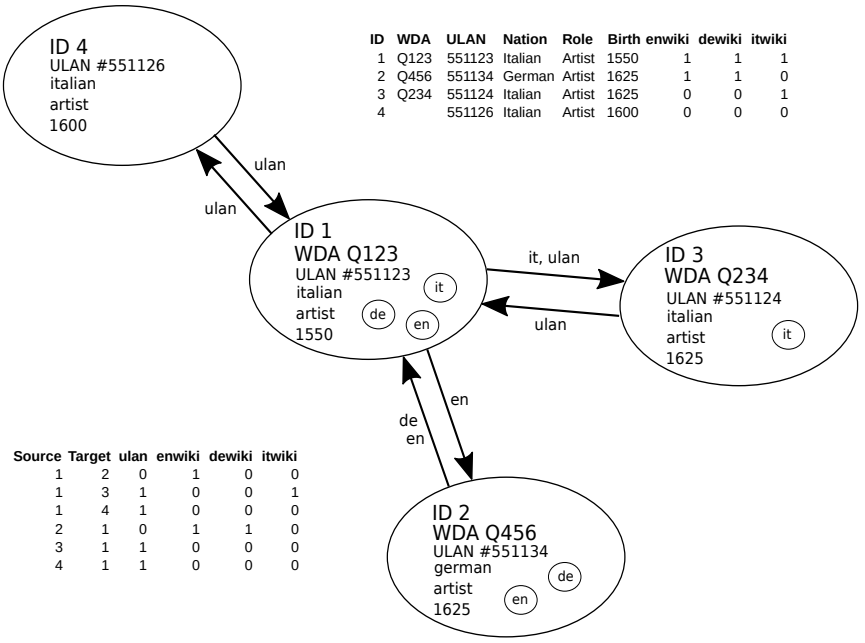
Figure 5.3: Wikipedia links between mapped ULAN biographies, extracted from DBpedia

5.3.4 Data infrastructure used for analysis

Figure 5.4 shows the conceptual structure of the hybrid dataset created by the procedure outlined in this section, based on four example person nodes from the ULAN, some of which having corresponding Wikidata records with references to Wikidata biographies in different Wikipedia language versions, interlinked via six unique links either present as ULAN associative relationship, as hyperlink between the respective biographies in one or more Wikipedia language versions, or both. It incorporates the link structures from both the ULAN and its Wikipedia counterparts in different language versions into one data model, enabling both the analysis of the original ULAN data as discussed in Chapter 4 as well as the comparison with its overlapping Wikipedia counterparts. The data structure consists of two tables, one node table, shown on the top right of the Figure, and one link table, shown at the bottom left.

Each row of the node table represents a unique person, assigned with an internal id which is linked to the person's ULAN identifier and, in case of the presence of a P245 link in a corresponding Wikidata record, the mapped Wikidata ID. All persons featured in the table thus have an ULAN ID, but not always a Wikidata ID. The ULAN attributes preferredNation, preferredRole and birth are stored for each person, followed by columns indicating the presence of a respective Wikipedia biography in one or more of the observed language versions. The link table in turn features rows representing

unique links between entries in the node table . Source and target columns contain the respective node IDs and additional columns indicate the presence of each link in the different sources. The network sketch in the center of Figure 5.4 shows how the four example nodes featured in the node table are connected via links in the link table.



ID	WDA	ULAN	Nation	Role	Birth	enwiki	dewiki	itwiki
1	Q123	551123	Italian	Artist	1550	1	1	1
2	Q456	551134	German	Artist	1625	1	1	0
3	Q234	551124	Italian	Artist	1625	0	0	1
4		551126	Italian	Artist	1600	0	0	0

Source	Target	ulan	enwiki	dewiki	itwiki
1	2	0	1	0	0
1	3	1	0	0	1
1	4	1	0	0	0
2	1	0	1	1	0
3	1	1	0	0	0
4	1	1	0	0	0

Figure 5.4: Hybrid ULAN-Wikidata-Wikipedia link network data structure

5.4 Coverage of ULAN records in Wikipedia languages

39.2% of the ULAN records mapped to multilingual Wikipedia had only one unique corresponding biography in one language version. It was therefore of interest to have a deeper look at the properties of ULAN records identified in the different language versions in order to find out if their individual coverages followed specific large-scale patterns of preference. This could be achieved by observing the distributions of the ULAN attributes nationality and role across the records covered in the different languages.

5.4.1 Coverage of nationalities across languages

Looking at the nationality specific coverage of ULAN records in the different language versions revealed interesting, culturally specific preferences. Figure 5.5 shows the distribution of nationalities in the 50 languages, the top and bottom parts representing absolute counts and relative fractions, respectively. The 26 highest counting nationalities were again colored according to Armytage's color alphabet from [Green-Armytage, 2010],

as used in section 4.3.1. It became evident that the distributions of covered nationalities were related to the respective language version, i.e. the English version (enwiki) featured by far the most Americans and Britons, the German Wikipedia (dewiki) most Germans, Austrians and Swiss, the French most French, etc. Similar observations could be made for smaller language versions, which tended to form culturally related groups, such as the Catalan (cawiki), Basque (euwiki), Galician (glwiki) and Asturian (astwiki) Wikipedias having relatively high presence of Spanish persons (yellow) besides the Spanish Wikipedia (eswiki) itself, which was similarly the case for Russians (purple) who were, besides in the Russian version (ruwiki), higher represented in Ukrainian (ukwiki), Belorussian (bewiki), but also in the Finnish (fiwiki), Hebrew (hewiki) and the Baltic (lawiki, etwiki, lvwiki) language versions.

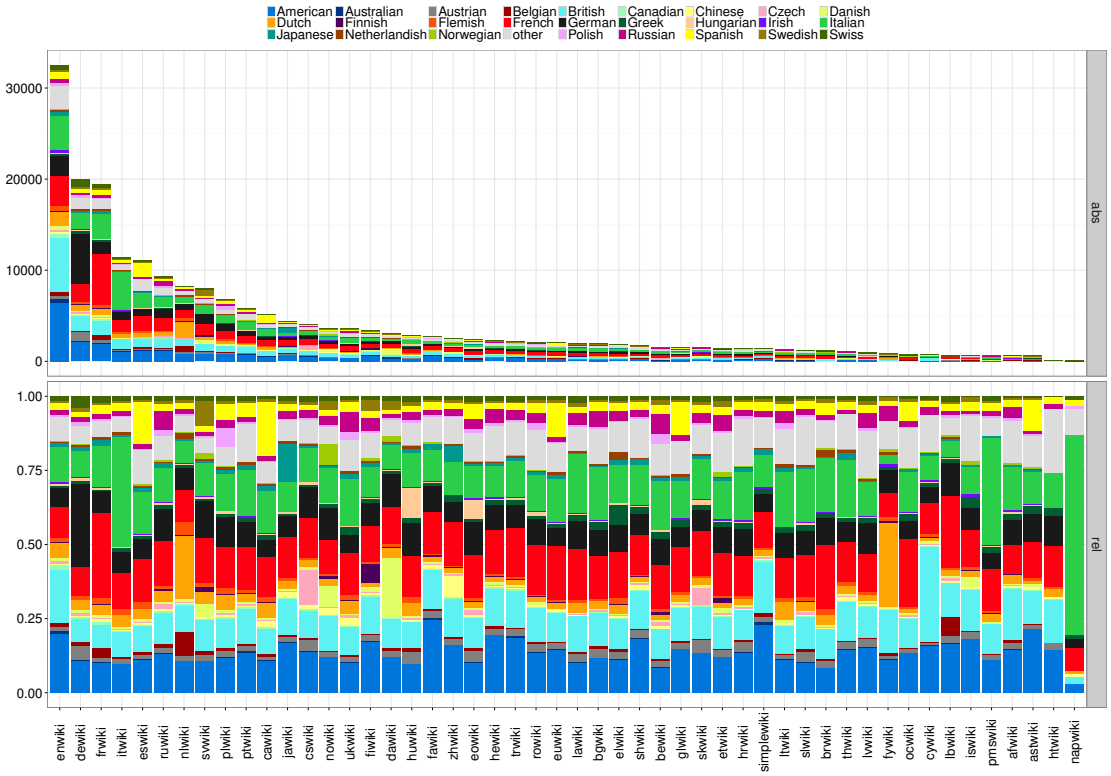


Figure 5.5: Nationality distributions across Wikipedia languages

Figures 5.6 and 5.7 provide the absolute overall and unique only biography counts for a selection of ULAN nationalities in the 20 largest Wikipedia languages. Grey shaded cells mark the language version having the highest coverage for each of the featured nationalities. This again highlighted that the highest covered overall counts of nationalities tended to be present in the respective language version, although with exceptions especially in the English Wikipedia which sometimes had highest coverage for other nationalities as well. In such cases, however, the most closely related language

version often featured the second highest coverage, such as for example for Chinese persons having highest coverage in the English, second highest in the Chinese Wikipedia (zhwiki).

enwiki	dewiki	frwiki	itwiki	eswiki	ruwiki	nlwiki	svwiki	plwiki	ptwiki	cawiki	jawiki	cswiki	ukwiki	nowiki	fiwiki	dawiki	huwiki	fawiki	zhwiki	
209	141	106	38	38	66	28	32	54	17	12	16	34	11	12	12	11	294	9	6	Hungarian
418	188	155	81	89	82	94	332	56	44	31	26	32	19	246	56	604	26	22	15	Danish
109	54	94	29	24	49	19	116	20	14	11	14	9	12	26	222	14	8	11	5	Finnish
175	77	56	27	41	47	25	154	18	9	9	13	12	9	250	32	58	11	14	5	Norwegian
233	202	125	51	57	133	36	24	62	21	13	18	489	27	14	16	7	22	9	10	Czech
419	238	419	112	144	150	53	54	81	94	82	584	41	61	21	56	16	32	54	136	Japanese
122	49	62	31	75	26	32	14	15	201	17	9	7	11	5	10	1	11	7	5	Brazilian
106	76	62	27	65	24	15	18	29	179	18	14	17	10	6	22	7	10	9	10	Portuguese
278	257	139	58	69	195	61	48	438	38	24	26	39	88	28	26	18	22	13	15	Polish
284	238	147	76	71	96	83	681	74	52	37	53	35	32	109	124	92	37	23	21	Swedish
392	207	636	105	123	113	663	101	65	42	56	38	22	29	20	21	27	36	18	16	Belgian
1542	664	669	433	368	346	1760	341	343	155	264	116	95	151	96	49	101	117	51	42	Dutch
499	316	348	175	164	585	140	128	181	110	84	122	101	241	104	148	43	85	55	55	Russian
116	41	35	23	148	16	9	9	11	34	20	12	5	7	4	7	4	3	7	6	Argentine
729	367	549	291	1576	220	172	142	380	262	962	95	80	118	73	77	41	55	66	56	Spanish
3775	1813	2702	4200	1584	1045	630	905	847	918	736	437	458	558	424	337	247	336	286	276	Italian
3295	2012	5567	1416	1579	1383	893	1229	949	836	698	615	550	500	411	422	374	404	389	373	French
419	949	273	173	178	186	117	153	154	101	67	70	158	61	56	62	63	109	70	48	Austrian
2188	5559	1405	776	767	1015	628	1008	690	486	295	325	430	226	403	266	347	322	241	168	German
197	241	122	140	100	126	62	55	72	75	101	36	33	88	43	46	24	39	25	50	Greek
485	785	412	190	134	139	97	118	99	81	52	66	78	39	57	46	46	54	32	32	Swiss
6424	2174	1954	1102	1218	1210	885	848	802	791	562	750	570	367	437	591	365	282	680	405	American
404	76	77	23	30	28	28	15	22	26	8	16	9	7	7	10	6	4	12	7	Australian
5900	1601	1488	937	935	990	754	857	673	688	442	537	384	330	425	429	294	251	358	325	British
474	90	236	42	55	41	39	22	44	37	18	23	19	11	17	25	8	7	15	15	Canadian
204	87	161	44	56	84	25	37	40	33	121	85	61	106	87	16	10	15	30	183	Chinese
531	206	347	199	255	112	374	102	150	68	101	34	28	68	28	12	26	51	13	13	Flemish
237	50	54	37	28	38	25	18	23	23	22	12	7	12	7	13	13	7	11	9	Irish
179	131	51	26	159	31	28	16	15	26	21	16	11	12	8	12	6	8	13	9	Mexican
189	133	139	77	85	57	161	55	87	61	48	31	22	35	18	16	13	28	6	17	Netherlandish

Figure 5.6: Coverage of ULAN nationalities in Wikipedia language versions

It was interesting to observe that a similar pattern was present in both overall and unique only biography counts, although the latter were visibly more concentrated on those culturally related to each language version, providing a strong indication that biographies in Wikipedia were subject to similar self-focus bias as identified by Hecht and Darren for geographic places in [Hecht and Gergle, 2009]. A notable observation were the 100 persons assigned with Spanish preferred nationality in the ULAN having unique presence in the Catalan Wikipedia, which highlighted the political and subjective dimension of the concept of nationalities in general.

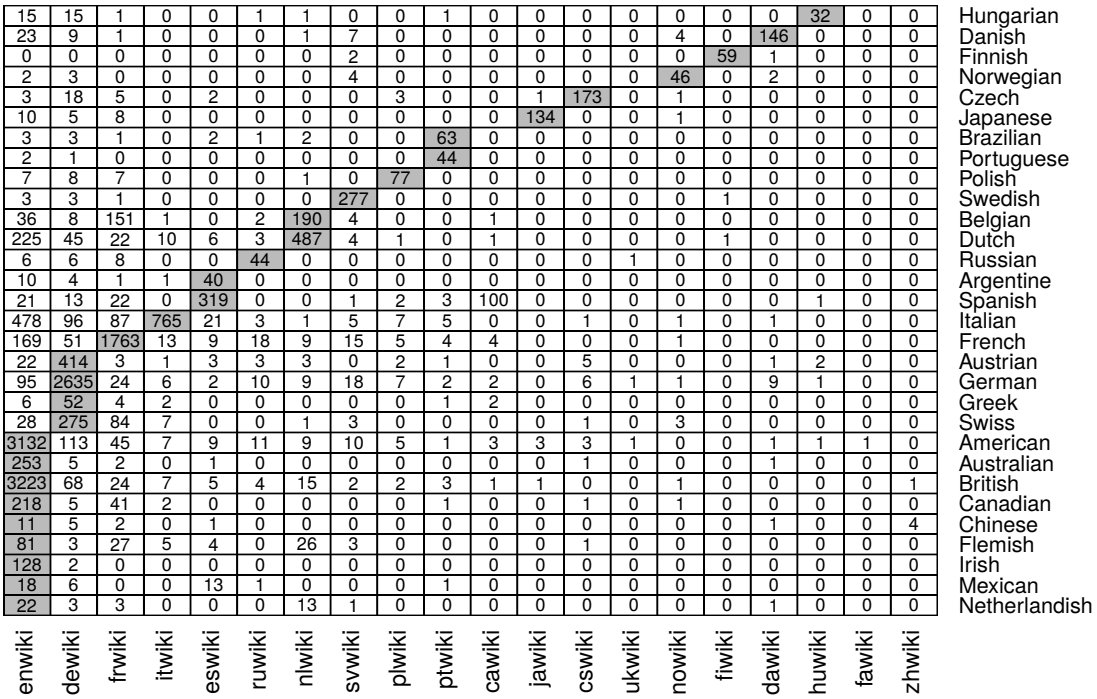


Figure 5.7: Coverage of ULAN nationalities of persons featured in only one Wiki language

5.4.2 Coverage of roles across languages

While the distributions of nationalities of covered ULAN records across Wikipedia language versions showed culturally specific biases, the distribution of roles revealed a different but related aspect. As shown in Figure 5.8, artists always represented the largest proportion of covered ULAN biographies in all observed Wikipedia languages but their overall proportion visibly decreased for Wikipedia versions covering smaller numbers of ULAN biographies. The smaller the coverage, the distribution of roles of covered ULAN persons appeared to shift towards other occupations such as emperors and kings, the nobility and the cleric, but also to other roles from the humanities and the arts, such as poets and philosophers.

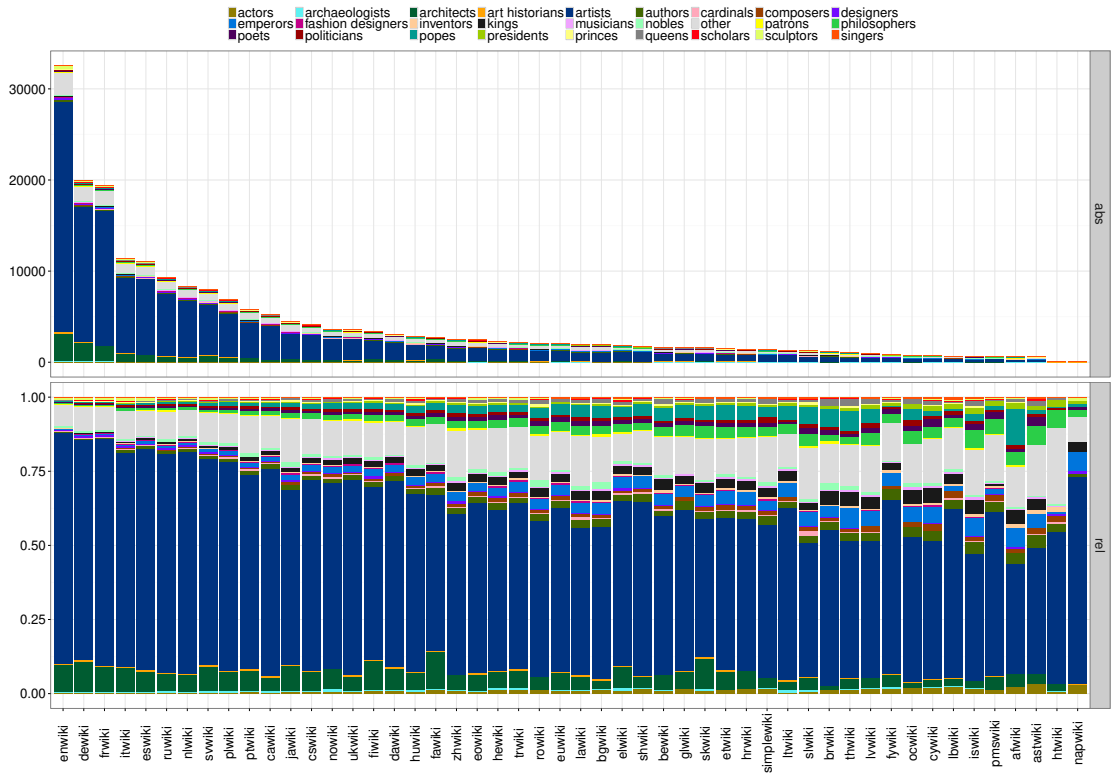


Figure 5.8: Role distributions across Wikipedia languages

This observation suggested that, as far as ULAN mapped biographies were concerned, the most important historical key figures present in the ULAN, which included the most famous artist "superstars" and important persons having the previously mentioned roles, tended to have high coverage across multiple Wikipedia versions, while many other ULAN artists appeared to be more of local relevance. One way to quantify this assumption was to calculate the mean and median number of Wikipedia languages covering the ULAN biographies found in each language version. As shown in Figure 5.9, this revealed a clear correlation where language versions with decreasing overall coverage of ULAN biographies tended to feature an increasing "prominence" of the covered biographies across multiple language versions in turn.

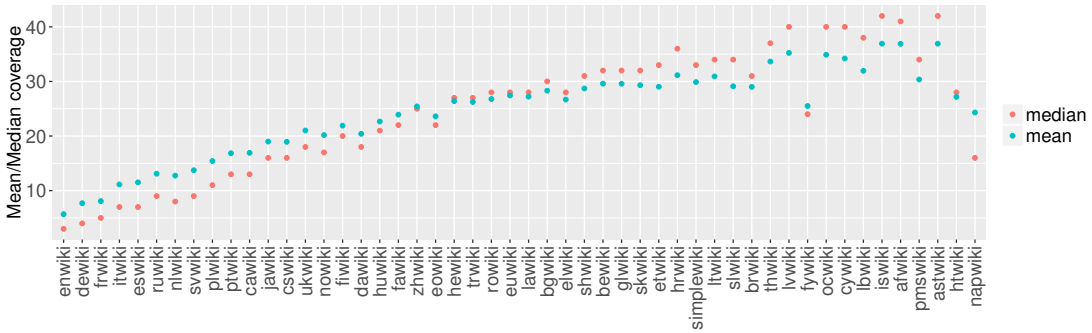


Figure 5.9: Mean/median coverage of covered ULAN bios in 50 Wikipedia languages

Table 5.1 shows the eight persons covered by all 50 and the 35 covered by 49 of the observed Wikipedia languages. In line with the previous findings, only 15 of them were assigned with the preferred role "artist" in the ULAN. Interestingly, however, only seven of these 15 would be recognized as fine artists by primary occupation, they are highlighted in the table and without doubt were amongst the most famous personalities from the domain. The remaining eight "artists" were mainly famous for other roles they played in history. For some of them, however, it is known that they were also active in the field of the fine arts, such as Sarah Bernhardt's parallel career as fine artist or the many drawings by Victor Hugo, while Walt Disney would probably be put under the term "creative industries". It was interesting to see Galileo Galilei, the famous scientist, to be listed as artist, since this topic was connected to a controversy regarding forged documents used as proof for the artistic aspect of his work[19]. Four other persons, Adolf Hitler, Winston Churchill, Benjamin Franklin and Nelson Mandela were listed as artists too, although they are mainly known for their political role in history. Hitler's ambition to become an artist and his rejection from the Academy of Fine Arts in Vienna, however, has been discussed in light of his later deeds while Winston Churchill was active as hobby painter[20], Mandela published a series of Lithographs[21] and Franklin is said to have published the first political cartoon in a newspaper[22]. It was interesting to see that these persons were assigned with the preferred role artist, which was most likely due to the intended function of the ULAN serving as database for museum documentation purposes and not as a general purpose knowledge base.

[19] http://www.zeit.de/2014/01/faelschung-zeichnungen-galileo-galilei-horst-bredekamp, retrieved Sept. 14th, 2020

[20] https://www.nationalchurchillmuseum.org/the-artist-winston-churchill.html, retrieved Sept. 14th, 2020

[21] http://www.nelsonmandelaart.com/, retrieved Sept. 14th, 2020

[22] http://www.historyofinformation.com/expanded.php?id=3394, retrieved Sept. 14th, 2020

label	role	nation	langs
Lincoln, Abraham	presidents	American	50
Confucius	philosophers	Chinese	50
Napoleon I, Emperor of the French	emperors	French	50
Gogh, Vincent van	artists	Dutch	50
Hitler, Adolf	artists	German	50
Leonardo da Vinci	artists	Italian	50
Mandela, Nelson	artists	South African	50
Chaplin, Charlie	actors	British	50
Luther, Martin	theologians	German	49
Elizabeth II, Queen of Great Britain	queens	British	49
Freud, Sigmund	psychiatrists	Austrian	49
Kennedy, John F.	presidents	American	49
Washington, George	presidents	American	49
John Paul II, Pope	popes	Polish	49
Lenin, V. I.	politicians	Russian	49
Alighieri, Dante	poets	Italian	49
Newton, Isaac	physicists	British	49
Einstein, Albert	physicists	German	49
Descartes, René	philosophers	French	49
Marx, Karl	philosophers	German	49
Leibniz, Gottfried Wilhelm	philosophers	German	49
Aristotle	philosophers	Greek	49

label	role	nation	langs
Plato	philosophers	Greek	49
Archimedes	inventors	Greek	49
Columbus, Christopher	explorers	Spanish	49
Watt, James	engineers	British	49
Mozart, Wolfgang Amadeus	composers	Austrian	49
Beethoven, Ludwig van	composers	German	49
Shakespeare, William	authors	British	49
Gagarin, Yuri Alekseyevich	astronauts	Russian	49
Franklin, Benjamin	artists	American	49
Warhol, Andy	artists	American	49
Disney, Walt	artists	American	49
Churchill, Winston	artists	British	49
Hugo, Victor	artists	French	49
Bernhardt, Sarah	artists	French	49
Caravaggio, Michelangelo Merisi da	artists	Italian	49
Michelangelo Buonarroti	artists	Italian	49
Galilei, Galileo	artists	Italian	49
Picasso, Pablo	artists	Spanish	49
Dalí, Salvador	artists	Spanish	49
Monroe, Marilyn	actors	American	49
Dietrich, Marlene	actors	German	49

Table 5.1: Person records covered in 49 and 50 Wikipedia languages

5.5 Analysis of the network of ULAN biographies in Wikipedia

The analysis of the ULAN biography coverage in different Wikipedia languages confirmed the existence of cultural preferences for biographies from a related cultural context, but also suggested a core group of persons which tended to be present in most of the observed language versions. It was thus expected to find similar effects in the network structures of the hyperlinks between Wikipedia biographies about ULAN persons. Moreover, it was of general interest to study the structure of the biographical Wikipedia networks in more detail and to compare them with the network of associative links from the ULAN.

5.5.1 Connected components

Table 5.2 shows, for the combination of all languages and each version separately (ordered by coverage), the distribution of component sizes for the DBpedia-hyperlink networks of Wikipedia biographies mapped to the ULAN via Wikidata. Components of size one represent biographies without any in- or outgoing links from or to other biographies and the column "gcc" expresses the size of the giant component in each network. In contrast to the associations analyzed directly in the ULAN which featured a variety of components of varying size, biographies in the Wikipedia hyperlink networks were either not interlinked at all, or they were part of the giant component, there were only very few separated components besides these two "extremes". A notable observation was that in 38 of the 50 observed languages and also in the combined network of all of them, biographies without hyperlinks to others tended to be more unique than biographies having links to others. In order to distinguish between the giant component in the ULAN network and the giant component in the combined Wikipedia network or individual language versions, the former will be referred to as the ULAN Giant Connected Component (UGCC) and the latter as the Wikipedia Giant Connected Component (WGCC), respectively.

In 40 language versions, the majority of covered ULAN biographies was part of the language's WGCC, the 10 language versions where this was not the case are highlighted in Table 5.2. Eight out of them were amongst the Wikipedias with the lowest coverage of ULAN person records, with the Haitian (htwiki) language version especially standing out, since the "network" of its covered ULAN biographies consisted of one single hyperlink between Aristotle and Plato. This observation went in line with the previous finding that the smaller the coverage of ULAN records in a Wikipedia, the more general historical key figures seemed to be covered, at least as far as the distribution of preferred ULAN roles was taken as an indicator in this regard.

set†	1	2	3	4	5	6	7	8	9	10	11	13	15	31	wgcc	set†	1	2	3	4	5	6	7	8	9	10	11	13	15	31	wgcc
fullwiki	6546	136	16	3	1	1				1					40564	**lawiki***	1013	66	8	5		1									742
enwiki	5833	139	20	8	4	2				1					26295	bgwiki	459	26	2												1415
dewiki	3249	111	16	2	2	1									16414	elwiki	595	21	7	2	1								1		1170
frwiki	3533	105	15	2	2										15612	shwiki	494	22	3	2											1145
itwiki	1866	53	12	3											9374	bewiki	480	38	5			1									1005
eswiki	1993	66	12	2	3	1		2							8826	glwiki	397	16	1					1							1132
ruwiki	1786	40	13	2											7387	skwiki	453	22	8		2										1041
nlwiki	1564	53	6	2			1								6599	etwiki	477	14	1	2				1	1						939
svwiki	2448	121	23	1	1			1							5205	hrwiki	335	4													1098
plwiki	1584	63	12	9	2		1								5057	simplewiki	609	20	9								1				700
ptwiki	1637	61	14	2											3999	ltwiki	403	18	4	3											858
cawiki	1334	41	4	4		1									3770	slwiki	524	17	4	1	1										658
jawiki	785	29	4		2	1									3589	brwiki	424	30	4	2			1	1							678
cswiki	964	36	11											1	3009	thwiki	282	21	5	3	2										702
ukwiki	881	30	12	3		1									2587	lvwiki	320	21	6												611
nowiki	1037	35	11	5	1	1			1			1			2444	fywiki	247	23	4		2										498
fiwiki	1166	45	11	1											2131	**ocwiki***	359	24	3	4											334
dawiki	846	41	5	2											2083	**cywiki***	374	19	6	4	2			1							287
huwiki	578	24	3	3											2225	**lbwiki***	335	29	11	2		1									243
fawiki*	1473	32	9	2				1	1						1157	**iswiki***	310	18	6	1	2	1	1	1		1		1			207
zhwiki	810	43	10	3	1										1585	afwiki	221	13	2	1											343
eowiki	697	29	7		1	2									1663	**pmswiki***	357	20	6		3		1								178
hewiki	423	12	5	1											1794	**astwiki***	438	10	2	2	1			2							97
trwiki	694	33	9	3											1365	**htwiki***	115														2
rowiki	621	19	4	1											1421	**napwiki***	53	4	1		1	1		1							10
euwiki	762	29	5	4	2										1203																

†) Ordered by total number of ULAN-mapped biographies, *) WGCC contains < 50% of ULAN-mapped biographies

Table 5.2: Components in hyperlink networks of ULAN-mapped Wikipedia biographies

5.5.2 Overlap between the ULAN and the Wikipedia networks

As stated in Section 5.3, 47,463 (24.49%) of the 193,823 person records in the ULAN were covered by at least one Wikipedia language and 18,891 (39.8%) of the 47,463 were unique contributions present in only one of the 50 observed languages. Separating between ULAN biographies with and without links in the original ULAN network revealed that 34,766 (20.23%) of the 171,881 ULAN biographies without links were present in at least one Wikipedia language, and 15,675 (45.09%) of them were unique contributions. The coverage of ULAN biographies having associative links was considerably higher than for those without links. 12,697 of the 21,942 ULAN biographies (57.87%) with such links were covered by the mapping via Wikidata and had a corresponding Wikipedia article in at least one language. Of these 12,697 biographies, 3,216 (25.33%) were unique contributions in Wikipedia, a clearly lower fraction than in the overall coverage. Looking at the 10,444 biographies from the UGCC in Wikipedia, their coverage was even higher than for all ULAN biographies with associative links. 7,256 (69.48%) of them were

represented in at least one language version of the free encyclopedia, and most of them were featured in more than one, since only 1,525 of the 7,256 (21.02%) contributions were unique.

As shown in Table 5.2, 40,943 (86.26%) of all 47,463 ULAN-mapped Wikipedia biographies had hyperlinks to other ULAN-mapped biographies in the combined Wikipedia data. Counting ULAN records simultaneously having associative links in the ULAN and hyperlinks in Wikipedia, this was the case for 12,239 (96.39%) of the 12,697 covered records with associative links in the ULAN and for 7,097 (97.81%) of the those covered 7,256 which were part of the UGCC. Since in Wikipedia there were very few components besides the WGCC, its coverage was almost the same: 7,088 (97.68%) of the 7,256 covered UGCC records were also part of the WGCC. Of the 6,520 mapped Wikipedia biographies without hyperlinks, 6,062 (92.98%) also had no associative links in the ULAN.

These findings could be interpreted as such that the more important a person featured in the ULAN was in the global canon of art history — Based on the assumption that persons with associative links in the ULAN were, overall, more important than those without, and that those in the UGCC were usually more important than interlinked persons from smaller components — the more likely he or she was covered by one of the Wikipedia languages and with growing importance, also more likely to be covered by more than one Wikipedia language. Similarly, although less pronounced, the more important a person from the ULAN was, the more likely his or her respective Wikipedia biography would have hyperlinks to or from other ULAN-mapped biographies.

Given the incomplete coverage of linked ULAN records in Wikipedia, it was of interest to check how many of the 178 famous artists from the scholarly rankings used in Section 4.3.3, shown in the Appendix in Table A.4, were represented there. As it turned out, 174 (97.75%) of the 178 persons were present in the WGCC, which was clearly higher than the UGCC's overlap of 136 (76.4%) and an indicator that the particularly outstanding representatives of art history tended to be represented in Wikipedia as well. Compared to the findings of Samoilenko and Yasseri in [Samoilenko and Yasseri, 2014] regarding the representativeness of Wikipedia biographies of living scientists, however, the selection of the 178 persons referred to historical persons only and not to living ones. An analysis of how much the Wikipedia presence of contemporary artists reflected their "real life" standing in the art world would thus be an interesting topic for future work.

Of the 174 covered persons only one, the French 19th century painter Théophile Auguste Vauchelet, was a unique contribution of the French Wikipedia, the remaining 173 persons were at least featured in four different languages, three quarters of them in more than 25, half of them in more than 35 languages. Of the four persons not found in the Wikipedia data mapped through Wikidata, one person, Harry Millson Hunt, could not be found in Wikipedia/Wikidata at all, while two persons, Hans Stethaimer the younger and Jan Boudolf, were actually present Wikipedia (The former only in the German, the latter in multiple languages) but not mapped to the ULAN in Wikidata. The last of the four, Tommaso Barlacchi, was present in Wikidata but without a corresponding article in Wikipedia. The latter was not present in the current dataset because all ULAN-Wikidata mappings without corresponding Wikipedia articles were discarded at the beginning.

Overlap between ULAN associative links and Wikpedia hyperlinks

Besides checking for the person-wise overlap between the ULAN and the Wikipedia networks, the overlap of the individual connections between them was another important aspect. In order to analyze which of the associative ULAN links had corresponding hyperlinks in Wikipedia, a subset of ULAN associative links had to be identified where every two connected ULAN records each had a corresponding Wikipedia article in one or more languages. This was found to be the case for 23,730 (47.39%) of the 50,076 unique ULAN links, connecting 11,005 persons with each other. 15,747 (66.36%) of these links were directly covered by Wikipedia hyperlinks and Figure 5.10 shows their distribution by ULAN link type, juxtaposing the full count of ULAN links for each type in the observed subset, represented as dark grey bar on the left side, with the count of overlapping Wikipedia hyperlinks to its right. The bars representing the latter are additionally subdivided into segments, representing the fractions of links which were covered by multiple Wikipedia versions (light grey) and those which were uniquely present in only one language version (colored by language).

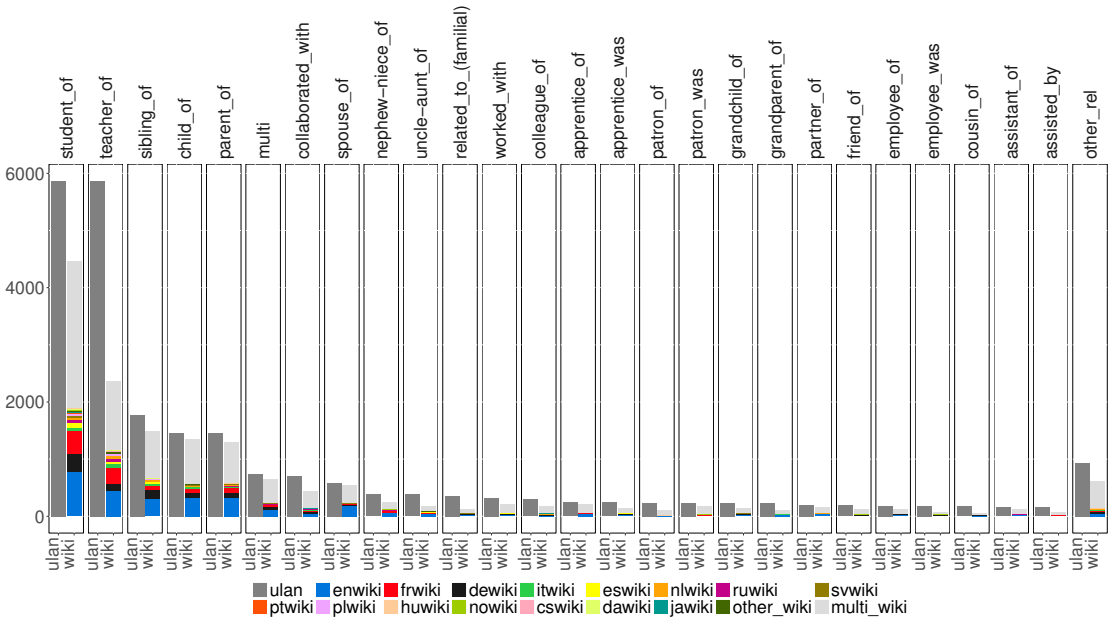


Figure 5.10: Coverage of ULAN links by Wikipedia hyperlinks

The plot reveals a number of interesting aspects which were encountered for the coverage of ULAN associative links in Wikipedia. The most immediate insight was that student and teacher relationships, which were always mirrored with each other in the ULAN, had a clear asymmetry in Wikipedia, with significantly more covered links pointing from students to teachers than vice versa. This was an interesting observation which suggested that teacher biographies did not feature links to all of their students,

whose own biographies nevertheless very well tended to link to their teachers in turn. In contrast to teaching, parent and child relationships were in turn highly mirrored in Wikipedia and together with other family relationships such as sibling and spouse also had a far higher coverage there. ULAN links which were simultaneously assigned with different relationship types ("multi") had high coverage as well.

Table 5.3 shows the identified overlap between ULAN associative links and Wikipedia hyperlinks by numbers for high-level link type, confirming that family relationships and links with multiple types had highest coverage. In contrast to the "single-type" links, those having multiple types assigned were represented to a high degree (86.83%). As shown in Table 4.7, many of these links included family ties of type parent, child, sibling and spouse, which already had a high coverage of 89.13%, 92.37%, 83.80% and 92.98% in the original Wikipedia network and thus were likely to contribute there. The lower overall coverage of family relationships (77.62%) could in turn be explained by non-immediate family ties such as nephew/niece (63.23%), grandchild (59.65%) or cousin (33.91%).

Since ULAN associative links were always mirrored, it was of interest to compare them to a similarly mirrored set of Wikipedia hyperlinks where each link from A to B without an existing inverse tie was augmented with its counterpart. This was based on the assumption that the presence of an unidirectional relationship in Wikipedia also implied the knowledge about its inverse. The resulting mirrored Wikipedia network featured a significantly higher coverage of ULAN links across all types, shown in the lower part of Table 5.3.

	teach.-learn.	**family**	**professional**	**multi**	**patronage**	**personal**	**other**
Wikipedia	58.87%	77.62%	65.91%	86.83%	59.82%	62.9%	77.56
	7331	5755	1247	646	530	117	121
ULAN	12452	7414	1892	744	886	186	156
	9816	6184	1498	692	694	142	128
Wikipedia mirrored	78.83%	83.41%	79.18%	93.01%	78.33%	76.34%	82.05

Table 5.3: Overlap of ULAN associative links with Wikipedia hyperlinks

Table 5.4 shows the 16 ULAN links covered by more than 43 Wikipedia languages. The tendency of non-artist biographies to be covered in more languages became also visible in the hyperlinks, where those between famous figures from antiquity, nobility and the cleric were more likely to be present in multiple Wikipedias. Table 5.5 therefore shows the subset of ULAN links between persons of type "artist", covered by more than 30 languages and representing relationships between famous artist figures across many relevant periods. It was again interesting to see that the most replicated links were between persons not immediately perceived as primary representatives for the fine arts, since John Lennon and the British authors and siblings Charlotte and Emily Brontë were mainly known for other achievements.

sourcelab	sourcerole	reltype	targetlab	targetrole	coverage
Aristotle	philosophers	student of	Plato	philosophers	48
Plato	philosophers	student of	Socrates	philosophers	47
Tiberius, Emperor of Rome	emperors	adoptive child of	Augustus, Emperor of Rome	emperors	47
Charles, Prince of Wales	princes	child of	Elizabeth II, Queen of Great Britain	queens	46
Plato	philosophers	teacher of	Aristotle	philosophers	46
Nero, Emperor of Rome	emperors	adoptive child of	Claudius, Emperor of Rome	emperors	46
Claudius, Emperor of Rome	emperors	adoptive parent of	Nero, Emperor of Rome	emperors	46
Socrates	philosophers	teacher of	Plato	philosophers	45
Augustus, Emperor of Rome	emperors	adoptive parent of	Tiberius, Emperor of Rome	emperors	45
Philip II, King of Spain	kings	child of	Charles V, Holy Roman Emperor	emperors	45
Antoninus Pius, Emperor of Rome	emperors	adoptive child of	Hadrian, Emperor of Rome	emperors	45
Elizabeth II, Queen of Great Britain	queens	parent of	Charles, Prince of Wales	princes	44
Charles, Prince of Wales	princes	spouse of	Diana, Princess of Wales	princesses	44
Victoria, Queen of Great Britain	queens	parent of	Edward VII, King of Great Britain	kings	44
John Paul I, Pope	popes	successor of	Paul VI, Pope	popes	44
Hadrian, Emperor of Rome	emperors	successor of	Trajan, Emperor of Rome	emperors	44

Table 5.4: ULAN links covered by hyperlinks in 44+ Wikipedia languages

sourcelab	sourcerole	reltype	targetlab	targetrole	coverage
Ono, Yoko	artists	spouse of	Lennon, John	artists	41
Brontë, Charlotte	artists	sibling of	Brontë, Emily	artists	39
Lennon, John	artists	spouse of	Ono, Yoko	artists	39
Brontë, Emily	artists	sibling of	Brontë, Charlotte	artists	38
Dyck, Sir Anthony van	artists	student of	Rubens, Peter Paul	artists	38
Verrocchio, Andrea del	artists	teacher of	Leonardo da Vinci	artists	37
Braque, Georges	artists	collaborated with	Picasso, Pablo	artists	37
Rivera, Diego	artists	spouse of	Kahlo, Frida	artists	37
Ingres, Jean-Auguste-Dominique	artists	student of	David, Jacques-Louis	artists	36
Dalí, Salvador	artists	collaborated with	Buñuel, Luis	artists	36
Kahlo, Frida	artists	spouse of	Rivera, Diego	artists	36
Picasso, Pablo	artists	collaborated with	Braque, Georges	artists	35
Buñuel, Luis	artists	collaborated with	Dalí, Salvador	artists	35
Leonardo da Vinci	artists	student of	Verrocchio, Andrea del	artists	34
Vasari, Giorgio	artists	student of	Michelangelo Buonarroti	artists	33
Cimabue	artists	teacher of	Giotto	artists	33
Dalí, Salvador	artists	associate of	Breton, André	artists	32
Giotto	artists	student of	Cimabue	artists	32
Bruegel, Pieter	artists	child of	Bruegel, Pieter	artists	31
Renoir, Jean	artists	child of	Renoir, Auguste	artists	31
Michelangelo Buonarroti	artists	apprentice of	Ghirlandaio, Domenico	artists	31
Bellini, Giovanni	artists	teacher of	Giorgione	artists	31
Bellini, Giovanni	artists	teacher of	Titian	artists	31

Table 5.5: ULAN links between artists, covered in 31+ Wikipedia languages

While the immediate 66.36% coverage of ULAN associative links in Wikipedia was already quite high, the significantly increased 80.72% link coverage found in the mirrored Wikipedia hyperlink network suggested that for those ULAN records which had corresponding Wikipedia biographies, the relational art history knowledge present in the ULAN was also represented in Wikipedia to a high degree. Since the Wikipedia hyperlink network was very dense compared to the much more sparse ULAN relationships, this left the question if the persons connected by the remaining 4,576 ULAN associative links for which no immediate overlap could be found in Wikipedia were probably nevertheless connected via paths there instead.

Alternative paths for ULAN links not covered in the Wikipedia network of mapped ULAN biographies

For the observed 11,005 persons whose 23,730 ULAN associative links overlapped with 15,747 (19,154 mirrored) hyperlinks in Wikipedia, there were 61,639 (103,418 mirrored) additional unique ties between ULAN-mapped Wikipedia biographies which had no ULAN counterparts. Assuming the mirrored case, the unique set of hyperlinks extracted from all Wikipedia languages combined thus featured about 5.17 as many (103,418 + 19,154 vs 23,730) unique inter-person ties than present in the ULAN for the same subset of persons. The presence of these additional links suggested that the 4,576 ULAN links which were not directly represented in the mirrored Wikipedia network might nevertheless be represented via alternative paths there instead. It was thus of interest to seek for such alternative paths and check if they varied in length for the different categories of the links without coverage.

Table 5.6 shows the distribution of identified alternative path lengths in the combined Wikipedia network with all mirrored links for high-level link type categories. On the left side of the table, multiple shortest alternative paths of similar length were counted only once for each ULAN link without coverage, enabling to see how many of the latter had at least one alternative of a specific length. The center of the table represents the 688 ULAN links for which not even an alternative path could be identified. The majority (666) of them had at least one endpoint which was not interlinked at all with any other ULAN biography in Wikipedia. The right side of the table shows all 25,468 available shortest alternative link paths for the 4,576 ULAN links without coverage, revealing that many of them had multiple shortest paths of the same length.

	2	**3**	**4**	**5**	**6**	**N/A**	**2**	**3**	**4**	**5**	**6**
other	60	12	20	6	0	26	100	58	258	14	0
professional	192	110	34	14	2	42	468	854	390	134	8
family	756	134	70	26	4	240	1336	528	534	402	236
patronage	126	38	8	2	2	16	616	638	90	10	2
teaching	926	896	366	74	10	364	1784	6084	8162	2358	176

Table 5.6: Lengths of alternative Wikipedia paths for ULAN associative links

Overall, alternative Wikipedia paths could be found for between 79.03% (links of type "other") and 91.67% (patronage ties) of the ULAN links without coverage. The 80.72% overlap with the mirrored and combined Wikipedia network already increased to 89.39% when adding shortest alternatives of length two and to 94.41% including those of length three. Regarding the two largest link categories, it was interesting to observe that there were length-two alternative paths for 756 (61.46%) out of the 1,230 family ties, but only for 926 (35.13%) of the 2,636 teaching relationships. This suggested that missing family ties were most likely bridged by mutual ties to immediate family members, while alternatives for teaching relationships tended to have more complicated routes.

Wikipedia links not covered in the ULAN network

The very high connectedness of ULAN biographies in Wikipedia made it possible to find alternative paths for the majority of the uncovered ULAN associative links in the combined Wikipedia network of all languages. It was therefore also of interest to see how far this was the case the other way round, if and how many of the Wikipedia hyperlinks which were not present in the original ULAN network had alternative paths there. Table 5.7 shows a selection of the 267,621 Wikipedia hyperlinks from the combined multilingual network which, although present in at least 30 Wikipedia versions, were not covered by associative relationships in the original ULAN network, limited to ULAN persons assigned with the role artist. It was interesting to find connections between very well known Renaissance or Modern artists and that many of them were alive at the same time, as seen in the column "bdiff", representing their birth date difference. Although not present in the ULAN, many of these links indeed turned out to be well documented direct contacts between the featured persons, such as the friendship between Van Gogh and Gauguin or the previously mentioned contact between Albrecht Dürer and Giovanni Bellini (Covered by 23 languages and not visible in the Table). Other Wikipedia links not covered in the ULAN, however, appeared to be more based on influence or mere contemporaneity, which were stated to be out of the scope for recording associative links in the ULAN. In this regard it became evident that the birth date difference of influential relationships often exceeded the lifetime of a person, as for example visible in the Table for the ties between Rubens and Titian or Manet and Velázquez.

For the two mentioned influential relationships, however, alternative shortest paths existed in the ULAN network, one of length three for the former, one of length 17 for the latter, representing a succession of persons spanning the 233 years between Manet and Velázquez. Moreover, there were alternatives for some of the uncovered ties with documented contact too, such as one of length eight for the teaching relationship between the Austrian artists Gustav Klimt and Egon Schiele, which appeared quite unusual given their close context. For other uncovered ties in the Table, the alternatives were much shorter, often relating two persons interlinked in Wikipedia, but not in the ULAN, via two associative links in the latter. The length of the shortest path is shown in Table 5.7 in the column "shortest" and as shown previously in the discussion of the ULAN network, some well known artists were located in different network components and there was thus no possible alternative path for Wikipedia hyperlinks between them, which is shown in the column as "Inf".

source	sbirth	snation	target	tbirth	tnation	bdiff	coverage	shortest
Raphael	1483	Italian	Michelangelo Buonarroti	1475	Italian	-8	42	2
Raphael	1483	Italian	Leonardo da Vinci	1452	Italian	-31	42	2
Dalí, Salvador	1904	Spanish	Picasso, Pablo	1881	Spanish	-23	41	4
Gogh, Vincent van	1853	Dutch	Gauguin, Paul	1848	French	-5	37	3
Renoir, Auguste	1841	French	Monet, Claude	1840	French	-1	37	2
Schiele, Egon	1890	Austrian	Klimt, Gustav	1862	Austrian	-28	36	8
Gauguin, Paul	1848	French	Gogh, Vincent van	1853	Dutch	5	36	3
Michelangelo Buonarroti	1475	Italian	Leonardo da Vinci	1452	Italian	-23	36	3
Sisley, Alfred	1839	French	Monet, Claude	1840	French	1	35	2
Pissarro, Camille	1831	French	Cézanne, Paul	1839	French	8	34	Inf
Cézanne, Paul	1839	French	Zola, Emile	1840	French	1	34	Inf
Monet, Claude	1840	French	Renoir, Auguste	1841	French	1	33	2
Signac, Paul	1863	French	Seurat, Georges	1859	French	-4	33	Inf
Leonardo da Vinci	1452	Italian	Michelangelo Buonarroti	1475	Italian	23	33	3
Gris, Juan	1887	Spanish	Picasso, Pablo	1881	Spanish	-6	33	Inf
Klee, Paul	1879	Swiss	Kandinsky, Vasily.	1866	Russian	-13	33	2
Cassatt, Mary	1844	American	Degas, Edgar	1834	French	-10	32	5
Rubens, Peter Paul	1577	Flemish	Titian	1483	Italian	-94	32	3
Morisot, Berthe	1841	French	Manet, Edouard	1832	French	-9	32	Inf
Giorgione	1477	Italian	Titian	1483	Italian	6	32	2
Tintoretto	1519	Italian	Titian	1483	Italian	-36	32	Inf
Titian	1483	Italian	Giorgione	1477	Italian	-6	32	2
Manet, Edouard	1832	French	Velázquez, Diego	1599	Spanish	-233	31	17

Table 5.7: Wikipedia links not present in the ULAN

Overall, it was possible to find alternative ULAN link paths for 35,387 of the 267,621 Wikipedia hyperlinks without direct ULAN counterparts, in the mirrored case for 58,910 of 455,124. Figure 5.11 shows the counts for the mirrored case, revealing that although the majority of them were shorter, alternative ULAN path lengths could reach 30 links. It was interesting to observe that although alternatives of length two occurred most often, there was a sudden change between them and those of length 3, from where on the counts increased linearly until reaching alternatives of length six which had the second highest count. From then on, there appeared to be an exponential decrease in the counts. Recalling the many small components existing in the ULAN network besides its GCC suggested them to be the reason for the encountered distribution.

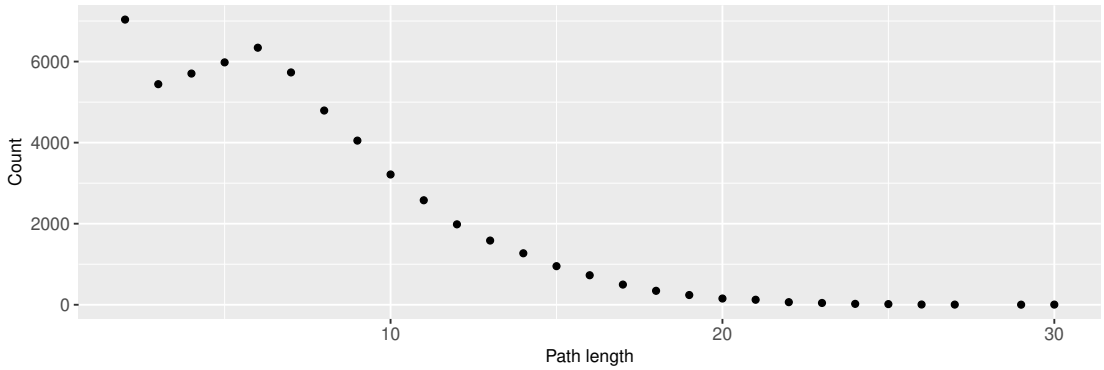


Figure 5.11: Alternative ULAN paths for uncovered Wikipedia hyperlinks

WGA-ULAN portrait relationships covered in Wikipedia

Another approach to comparing the coverage of the ULAN associative links and Wikipedia hyperlinks was to use a "third party" data source in form of the portrait relationships identified between creators of WGA pictures and persons depicted there, discussed in Section 3.5 (Artifact **A2**, Figure 3.8). While the ULAN associative relationships covered only 38 (10.08%) out of 377 unique creator-subject pairs of persons, Wikipedia hyperlinks had a far higher coverage, featuring links for 240 (63.66%) of the 377 creator-subject pairs. 208 of these pairs were not represented by ULAN associative relationships, which in turn only covered six unique pairs which were not represented in form of Wikipedia hyperlinks as well.

Table 5.8 shows that the Wikipedia coverage was also much higher for creator-subject pairs which appeared only once, while it missed one of the two pairs for which four portraits could be identified. As it turned out, this missing four-fold relationship between the painter Anthonis van Dyck and King Charles I. of England was very well present in van Dyck's Wikipedia biography but was not identified due to an ambiguous ULAN-Wikidata mapping for Charles I. Another interesting observation was that about a quarter (54) of the creator-subject pairs additionally covered in Wikipedia represented relationships where the involved persons were born more than 75 years apart and thus did not have associative relationships in the ULAN, which included many (imaginary) portraits of persons from antiquity which were painted during or after the Renaissance. The remaining 154 of the additionally covered creator-subject pairs were contemporaries, which underscored that Wikipedia contained art history knowledge that was not represented in the ULAN via its associative links.

Num. occ.	4	3	2	1
Complete	2	11	31	333
Matched ULAN	2	4	5	27
M/C	1.00	0.36	0.16	0.08
Matched Wikipedia	1*	11	19	209
M/C	0.5*	1.0	0.61	0.63
Bdiff > 75y	0	1	6	47

* Due to ambiguous Wikidata-ULAN mapping

Table 5.8: Recurring WGA portrait creator/subject pairs covered in ULAN and Wikipedia relationships

5.5.3 Analyzing the Wikipedia network via ULAN person attributes

As shown Subsection 5.5.2, there was a substantial overlap between the subset of the ULAN associative link network whose persons were also covered in Wikipedia and the respective hyperlink network there, especially when considering the latter irrespective of the links' direction. It was therefore assumed that for covered biographies, most of

the knowledge represented in the ULAN was also represented in the different Wikipedia hyperlink networks. Since the combined Wikipedia network was much more dense than its ULAN counterpart and featured a much larger GCC, it was of great interest to study it in more depth and to compare its overall structure with the one present in the ULAN. Moreover, a comparison of the individual networks present in the different language versions seemed promising as well.

In contrast to the ULAN associative links, however, Wikipedia hyperlinks were not explicitly typed in DBpedia since their semantics were embedded in Wikipedia's free text and couldn't be extracted in a straight-forward manner. Although lots of existing research has been dealing with related questions, it was out of the scope of this work to consider the underlying semantic types of hyperlinks by processing the respective Wikipedia text and they were analyzed based on the ULAN attributes assigned to the involved persons instead.

Assortative mixing in Wikipedia hyperlink networks

Like the associative links in the ULAN, Wikipedia hyperlinks tended to connect persons having similar attributes. This tendency, however, was less pronounced than in the ULAN network: About 59.95% of the hyperlinks in the Wikipedia network of all languages combined were links between persons of the same nationality (ULAN: 81.71%) and 67.65% linked persons of same role (ULAN: 88.35%). Using the previously defined cutoff of 75 years as dichotomous measure for birth date difference, 85.24% of the links connected persons whose birth dates were not more than 75 years apart (ULAN: 95.55%). Regarding links between persons of same gender, Wikipedia was quite equal to the ULAN (86.61% vs. 86.15%). Taking the underlying distributions of attribute values into account, the respective assortativity coefficients in the combined Wikipedia network yielded 0.554 for nationalities (ULAN: 0.8), 0.257 for role (ULAN: 0.53) and 0.155 for gender (ULAN: 0.16), while birth dates had a coefficient of 0.63 (ULAN: 0.97).

Nationality distributions of interlinked persons across language versions

As it could be expected from the language specific distributions of nationalities in the different Wikipedias, they had a strong effect on the distribution of nationalities in hyperlinks between the respective persons. The preference of language versions for links connecting culturally related persons could also be quantified via standardized Pearson residuals calculated as described in [Agresti, 2007, p. 38] via tabulation of nationality combinations against Wikipedia language versions, which took the individual sizes of the different Wikipedia versions better into account. Table 5.9 shows the results for 34 aggregations of same-nationality links in 12 Wikipedia languages. The remaining languages and same-nationality link aggregations were grouped into respective "other" categories. 11 of the featured languages had significant residual values for culturally related nationalities.

	enwiki	dewiki	frwiki	itwiki	eswiki	ruwiki	nlwiki	svwiki	plwiki	ptwiki	cawiki	jawiki	otherwiki
American_American	**113.3**	-11.5	-40.8	-31.2	-21.2	-7.3	-15.3	-14.6	-8.8	-8.9	-16.6	-0.5	-10.1
Argentine_Argentine	1.7	-5.7	-6.3	-4.1	**33.7**	-2.1	-3.6	-1.8	-2.7	-1.6	-2.3	-2.5	-4
Australian_Australian	**40.4**	-7.3	-8.8	-8.5	-7.5	-5.7	-3.4	-4.1	-3.6	-4	-4.8	-4.2	-7.3
Austrian_Austrian	-18.5	**55.7**	-14.3	-9.6	-6.2	0.4	-6.1	1.5	2.3	-2.3	-3.9	-0.5	1.9
Belgian_Belgian	-18.7	-15.1	24.1	-16.8	-12	-7.7	**118.6**	-7.7	-9.1	-9.3	-9	-9	-14.4
Brazilian_Brazilian	-3.3	-5.3	-6.2	-6.2	2.6	-2.7	-4.1	-3.5	-4.3	**76.8**	-3.1	-3.5	-5.4
British_British	**118.2**	-28.4	-47.9	-28.6	-20	-8.2	-9.9	-1.6	-6.3	0.9	-17.9	2.7	-11
Canadian_Canadian	**26.8**	-9.6	9.7	-8.9	-7.1	-6.3	-6.3	-4.8	1.4	-3.6	-5.1	-4.2	-7.4
Chinese_Chinese	-10.9	-12.3	11.3	-10.6	-8.3	6.9	-8.8	-4.2	-4.5	-5.9	9.2	20.1	**28.5**
Czech_Czech	-11.4	-3.5	-10.8	-12.1	-9.9	6.5	-7	-5.9	-4.6	-5.9	-6.8	-6.3	**75.6**
Danish_Danish	0.3	-12.4	-19.1	-16.6	-12.2	-9.1	-5.2	21.6	-8.4	-6	-10.1	-8.8	**87.3**
Dutch_Dutch	5.1	-18.9	-24.5	-14.9	-12	-7.4	**134.1**	-1.8	8.6	-13.3	-8.8	-9	-9.8
Finnish_Finnish	-2.5	-5.1	-1.1	-5.6	-5.9	-1.6	-3.4	15.6	-3.6	-2.3	-4	-2.8	**25.5**
Flemish_Flemish	11.3	-15.1	-6	-5.2	7.8	-5	**37.1**	-6	10.1	-5.4	-6.9	-5.9	-9.9
French_French	-61.9	-38.5	**168**	-25.4	-9.3	2.2	-18.3	-2	-4	-2.5	-12.1	-4.1	-13.8
German_German	-51	**204.4**	-41.6	-36.5	-25	0.4	-18.5	5.2	-6.6	-8.9	-19.4	-10.4	-14.7
Greek_Greek	-13.8	1.3	-7.8	4.4	1.1	3.1	-3.2	0.5	2.4	9.4	8.2	0.8	**12.1**
Hungarian_Hungarian	-6.2	-1.9	-9.1	-10.8	-9.4	-4.4	-7.9	-5.7	-3.1	-5.9	-6.2	-5.9	**69.5**
Indian_Indian	**14.3**	-2.3	-4.6	-4.5	-2.8	-0.8	-2.8	0.7	-2.3	-2	-1.9	-1.3	0.7
Irish_Irish	**13**	-4.8	-4.8	-3.1	-2	2.9	-1.8	-1	-0.2	1.6	-2	-1	-2.5
Italian_Italian	-44.9	-47.2	-20.9	**163.1**	21.4	-10.4	-34.4	-7.8	-3.9	21.9	20	-10.8	-11.7
Japanese_Japanese	-11.3	-15.4	-1.6	-15.2	-8	-6	-11.7	-7.1	-6.4	-4.6	-2.4	**163.1**	-7.1
Mexican_Mexican	13.8	3.8	-11.2	-8.4	**25.5**	-4.8	-3.8	-3.1	-5.4	-3.8	-5.1	-3.5	-7.5
Netherl._Netherl.	-6.3	-3.2	-1.4	-4.3	0.5	-0.5	**13.3**	-0.6	10.8	4.7	2.7	1.4	-1.8
Norwegian_Norwegian	-4	-2.5	-11	-9.2	-3.6	0.8	-4.7	23.3	-4.7	-5.3	-5.8	-3.9	**37.9**
Polish_Polish	-10.4	5.7	-13.9	-13	-10	15.4	-5.5	-6.4	**90.1**	-5.4	-6	-5.6	0.4
Portuguese_Portuguese	-5.1	-4	-6.6	-5.7	-1.5	-2.1	-3.1	-1.3	5.1	**65.8**	-3	-1.1	-3.8
Roman_Roman	-13.6	-4.9	-6.8	0.5	1.3	3.3	2.4	4	3.7	**12.9**	7.7	8.6	8.3
Romanian_Romanian	-0.3	-4	-5.5	-4.9	-4.1	2.2	-2.3	-2.9	-3.7	-3.4	-3.7	-3.5	**31.7**
Russian_Russian	-14	-16.2	-6.2	-17.8	-11.2	**95.1**	-4.6	-7.6	1.1	-1.7	-8.1	-1.3	15.1
Spanish_Spanish	-38.8	-30.2	-21.8	-22.8	**126.2**	-14.4	-17.2	-11	18.1	-2.8	117.4	-11.5	-12
Swedish_Swedish	-12.7	-2.7	-15.9	-11.3	-10	-5.3	-3.8	**139.8**	-0.6	-4.5	-6.3	-0.4	-1.3
Swiss_Swiss	-11.8	**49.8**	0.2	-6.8	-7.7	-3.2	-6.2	-2.5	-4.3	-5.3	-6.5	-5.1	-6.6
other_other	9	-4.6	-12.5	-11.7	11.2	4.7	-7.3	-5.4	-5.2	-5.3	-5.8	-5.5	**27.4**

Table 5.9: Standardized Pearson residuals for same-nationality hyperlink counts by Wikipedia language

The higher presence of hyperlinks between language related nationalities was immediately tied to the different versions' higher coverage of the respective biographies and it was thus of interest to look at this aspect from a different point of view. One approach was to observe only biographies common to a number of Wikipedia languages, since the distribution of nationalities would be constant there and any language version still featuring higher link counts between specific nationalities would suggest that its individual biographies themselves provided more culturally related information. Figure 5.12 shows the aggregated link counts between the 15 most present ULAN nationalities for the set of 1,809 biographies present in each of the 12 language versions with the highest ULAN coverage. Link counts are shown in form of small multiple bar charts, one for each nationality combination, each bar representing the aggregated link count for that nationality combination in one of the 12 languages. In order to highlight languages with highest count, those bars having maximum counts for each combination have full

opacity. Even for the shared set of biographies, the majority of Wikipedia language versions showed higher counts for related same-nation links and in part also for related international ones, suggesting that it was indeed the case that biographies about persons which were culturally closer related to the respective language version featured richer content than those about less related ones, based on the assumption that the number of links to and from the respective biographies could serve as an indicator in this regard.

Figure 5.12: Aggregated hyperlink counts between nationalities for biographies common to 12 Wikipedia languages

Birth date differences between Wikipedia biographies

Since almost 15% of the hyperlinks in the combined Wikipedia network connected persons whose birthdates were more than 75 years apart, it was worth to have a deeper look at the distribution of their birth date differences in each language version. Figure 5.13 shows density plots for the combined network (label: "wiki") and for all language versions having at least 1,000 hyperlinks, sorted from top to bottom by ascending number of hyperlinks. The density plot for the ULAN network (label: "ulan") is shown below for comparison. Each density plot is divided into regions representing different quantiles specified in the legend on top of the Figure.

The temporal distributions of hyperlinks in the different Wikipedia language versions appeared remarkably similar to the ULAN at first sight, but were clearly skewed when

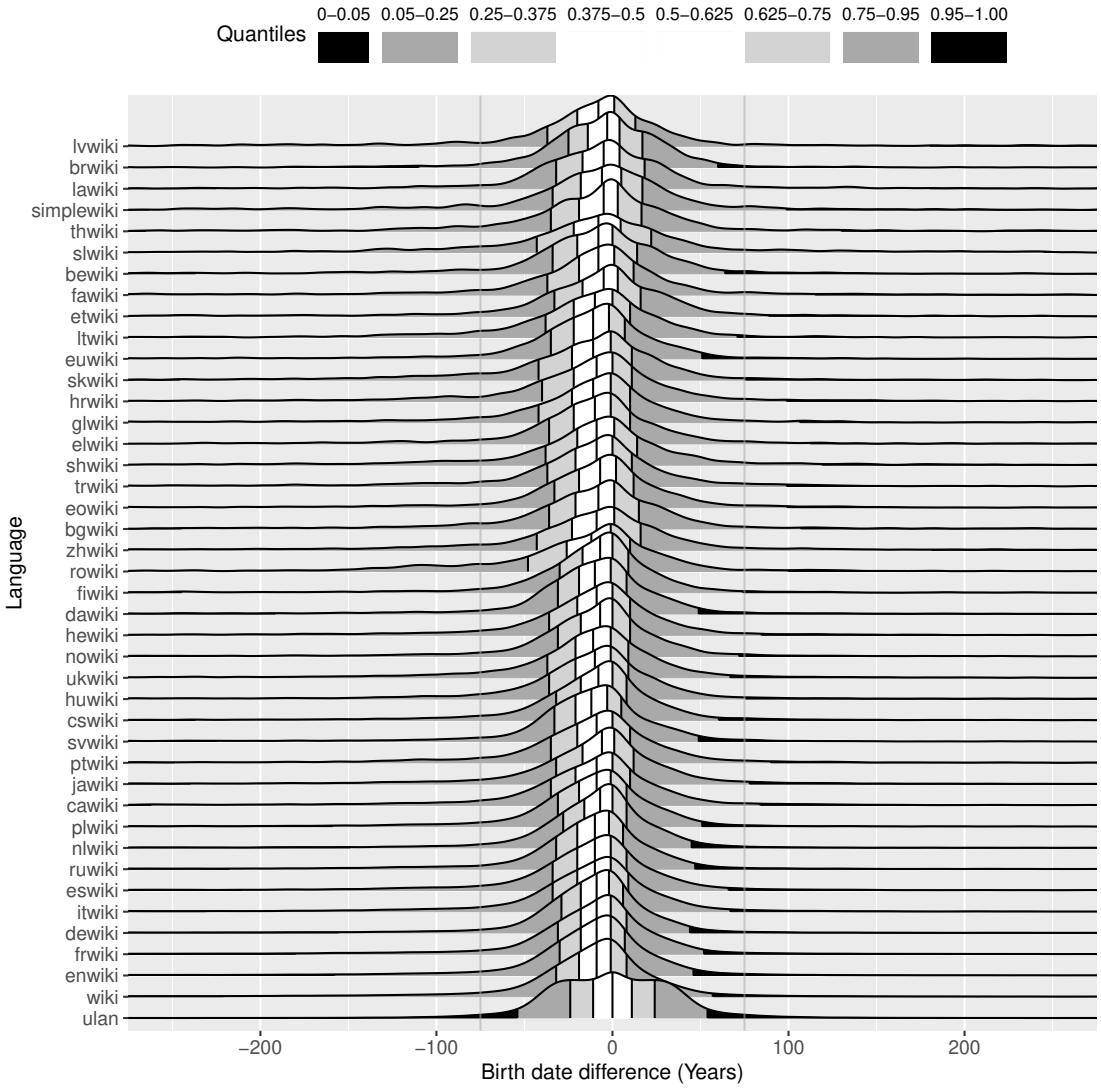


Figure 5.13: Birth date differences between linked persons in different Wikipedias

compared with the symmetry of the distribution of ULAN associative relationships. In contrast to the latter, the median birth date difference was below zero for all Wikipedia language versions, i.e. more hyperlinks pointed "into the past", from younger to older persons, than vice versa: The combined network had 63.43% of its links pointing to the past with a median of -9 years, the individual language versions ranged from 55.04% with median -3 to 66.34% with median -12 years. This coincided with the previous observation shown in Figure 5.10, where the overlap of teaching links for example was based on a much higher number of links from younger students to older teachers than vice versa.

This temporal asymmetry became even more visible when considering the birth date differences left of the 0.05 and right of the 0.95 quantile, shown as black regions in the plot. In many cases, the top/bottom 5% birth date differences were located in the outer tails of the distributions so that their respective regions were barely visible. While the top and bottom 5% of the ULAN links all had an absolute birth date difference above 54 years, the bottom 5% of the hyperlinks in the combined Wikipedia network had birth date differences below -221 years, while the top 5% of the hyperlinks were more like the ULAN, having birth date differences above 57 years. This clearly showed that the majority of "long distance" Wikipedia links pointed from younger to older persons born many years earlier and a check for role distributions revealed that more than half of the links involved non-artists, which was remarkable since the proportion of non-artists was low in the overall dataset. For the individual languages, the boundary birth date differences for the bottom 5% ranged from -110 down to -1,125 years, those for the upper 5% from 44 to 245 years, the largest individual English language version for example had the bottom 5% below -158 and top 5% above 46 years. Overall it appeared that especially the top 5% of the hyperlinks in the larger Wikipedia versions tended to have relatively similar boundary birth date differences like the ULAN, while this was less the case for smaller Wikipedias.

The variety of different birth date differences and the observed asymmetry suggested to look at the joint temporal distribution of the birth dates of connected persons in the combined language network. Figure 5.14 shows the respective 2D histogram, revealing that the range of source/target birth dates with the highest concentration of hyperlinks coincided with the modern age, starting at around 1250 AD and culminating during the late 19th / early 20th century. This range also clearly showed the highest density of links with small birth date differences, visible in the diagonal of the plot (seen from the lower left to the upper right) and representing their substantial contribution to the previously made observations. The densities at off-diagonal positions in the same range again clearly revealed the temporal asymmetry in hyperlinks, being clearly higher in the area above the diagonal representing links from more recent to older persons .

The high number of hyperlinks in this range was closely tied to the underlying distribution of ULAN biographies across the ages, shown as marginal line plots where the grey line represents over time the counts of birth dates of ULAN biographies having at least one hyperlink in one Wikipedia version, while the black line shows the temporal distribution of birth dates for sources (right) and targets (top) of the hyperlinks in the combined Wikipedia network. Especially the two peaks at around 1500AD and 1600AD, mainly representing biographies related to the Italian Renaissance and the Dutch Golden age, showed that the number of hyperlinks starting or ending at a specific birth date was not always tied to the number of biographies having that birth date since obviously, biographies about famous historical key persons attracted many links individually.

Providing a link-based perspective on the clear gap in ULAN biographies for persons born between antiquity and the modern age discussed in Section 4.2 and shown in Figure 4.5, the plot in Figure 5.14 served as additional means to compare the temporal structure of Wikipedia links with their ULAN counterparts. This also highlighted one of the

key differences between the two data sources, showing that Wikipedia contained many influential and referential links between persons which were not based on direct contact. Two separate "clusters" at the diagonal represented contemporary links between persons from mainly Greek (around 500 BC) and Roman (around 0AD) origin and again revealed the interrupted succession between these two groups of persons and the main group. In contrast to the ULAN, whose links were limited to direct contact and would thus only appear close to the diagonal of the plot, however, Greek and Roman antiquity were interconnected through "medium distance" Wikipedia links having an average birth date difference of about 500 years, represented in the plot by two complementary off-diagonal clusters. Antiquity and the modern age were in turn bridged through "long distance" Wikipedia links, spanning birth date differences between about 1,000 and 3,000 years. While the temporal asymmetry in long distance links was even stronger than for "shorter" links (28.34% pointed "into the future"), some exceptions became visible as well. One notable example was the antique Greek painter Peiraikos, who was connected to other ULAN biographies via 23 hyperlinks, 21 of which were exclusively present in the English Wikipedia. 19 of the 23 links exceeded 2,500 years in distance and pointed from his biography to mainly Spanish, Italian, Dutch and French persons from the modern age, visible as succession of dots in the lowest right of the plot.

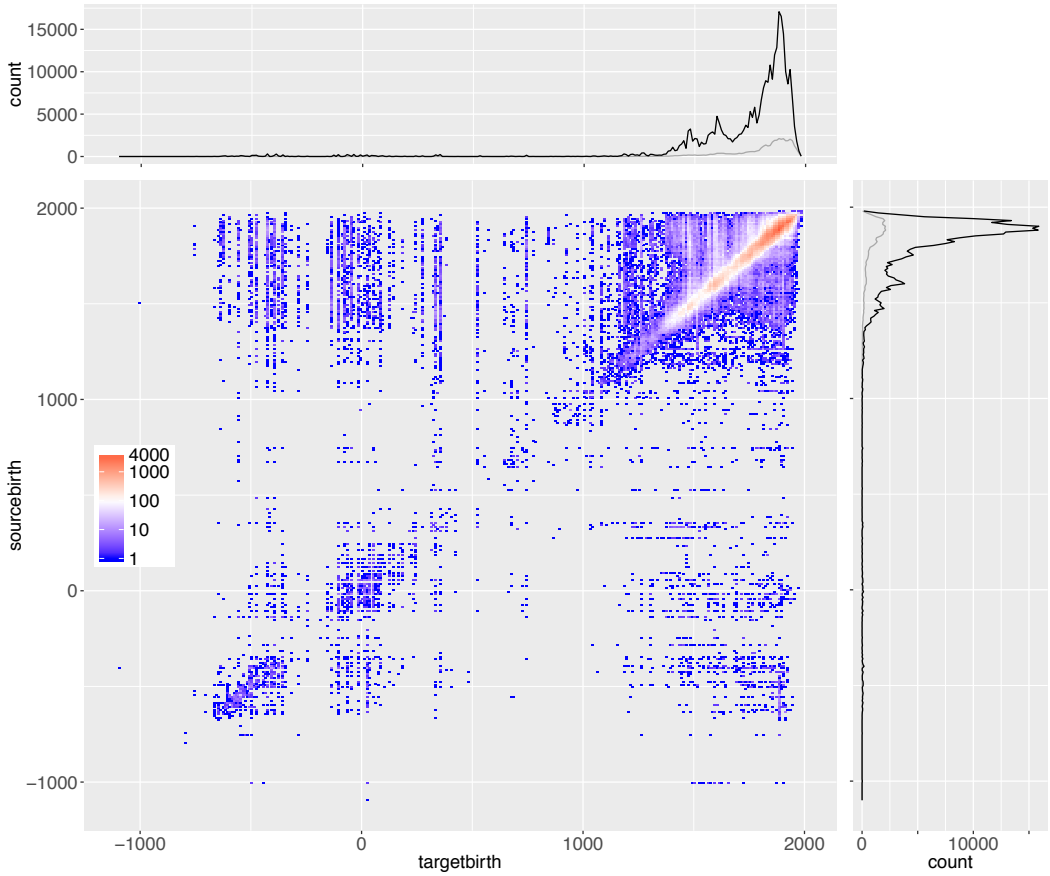


Figure 5.14: 2D Histogram of linked birth dates in the combined Wikipedia network

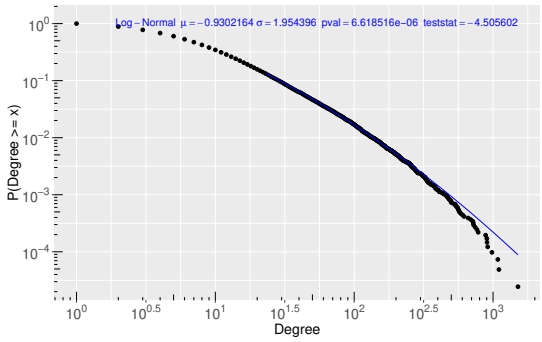
5.5.4 Degree distribution in the Wikipedia person-to-person network

The asymmetry in Wikipedia hyperlinks was also expressed in the respective degree distribution of the nodes in the hyperlink network. In the combined Wikipedia for example, the most highly connected person was Pablo Picasso, having 186 outgoing and 1,326 incoming hyperlinks to and from other ULAN biographies in Wikipedia, a ratio of about 7.14. Figure 5.15 shows the respective complementary cumulative degree distributions (CCDF) for the combined Wikipedia network, separated by full- (a) , in- (b) and out-degree (c). The results of fitting various distribution models to the observed data, using the same procedure as previously discussed for the ULAN associative relationships and shown in Figure 4.13, indicated that all the three distributions were heavy-tailed, although only the out-degree distribution appeared to follow a power-law in this regard. In contrast to the degree distribution of the ULAN-network, where the contributions of the different link types resulted in its mixed appearance, the shapes of Wikipedia degree distributions appeared to be quite homogeneous instead. Since the overlap analysis between ULAN and Wikipedia links revealed a high coverage of both family and teaching ULAN relationships in the latter, the difference was most likely rooted in the many additional hyperlinks present in the free encyclopedia which were not solely based on documented direct contact but also represented broader and externally conceived relations such as perceived influence or statements about other common aspects between the referencing and the referenced person. Considering the individual language versions revealed that the shapes of the respective distributions were relatively similar, as it was also shown by Zlatic et. al in [Zlatić et al., 2006] for the complete Wikipedia hyperlink networks (including non-ULAN biographies and all the other non-person articles).

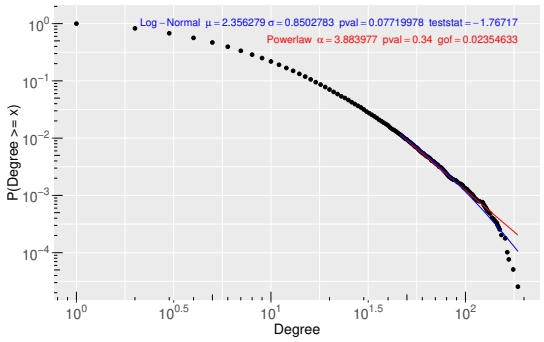
As far as the correlation between in- and out-degrees was concerned, Pearson correlation measures obtained for each language version and the combined network yielded a strong correlation of 0.74 for the combined network and moderate to strong correlation for the individual languages usually ranging between 0.4 and 0.6. The correlation coefficient of 0.54 obtained for the ULAN subnetwork in the English Wikipedia stood in contrast to the results obtained by Kamps and Koolen in [Kamps and Koolen, 2009] for the full English Wikipedia corpus from 2006, for which they reported a weak correlation of only 0.19. A replication of this measurement with the English Wikipedia network as of early 2016 yielded a correspondingly low coefficient of 0.12. This confirmed that the Wikipedia subnetwork of ULAN mapped biographies was significantly different than the overall Wikipedia hyperlink network and a reasonable explanation was that the latter contained many additional articles about entities such as places or general concepts, which were very often referenced from other articles but contained only a limited number of links themselves.

Degree rankings

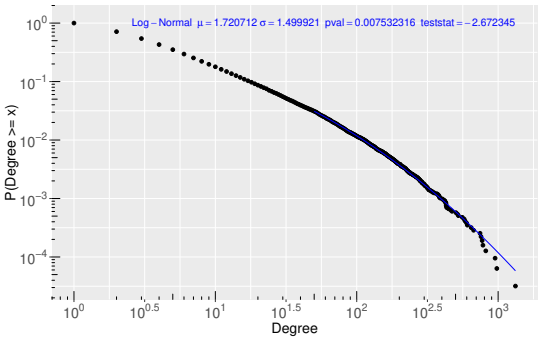
The relatively high correlation between in- and outdegree suggested that biographies featuring important historical figures both tended to have longer articles with many outgoing links as well as many inbound references from other biographies in turn. Given



(a) Degree



(b) Out-Degree



(c) In-Degree

Figure 5.15: Degree-CCDF for the combined Wikipedia network

the observations regarding self-focus bias in different Wikipedia language versions, it was of interest to give a closer look at the top-ranked persons in each language and to check for the presence of culturally related nationalities there. Moreover, the observation that smaller Wikipedias tended to feature many non-artists suggested to look at the roles of the top-ranked persons as well.

Top-100 in- and out degree rankings were thus generated for all language versions having at least 1,000 hyperlinks and the proportions of persons culturally related to the respective language calculated based on their ULAN preferred nationality, assuming cultural ties mainly based on common language as well as well-known historical/colonial ties. Proportions of non-artists in the top-ranked persons were calculated by the ULAN preferred role instead. Figure 5.16 shows the results for the languages ranked by descending number of hyperlinks. The plot again confirmed the growing proportion of non-artist ULAN biographies in smaller Wikipedia language versions, which coincided with a decreasing proportion of culturally related ULAN persons there. It interestingly also revealed that in most cases, persons top-ranked by out-degree seemed to be slightly more culturally related and a little more likely to be artists compared to the in-degree ranks, the latter even in smaller language versions.

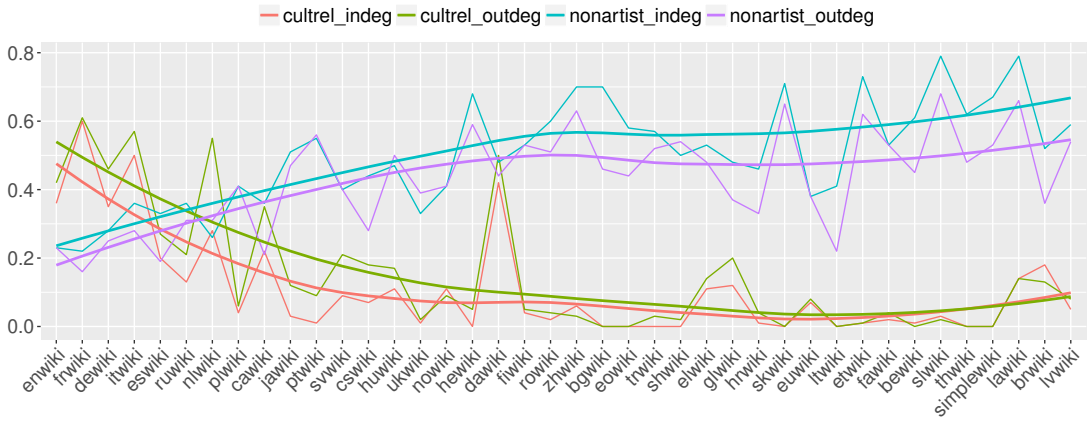


Figure 5.16: Proportions of culturally related and non-artist persons in top 100 degree rankings derived from ULAN biography networks in 40 Wikipedias

Although following the above described trends, the proportions of culturally related and non-artist persons fluctuated quite sharply even across larger Wikipedias. Especially the Wikipedias with languages spoken in the "classical" contributing regions to art history, such as France, Italy and the Netherlands, had high proportions of culturally related, high ranking persons. Since many of these persons represented the "all stars" of art history and other historically important persons featured in the ULAN, they were also present in the top-rankings of other large Wikipedias with many unique contributions, such as the English, which resulted in relatively lower proportions of culturally related top-ranked persons there. A clear outlier in this regard was the Danish Wikipedia (dawiki), having very large proportions of top-ranked culturally related ULAN biographies for both in-and out-degrees. Indeed, although only representing a small fraction in the ULAN overall (1.01% for the whole ULAN, 1.45% for the subset identified in Wikidata), Danes represented about 20% of the identified ULAN biographies in the Danish Wikipedia and their high in- and out-degree ranks suggested that they were all very tightly interlinked.

Looking at the names of the highest ranking persons in the different language versions again reflected their differing correlations between in- and out-degree. Table 5.10 shows the top-20 persons ranked by in- (upper half) and out-degree (lower half) for the combined Wikipedia network and for a selection of individual languages. For each ranking, persons culturally related to the respective language are highlighted in bold. The top-20 in- and out-degree rankings overlapped by 10 persons in the combined Wikipedia and the Japanese language network, the other languages had less agreement there. It was interesting to see that persons high-ranked by in-degree were well known historical figures which were top-ranked in a number of languages. A number of persons highest-ranked by out-degree were in turn quite unknown for a general audience, which was also reflected in the clearly lower agreement between out-degree rankings across languages. Such "locally important" figures often turned out to be art professionals such as curators, critics or collectors having numerous artist references in their respective Wikipedia articles.

As indicated in Figure 5.16, the top 20 rankings for the Danish Wikipedia contained many Danish artists who were often higher ranked than the much better known "all stars", while those for the Chinese Wikipedia (zhwiki) stood as an example for the many language versions with small ULAN coverage which featured many famous but culturally unrelated non-artist figures. In this regard it was notable that the German/Austrian dictator Adolf Hitler quite frequently appeared as high ranked person especially in in-degree rankings (30 out of 51). Given the global changes imposed by WWII, but also the explicit National-Socialist policy of banning so-called "degenerate art", his influence on art history couldn't be denied and his biography thus quite often referenced by ULAN artist biographies in Wikipedia. Notable exceptions for small Wikipedias with little ULAN coverage but relatively high proportions of culturally related persons were the Greek (elwiki) and the Latin (lawiki) Wikipedias, obviously containing many globally recognized figures from Antiquity.

In-degree

combined	enwiki	dewiki	frwiki	itwiki
Picasso (Spanish artist)	Picasso (Spanish artist)	Picasso (Spanish artist)	Picasso (Spanish artist)	**Vasari (Italian artist)**
Napoleon I (French emperor)	Napoleon I (French emperor)	**Goethe (German artist)**	**Napoleon I (French emperor)**	**Raphael (Italian artist)**
Raphael (Italian artist)	Raphael (Italian artist)	Napoleon I (French emperor)	Vasari (Italian artist)	**Michelangelo Buonarroti (Italian artist)**
Michelangelo Buonarroti (Italian artist)	Matisse (French artist)	Raphael (Italian artist)	**Louis XIV (French king)**	Napoleon I (French emperor)
Rembrandt van Rijn (Dutch artist)	**Victoria (British queen)**	Matisse (French artist)	Raphael (Italian artist)	Picasso (Spanish artist)
Cézanne (French artist)	Rubens (Flemish artist)	Cézanne (French artist)	**David (French artist)**	**Titian (Italian artist)**
Rubens (Flemish artist)	Rembrandt van Rijn (Dutch artist)	**Klee (Swiss artist)**	**Matisse (French artist)**	**Leonardo da Vinci (Italian artist)**
Matisse (French artist)	Michelangelo Buonarroti (Italian artist)	Rembrandt van Rijn (Dutch artist)	**Cézanne (French artist)**	Louis XIV (French king)
Titian (Italian artist)	Vasari (Italian artist)	**Hitler (German artist)**	Rubens (Flemish artist)	**Caravaggio (Italian artist)**
Shakespeare (British author)	Cézanne (French artist)	Michelangelo Buonarroti (Italian artist)	Michelangelo Buonarroti (Italian artist)	Rubens (Flemish artist)
Gogh (Dutch artist)	**Shakespeare (British author)**	Kandinsky (Russian artist)	Rembrandt van Rijn (Dutch artist)	Rembrandt van Rijn (Dutch artist)
Goethe (German artist)	Titian (Italian artist)	Rubens (Flemish artist)	**Napoleon III (French emperor)**	Shakespeare (British author)
Louis XIV (French king)	Louis XIV (French king)	**Dürer (German artist)**	**Louis XVI (French king)**	Gombrich (British art historian)
Vasari (Italian artist)	Gogh (Dutch artist)	Vasari (Italian artist)	**Delacroix (French artist)**	**Alighieri (Italian poet)**
Leonardo da Vinci (Italian artist)	**Warhol (American artist)**	Louis XIV (French king)	**Hugo (French artist)**	Charles V (Holy Roman Imperial emperor)
Le Corbusier (Swiss artist)	Le Corbusier (Swiss artist)	Gogh (Dutch artist)	Titian (Italian artist)	**Correggio (Italian artist)**
Hitler (German artist)	Leonardo da Vinci (Italian artist)	Titian (Italian artist)	**Breton (French artist)**	**Bernini (Italian artist)**
Victoria (British queen)	Goethe (German artist)	**Le Corbusier (Swiss artist)**	**Le Corbusier (Swiss artist)**	David (French artist)
Dürer (German artist)	Dyck (Flemish artist)	Shakespeare (British author)	Leonardo da Vinci (Italian artist)	Cézanne (French artist)
Rodin (French artist)	**Pollock (American artist)**	Leonardo da Vinci (Italian artist)	**Ingres (French artist)**	Goethe (German artist)
eswiki	**ruwiki**	**jawiki**	**dawiki**	**zhwiki**
Picasso (Spanish artist)	Picasso (Spanish artist)	Napoleon I (French emperor)	**Thorvaldsen (Danish artist)**	Napoleon I (French emperor)
Vasari (Italian artist)	Napoleon I (French emperor)	Picasso (Spanish artist)	Napoleon I (French emperor)	Shakespeare (British author)
Michelangelo Buonarroti (Italian artist)	Raphael (Italian artist)	Michelangelo Buonarroti (Italian artist)	**Eckersberg (Danish artist)**	Aristotle (Greek philosopher)
Raphael (Italian artist)	Rembrandt van Rijn (Dutch artist)	Gogh (Dutch artist)	Goethe (German artist)	Louis XIV (French king)
Napoleon I (French emperor)	Michelangelo Buonarroti (Italian artist)	Raphael (Italian artist)	Hitler (German artist)	Picasso (Spanish artist)
Rubens (Flemish artist)	Cézanne (French artist)	Shakespeare (British author)	Shakespeare (British author)	Kant (German philosopher)
Velázquez (Spanish artist)	Vasari (Italian artist)	Le Corbusier (Swiss artist)	Michelangelo Buonarroti (Italian artist)	Marx (German philosopher)
Goya (Spanish artist)	Titian (Italian artist)	Hitler (German artist)	**Krøyer (Danish artist)**	Plato (Greek philosopher)
Titian (Italian artist)	Matisse (French artist)	Leonardo da Vinci (Italian artist)	Rembrandt van Rijn (Dutch artist)	Michelangelo Buonarroti (Italian artist)
Louis XIV (French king)	Louis XIV (French king)	Vasari (Italian artist)	Louis XIV (French king)	Hitler (German artist)
Philip II (Spanish king)	Rubens (Flemish artist)	Louis XIV (French king)	**Marstrand (Danish artist)**	George III (British king)
Leonardo da Vinci (Italian artist)	Goethe (German artist)	Victoria (British queen)	Picasso (Spanish artist)	Victoria (British queen)
Rembrandt van Rijn (Dutch artist)	Leonardo da Vinci (Italian artist)	Plato (Greek philosopher)	**Lund (Danish artist)**	Nietzsche (German philosopher)
Palomino de Castro y Velasco (Spanish artist)	Shakespeare (British author)	Kant (German philosopher)	Marx (German philosopher)	Voltaire (French author)
Charles V (Holy Roman Imperial emperor)	**Catherine II (Russian emperor)**	Rubens (Flemish artist)	Louis XVI (French king)	Hegel (German philosopher)
Caravaggio (Italian artist)	Rodin (French artist)	Goethe (German artist)	**Zahrtmann (Danish artist)**	Freud (Austrian psychiatrist)
Dürer (German artist)	Gogh (Dutch artist)	Napoleon III (French emperor)	**Abildgaard (Danish artist)**	Louis XVI (French king)
Shakespeare (British author)	David (French artist)	Rembrandt van Rijn (Dutch artist)	**Andersen (Danish artist)**	Alighieri (Italian poet)
Cézanne (French artist)	Manet (French artist)	Charles V (Holy Roman Imperial emperor)	David (French artist)	Leonardo da Vinci (Italian artist)
Matisse (French artist)	**Lenin (Russian politician)**	Cézanne (French artist)	**Hetsch (Danish architect)**	Rousseau (French philosopher)

Out-degree

combined	enwiki	dewiki	frwiki	itwiki
Picasso (Spanish artist)	Cavellini (Italian artist)	Solly (British merchant)	**Laurencin (French artist)**	**Cavellini (Italian artist)**
David (French artist)	Gool (Dutch artist)	Picasso (Spanish artist)	**David (French artist)**	David (French artist)
Leonardo da Vinci (Italian artist)	**Dafforne (British artist)**	Vasari (Italian artist)	**Cabanel (French artist)**	**Leonardo da Vinci (Italian artist)**
Michelangelo Buonarroti (Italian artist)	Ticozzi (Italian politician)	**Klee (Swiss artist)**	**Delacroix (French artist)**	**Michelangelo Buonarroti (Italian artist)**
Gogh (Dutch artist)	Tzara (Romanian artist)	**Liebermann (German artist)**	Hubbard (American artist)	**Caravaggio (Italian artist)**
Mander (Dutch artist)	**Sharp (American artist)**	Brancusi (Romanian artist)	Leonardo da Vinci (Italian artist)	**Raphael (Italian artist)**
Matisse (French artist)	**Ruskin (British artist)**	**Graff (Swiss artist)**	**Cogniet (French artist)**	Gogh (Dutch artist)
Duchamp (French artist)	Boschini (Italian artist)	Man Ray (American artist)	Velázquez (Spanish artist)	**Pascoli (Italian lawyer)**
Laurencin (French artist)	**Soane (British artist)**	Cézanne (French artist)	Caravaggio (Italian artist)	Dyck (Flemish artist)
Cavellini (Italian artist)	**Miller (American administrator)**	**Ernst (German artist)**	Mander (Dutch artist)	**Titian (Italian artist)**
Dalí (Spanish artist)	Hondius (Dutch artist)	Rembrandt van Rijn (Dutch artist)	**Duchamp (French artist)**	Piles (French artist)
Dürer (German artist)	Dominici (Italian artist)	Miró (Spanish artist)	**Louis XVI (French king)**	**Giotto (Italian artist)**
Rubens (Flemish artist)	**Larmon (American artist)**	Cabanel (French artist)	Dalí (Spanish artist)	Breton (French artist)
Caravaggio (Italian artist)	Matisse (French artist)	Cage (American artist)	Gogh (Dutch artist)	Paalen (Austrian artist)
Vasari (Italian artist)	**Coetzee (South African artist)**	Dalí (Spanish artist)	**Rigaud (French artist)**	**Ferretti (Italian artist)**
Le Corbusier (Swiss artist)	Duchamp (French artist)	**Giacometti (Swiss artist)**	Miró (Spanish artist)	**Carracci (Italian artist)**
Rembrandt van Rijn (Dutch artist)	Breton (French artist)	Duchamp (French artist)	**Braque (French artist)**	**Brunelleschi (Italian artist)**
Cabanel (French artist)	Leonardo da Vinci (Italian artist)	Matisse (French artist)	**Pignon (French artist)**	Louis XIV (French king)
Miró (Spanish artist)	Louis XIV (French king)	**Friedrich (German artist)**	**Poussin (French artist)**	**Pinturicchio (Italian artist)**
Velázquez (Spanish artist)	Gogh (Dutch artist)	Mondrian (Dutch artist)	Rubens (Flemish artist)	Voltaire (French author)
eswiki	**ruwiki**	**jawiki**	**dawiki**	**zhwiki**
Mander (Dutch artist)	Ingres (French artist)	Houbraken (Dutch artist)	Friedrich (German artist)	Marx (German philosopher)
Picasso (Spanish artist)	Dürer (German artist)	Mander (Dutch artist)	Louis XIV (French king)	Goethe (German artist)
Velázquez (Spanish artist)	Liebermann (German artist)	Gogh (Dutch artist)	**Marstrand (Danish artist)**	Shakespeare (British author)
Lorrain (French artist)	Gérôme (French artist)	Leonardo da Vinci (Italian artist)	Vermeer (Dutch artist)	Faulkner (American author)
Michelangelo Buonarroti (Italian artist)	Ernst (German artist)	Nietzsche (German philosopher)	**Eckersberg (Danish artist)**	Hegel (German philosopher)
Murillo (Spanish artist)	Bouguereau (French artist)	Cézanne (French artist)	**Bindesboll (Danish artist)**	William III (British king)
Leonardo da Vinci (Italian artist)	**Repin (Russian artist)**	**Domon (Japanese artist)**	**Slott-Møller (Danish artist)**	Kant (German philosopher)
Thyssen-Bornemisza (Swiss collector)	Deleuze (French philosopher)	Louis XIV (French king)	Visconti (Italian artist)	Louis XIV (French king)
Fuseli (Swiss artist)	Matisse (French artist)	Goering (German collector)	**Haslund (Danish artist)**	Nietzsche (German philosopher)
Goya (Spanish artist)	**Vrubel' (Russian artist)**	Churchill (British artist)	Courbet (French artist)	Virgil (Roman poet)
Nietzsche (German philosopher)	Napoleon I (French emperor)	Hegel (German philosopher)	**Thorvaldsen (Danish artist)**	Lévi-Strauss (French anthropologist)
Dalí (Spanish artist)	Corinth (German artist)	Marx (German philosopher)	Nietzsche (German philosopher)	Sargent (American artist)
Weyden (Flemish artist)	Caravaggio (Italian artist)	Napoleon I (French emperor)	**Bentsen (Danish architect)**	Matisse (French artist)
Titian (Italian artist)	Alma-Tadema (Dutch artist)	Goethe (German artist)	**Meldahl (Danish architect)**	Rousseau (French philosopher)
Gaudí (Spanish artist)	Siemiradzki (Polish artist)	Picasso (Spanish artist)	**Jerndorff (Danish artist)**	Freud (Austrian psychiatrist)
Rubens (Flemish artist)	Rembrandt van Rijn (Dutch artist)	Kant (German philosopher)	Houbraken (Dutch artist)	Ernst (German artist)
Joyce (Irish author)	**Morozov (Russian collector)**	**Ikeda (Japanese artist)**	Mendelssohn-Bartholdy (German artist)	Napoleon I (French emperor)
Canova (Italian artist)	Leonardo da Vinci (Italian artist)	Victoria (British queen)	Nadar (French artist)	Beethoven (German composer)
Rigaud (French artist)	**Dubovskoy (Russian artist)**	Rubens (Flemish artist)	Spinoza (Dutch philosopher)	Henry VIII (British king)
Borges (Argentine artist)	Leibniz (German philosopher)	Verrocchio (Italian artist)	**Bloch (Danish artist)**	Ferdinand I (Italian king)

Table 5.10: Top 20 in- and out-degree rankings of ULAN persons in the combined Wikipedia and nine selected languages

The degree rankings shown in Table 5.10 were solely derived from the hyperlink networks between ULAN biographies covered in the respective Wikipedia language versions. It was therefore of interest to observe the global degree of ULAN biographies in each language, also taking all the other non-ULAN Wikipedia articles they were interlinked with into account. In order to distinguish these two different rankings from each other, ranking ULAN biographies in Wikipedia by their references to and from other ULAN biographies in the same language was referred to as "ULAN only", while ranking ULAN biographies in Wikipedia by all other Wikipedia articles in that language was as "global", respectively. Figure 5.17 again shows the proportions of culturally related persons and those of non-artists, this time for the respective global top-100 in- and out-degree rankings. In the global ranking, non-artist figures were now clearly more often referenced in all language versions regardless of size, while the proportion of culturally related persons appeared to decrease with size, similar to the ULAN only rankings. Since only ULAN mapped biographies were considered in this study and the ULAN itself represented an obviously western oriented data source, it could nevertheless be expected that those languages featuring little culturally related ULAN persons in turn contained other persons specifically relevant to their individual histories, which were not observed here.

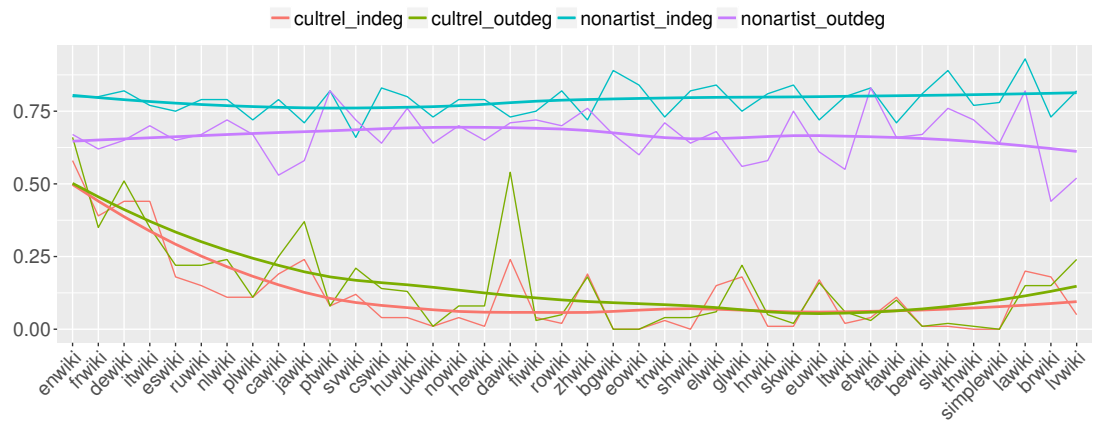


Figure 5.17: Proportions of culturally related and non-artist ULAN persons in global top 100 degree rankings derived from the complete networks in 40 Wikipedias

Table 5.12 shows the globally derived top 20 in- and out-degree rankings for the same language versions as Table 5.10. Especially the in-degree rankings again overlapped between language versions for certain common historical figures and there was again less overlap regarding out-degree. For the languages shown in the Table, it became visible that especially the global top-20 rankings for the Japanese and the Chinese Wikipedias now contained many more culturally relevant persons. It was interesting to see two Japanese military officers to be top-ranked in the Japanese Wikipedia, which was especially notable since they were the only Japanese military officers present in the ULAN. Besides persons

from nobility and the cleric, other languages too seemed to contain globally top-ranked roles specific to their respective cultures. The English Wikipedia featured high-ranked American presidents, but also persons from rather popular culture, such as film directors or musicians, while others seemed to emphasize famous composers and philosophers. The ULAN focus on artistic activity, however, again became clearly visible in the role descriptions of some of the globally top-ranked persons, where figures most known for non-fine-arts related achievements were nevertheless assigned the role artist, as long as they had some artistic output documented in their biographies.

The observation that some persons were featured in the top in- and out-degree ranks from multiple Wikipedia languages suggested to quantify their co-occurrence across similar ranking types in different languages. Table 5.11 lists the 20 persons with the highest cross-Wikipedia co-occurrences for each of the top-100 ULAN only and global in- and out-degree rankings, excluding those languages whose ULAN only network consisted of less than 1,000 hyperlinks. Each list shows the name of the person, the number of language versions he or she appeared in the respective top-100 ranking and the number of languages generally featuring that person, regardless of the individual rank. This confirmed that persons ranked by in-degree co-occurred much more frequently in both ULAN only and global top rankings. While only about half of the top-20 co-occurring persons in both ULAN only in- and out-degree rankings were artists, the global rankings featured even less of them, where just the top-20 out degree ranking included artists at all, Leonardo and Michelangelo.

Wikipedia networks of ULAN biographies						Full Wikipedia networks					
In-degree			Out-degree			In-degree			Out-degree		
Napoleon I, Emperor of the French	40	40	Hegel, Georg Wilhelm Friedrich	20	40	Shakespeare, William	40	40	Hitler, Adolf	35	40
Michelangelo Buonarroti	39	40	Louis XIV, King of France	18	40	Napoleon I, Emperor of the French	40	40	John Paul II, Pope	19	40
Picasso, Pablo	38	40	Kant, Immanuel	18	40	Hitler, Adolf	40	40	Churchill, Winston	18	40
Shakespeare, William	35	40	Leonardo da Vinci	18	40	Aristotle	36	40	Napoleon I, Emperor of the French	18	40
Louis XIV, King of France	33	40	Michelangelo Buonarroti	16	40	Mozart, Wolfgang Amadeus	29	40	Einstein, Albert	17	40
Raphael	33	40	Dalí, Salvador	15	40	Alexander, 356-323 B.C.	26	40	Alexander, 356-323 B.C.	14	40
Leonardo da Vinci	31	40	Rubens, Peter Paul	14	40	Augustus, Emperor of Rome	26	40	Elizabeth II, Queen of Great Britain	13	40
Vasari, Giorgio	29	39	Nietzsche, Friedrich	14	40	Elizabeth II, Queen of Great Britain	25	40	Louis XIV, King of France	13	40
Goethe, Johann Wolfgang von	28	40	Marx, Karl	14	40	John Paul II, Pope	25	40	Mussolini, Benito	13	40
Rubens, Peter Paul	27	40	Gogh, Vincent van	12	40	Louis XIV, King of France	22	40	Darwin, Charles	11	40
Aristotle	27	40	Matisse, Henri	12	40	Plato	22	40	Leonardo da Vinci	11	40
Hitler, Adolf	26	40	Titian	12	40	Charlemagne, Emperor	21	40	Columbus, Christopher	10	40
Charles V, Holy Roman Emperor	26	40	Rousseau, Jean-Jacques	10	40	Goethe, Johann Wolfgang von	13	40	Augustus, Emperor of Rome	10	40
Titian	25	40	Caravaggio, Michelangelo Merisi da	10	40	Lenin, V. I.	13	40	Marx, Karl	9	40
Plato	24	40	Manet, Edouard	9	40	Charles V, Holy Roman Emperor	13	40	Michelangelo Buonarroti	9	40
Cézanne, Paul	23	40	Napoleon I, Emperor of the French	8	40	Tolkien, J. R. R.	12	40	Lenin, V. I.	9	40
Rembrandt van Rijn	20	40	Hugo, Victor	8	40	Einstein, Albert	12	40	Charlemagne, Emperor	9	40
Matisse, Henri	19	40	Goethe, Johann Wolfgang von	8	40	Beethoven, Ludwig van	12	40	Dylan, Bob	8	40
Kant, Immanuel	18	40	Raphael	8	40	Herschel, Sir John	11	36	Spielberg, Steven	8	40
Victoria, Queen of Great Britain	16	40	Velázquez, Diego	8	40	Wagner, Richard	11	40	Victoria, Queen of Great Britain	8	40

Table 5.11: Persons most often co-occurring in Top 100 in- and out-degree rankings in different Wikipedia language versions

	combined	enwiki	dewiki	frwiki	itwiki
In-degree	Napoleon I (French emperor)	**Shakespeare (British author)**	Napoleon I (French emperor)	**Napoleon I (French emperor)**	Napoleon I (French emperor)
	Shakespeare (British author)	**Elizabeth II (British queen)**	John Paul II (Polish pope)	**Louis XIV (French king)**	John Paul II (Polish pope)
	Hitler (German artist)	Napoleon I (French emperor)	**Hitler (German artist)**	Shakespeare (British author)	Shakespeare (British author)
	John Paul II (Polish pope)	Hitler (German artist)	**Goethe (German artist)**	**Napoleon III (French emperor)**	Hitler (German artist)
	Mozart (Austrian composer)	**Victoria (British queen)**	Shakespeare (British author)	Hitler (German artist)	**Alighieri (Italian poet)**
	Elizabeth II (British queen)	**Roosevelt (American president)**	**Mozart (Austrian composer)**	**Hugo (French artist)**	**Mussolini (Italian politician)**
	Louis XIV (French king)	**Dylan (American artist)**	**Luther (German theologian)**	**Louis XVI (French king)**	**Augustus (Roman emperor)**
	Beethoven (German composer)	**Kennedy (American president)**	**Frederick II (Prussian king)**	John Paul II (Polish pope)	**Benedict XIV (Italian pope)**
	Goethe (German artist)	**Lincoln (American president)**	**Schiller (German author)**	**Mitterrand (French president)**	Charles V (Holy Roman Imperial emperor)
	Aristotle (Greek philosopher)	John Paul II (Polish pope)	**Wagner (German composer)**	Mozart (Austrian composer)	**Paul VI (Italian pope)**
	Lenin (Russian politician)	**Washington (American president)**	Paul VI (Italian pope)	**Louis XIII (French king)**	**Benedict XIII (Italian pope)**
	Victoria (British queen)	Mozart (Austrian composer)	**Charlemagne (Frankish emperor)**	**Francis I (French king)**	Mozart (Austrian composer)
	Roosevelt (American president)	**Churchill (British artist)**	**Beethoven (German composer)**	Charles V (Holy Roman Imperial emperor)	**Francis of Assisi (Italian religious leader)**
	Charles V (Holy Roman Imperial emperor)	**Henry VIII (British king)**	**Dehio (German architectural historian)**	**Charlemagne (Frankish emperor)**	**Pius IX (Italian pope)**
	Wagner (German composer)	**McCartney (British musician)**	Aristotle (Greek philosopher)	Aristotle (Greek philosopher)	Aristotle (Greek philosopher)
	Dylan (American artist)	**Bush (American president)**	**Franz Joseph I (Austrian emperor)**	**Voltaire (French author)**	Frederick II (Holy Roman Imperial emperor)
	Plato (Greek philosopher)	**Eisenhower (American artist)**	Louis XIV (French king)	**Louis XI (French king)**	**Clement XIII (Italian pope)**
	Charlemagne (Frankish emperor)	**Lennon (British artist)**	**Charles V (Holy Roman Imperial emperor)**	**Louis Philippe (French king)**	Charlemagne (Frankish emperor)
	Kennedy (American president)	**Sinatra (American singer)**	**Brecht (German playwright)**	**Louis IX (French king)**	Louis XIV (French king)
	Marx (German philosopher)	Beethoven (German composer)	**Marx (German philosopher)**	Augustus (Roman emperor)	**Pliny the Elder (Roman author)**
	eswiki	**ruwiki**	**jawiki**	**dawiki**	**zhwiki**
	Napoleon I (French emperor)	**Peter I (Russian emperor)**	**Toyotomi Hideyoshi (Japanese military officer)**	Hitler (German artist)	**Mao Zedong (Chinese theorist)**
	Philip II (Spanish king)	**Catherine II (Russian emperor)**	**Oda Nobunaga (Japanese military officer)**	Napoleon I (French emperor)	**Confucius (Chinese philosopher)**
	Charles V (Holy Roman Imperial emperor)	**Lenin (Russian politician)**	Napoleon I (French emperor)	**Andersen (Danish artist)**	Napoleon I (French emperor)
	Hitler (German artist)	Napoleon I (French emperor)	Hitler (German artist)	**Margrethe II (Danish queen)**	**Tang Xuanzong (Chinese emperor)**
	John Paul II (Polish pope)	Shakespeare (British author)	Mozart (Austrian composer)	**Frederick IV (Danish king)**	Hitler (German artist)
	Shakespeare (British author)	**Paul I (Russian emperor)**	Beethoven (German composer)	Shakespeare (British author)	Toyotomi Hideyoshi (Japanese military officer)
	Pliny the Elder (Roman author)	**Nicholas I (Russian emperor)**	Shakespeare (British author)	**Christian VII (Danish king)**	Elizabeth II (British queen)
	Philip V (Spanish king)	Hitler (German artist)	**Kitano (Japanese artist)**	Luther (German theologian)	**Yongle (Chinese emperor)**
	Bolívar (Venezuelan president)	John Paul II (Polish pope)	**Tezuka (Japanese artist)**	Mozart (Austrian composer)	**Kublai Khan (Chinese emperor)**
	Mozart (Austrian composer)	Mozart (Austrian composer)	**K?kai (Japanese artist)**	Wagner (German composer)	Oda Nobunaga (Japanese military officer)
	Aristotle (Greek philosopher)	Herschel (British artist)	Lennon (British artist)	Elizabeth II (British queen)	**Lu Xun (Chinese artist)**
	Augustus (Roman emperor)	Beethoven (German composer)	Wagner (German composer)	Aristotle (Greek philosopher)	**Huizong (Chinese emperor)**
	Isabella II (Spanish queen)	**Lermontov (Russian artist)**	McCartney (British musician)	Charlemagne (Frankish emperor)	Shakespeare (British author)
	Tolkien (British author)	Aristotle (Greek philosopher)	**Kurosawa (Japanese artist)**	Goethe (German artist)	**Su Shi (Chinese artist)**
	Alfonso X (Spanish king)	Nicholas of Myra (Asia Minor saint)	Roosevelt (American president)	Victoria (British queen)	Tolkien (British author)
	Alexander (Greek king)	Elizabeth II (British queen)	Elizabeth II (British queen)	Churchill (British artist)	**Liang (Chinese scholar)**
	Columbus (Spanish explorer)	Louis XIV (French king)	Kennedy (American president)	Louis XIV (French king)	Einstein (German physicist)
	Dylan (American artist)	Wagner (German composer)	Lincoln (American president)	**Thorvaldsen (Danish artist)**	Aristotle (Greek philosopher)
	Louis XIV (French king)	Marx (German philosopher)	Aristotle (Greek philosopher)	Marx (German philosopher)	Roosevelt (American president)
	James I (Aragonese king)	Goethe (German artist)	Mao Zedong (Chinese theorist)	Dylan (American artist)	Lenin (Russian politician)
	combined	**enwiki**	**dewiki**	**frwiki**	**itwiki**
Out-degree	Hitler (German artist)	**McCartney (British musician)**	**Hitler (German artist)**	**Napoleon III (French emperor)**	**Mussolini (Italian politician)**
	John Paul II (Polish pope)	**Churchill (British artist)**	**Goethe (German artist)**	Hitler (German artist)	John Paul II (Polish pope)
	Churchill (British artist)	**Welles (American artist)**	Arendt (American philosopher)	**Louis XVI (French king)**	Dylan (American artist)
	Napoleon I (French emperor)	John Paul II (Polish pope)	Mussolini (Italian politician)	**Robespierre (French revolutionarie)**	**Leonardo da Vinci (Italian artist)**
	Mussolini (Italian politician)	**Dylan (American artist)**	**Marx (German philosopher)**	**Mitterrand (French president)**	Adams (Canadian musician)
	Lenin (Russian politician)	**Eisenhower (American artist)**	Prince (American musician)	Churchill (British artist)	Sinatra (American singer)
	Marx (German philosopher)	**Sinatra (American singer)**	Napoleon I (French emperor)	**Napoleon I (French emperor)**	**Augustus (Roman emperor)**
	Leonardo da Vinci (Italian artist)	Tzara (Romanian artist)	**Adorno (German philosopher)**	**Laurencin (French artist)**	**Pasolini (Italian artist)**
	Einstein (German physicist)	**Scorsese (American artist)**	Augustus II (Saxon king)	Leonardo da Vinci (Italian artist)	Napoleon I (French emperor)
	Alexander (Greek king)	**Spielberg (American artist)**	**Graff (Swiss artist)**	Mussolini (Italian politician)	**Michelangelo Buonarroti (Italian artist)**
	Freud (Austrian psychiatrist)	**Kennedy (American president)**	Plato (Greek philosopher)	Lenin (Russian politician)	Nicholas of Myra (Asia Minor saint)
	Goethe (German artist)	**Ono (Japanese-American artist)**	**Luther (German theologian)**	**Louis IX (French king)**	Hitler (German artist)
	Columbus (Spanish explorer)	**Elizabeth II (British queen)**	**Beuys (German artist)**	Hitchcock (British film director)	Monroe (American actor)
	McCartney (British musician)	**Franklin (American artist)**	Aristotle (Greek philosopher)	**Charlemagne (Frankish emperor)**	**Raphael (Italian artist)**
	Dylan (American artist)	Hitler (German artist)	**Keller (Swiss artist)**	**Charles V (French king)**	Sartre (French philosopher)
	Charlemagne (Frankish emperor)	**Harrison (British musician)**	**Mozart (Austrian composer)**	Welles (American artist)	**De Sica (Italian artist)**
	Shakespeare (British author)	Napoleon I (French emperor)	Louis XIV (French king)	**Moreau (French actor)**	Allen (American artist)
	Elizabeth II (British queen)	Mussolini (Italian politician)	Alexander (Greek king)	**Motier (French noble)**	Callas (Greek singer)
	Louis XIV (French king)	Louis XIV (French king)	**Hegel (German philosopher)**	**Louis XV (French king)**	Spielberg (American artist)
	Mozart (Austrian composer)	**Mandela (South African artist)**	John Paul II (Polish pope)	Spielberg (American artist)	Freud (Austrian psychiatrist)
	eswiki	**ruwiki**	**jawiki**	**dawiki**	**zhwiki**
	Jung (Swiss psychiatrist)	**Lenin (Russian politician)**	**Oda Nobunaga (Japanese military officer)**	Churchill (British artist)	**Yongle (Chinese emperor)**
	Gaudí (Spanish artist)	Einstein (German physicist)	Churchill (British artist)	Hitler (German artist)	Oda Nobunaga (Japanese military officer)
	Dylan (American artist)	McCartney (British musician)	**Tezuka (Japanese artist)**	Frederick II (Holy Roman Imperial emperor)	**Mao Zedong (Chinese theorist)**
	Columbus (Spanish explorer)	Napoleon I (French emperor)	**Toyotomi Hideyoshi (Japanese military officer)**	Roosevelt (American president)	Roosevelt (American president)
	Hitler (German artist)	**Roerich (Russian artist)**	**Kitano (Japanese artist)**	**Wenck (Danish architect)**	**Liang (Chinese scholar)**
	Bolívar (Venezuelan president)	**Peter I (Russian emperor)**	Hitler (German artist)	Vermeer (Dutch artist)	Kennedy (American president)
	Borges (Argentine artist)	**Catherine II (Russian emperor)**	Victoria (British queen)	**Margrethe II (Danish queen)**	**Li (Chinese artist)**
	Subirachs (Spanish artist)	Deleuze (French philosopher)	Chaplin (British actor)	**Andersen (Danish artist)**	Toyotomi Hideyoshi (Japanese military officer)
	Philip II (Spanish king)	Jobs (American inventor)	Goering (German collector)	Speer (German architect)	Hitler (German artist)
	Charlemagne (Frankish emperor)	Leibniz (German philosopher)	Marx (German philosopher)	Columbus (Spanish explorer)	Marx (German philosopher)
	Arendt (American philosopher)	**Nicholas I (Russian emperor)**	**Kurosawa (Japanese artist)**	**Nyrop (Danish architect)**	**Lu Xun (Chinese artist)**
	Eisenhower (American artist)	Harrison (British musician)	Napoleon I (French emperor)	**Brummer (Danish architect)**	**Kublai Khan (Chinese emperor)**
	Liszt (Hungarian composer)	Mussolini (Italian politician)	Gogh (Dutch artist)	Mendelssohn-Bartholdy (German artist)	Süleyman (Ottoman emperor)
	Velázquez (Spanish artist)	**Lomonosov (Russian scientist)**	Edward VII (British king)	Luther (German theologian)	Lenin (Russian politician)
	Elizabeth II (British queen)	Humboldt (German artist)	Disraeli (British statesmen)	**Henningsen (Danish designer)**	**Confucius (Chinese philosopher)**
	Darwin (British naturalist)	John Paul II (Polish pope)	**Tsuge (Japanese cartoonist)**	Bernstorff (German statesmen)	Goering (German collector)
	Augustus (Roman emperor)	Churchill (British artist)	Mussolini (Italian politician)	**Storck (Danish architect)**	**Tang Xuanzong (Chinese emperor)**
	Charles V (Holy Roman Imperial emperor)	Burroughs (American artist)	**Mori (Japanese physician)**	**Dahlerup (Danish architect)**	Churchill (British artist)
	Joyce (Irish author)	Freud (Austrian psychiatrist)	Einstein (German physicist)	**Koch (Danish architect)**	McCartney (British musician)
	Michelangelo Buonarroti (Italian artist)	Columbus (Spanish explorer)	Napoleon III (French emperor)	Beethoven (German composer)	Einstein (German physicist)

able 5.12: Top 20 in- and out-degree global rankings of ULAN persons in the complete hyperlink network of the combined Wikipedia and nine selected languages

5.5.5 Rank comparisons with ULAN centralities and scholarly rankings

Additional centrality measures were calculated for the networks in the different Wikipedia language versions and their combination in order to compare the resulting person rankings with the centrality rankings from the ULAN network and the scholarly rankings made by Jensen, Elkins and Dilly introduced in Section 4.3.2. While PageRank, Betweenness and Closeness were calculated for the networks of ULAN-mapped Wikipedia biographies, only the first was calculated for the global Wikipedia network as well, since Betweenness and Closeness imposed very high computational cost for such big networks. Spearman-correlations were subsequently calculated in order to check for the agreement between the Wikipedia rankings and those obtained for the original ULAN network and extracted from the scholarly literature.

Regarding the comparison with the ULAN network, only the rankings for those biographies present in both the ULAN giant component and the giant component of the respective Wikipedia version were considered for correlations, limited to those language versions having an overlap of at least 1,000 persons. Moreover, additional "ULAN only" rankings based on links not exceeding a birth date difference of 75 years were created in order to mimic the link structure of the original ULAN network. For the scholarly rankings, only those language versions having full overlap with the 50 biographies each contained in Elkins' and Jensen's rankings and with at least 130 of the 147 present in Dilly's ranking were concerned.

Tables 5.13 and 5.14 show the results of this comparison. The four sections in Table 5.13 are dedicated to the correlation results between rankings of the various ULAN network centrality measures and the highest correlated centrality type for each Wikipedia language, while those in Table 5.14 show a similar comparison between the scholarly rankings and Wikipedia. Each section is divided into three columns. The first column contains strings divided into three parts, representing the type of centrality measure, the originating network type (u = full network between mapped ULAN biographies, u75 = type u with links exceeding 75 years pruned, f = full Wikipedia network including non-person articles) and the Wikipedia language version ("combined" referring to the combined network of all languages). The second and the third column show the number of overlapping biographies considered for the correlation and the resulting Spearman correlation coefficient.

For the comparisons with the ULAN centrality rankings, the results lead to a number of interesting observations. First of all, the highest correlations were achieved when comparing the rankings obtained from the Wikipedia networks consisting only of mapped ULAN biographies (type "u" and "u75"). Here it moreover became visible that more than half of the best correlated centrality rankings were derived from the pruned Wikipedia biography networks, confirming the effect of removing links exceeding 75 years birth date difference to approximate the structure of the ULAN network. The second observation was that for all observed language versions, in-degree yielded the best correlations with

ULAN degree, PageRank and betweenness, while ULAN closeness had the best agreement with closeness centrality in the Wikipedia network. This went in line with the observation made for the ULAN network that the first three centralities showed relatively high mutual correlation, with closeness centrality rather distinct from them. The degree based measures moreover had highest correlation for the combined Wikipedia network and the English Wikipedia, both having the highest overlap with the ULAN giant component, yielding a moderate positive correlation coefficient between 0.42 and 0.45. Closeness on the other hand was more correlated with the French Wikipedia, yielding a relatively strong correlation coefficient of about 0.58. This suggested that the hyperlink network of ULAN persons in the French Wikipedia had a high structural agreement with the original ULAN network, which was likely related to the strong overall French focus present in both the ULAN giant component and this particular Wikipedia language version.

degs_u_ulan			pagerank_u_ulan			betweenness_u_ulan			closeness_u_ulan		
indegs_u_enwiki	5662	0.43	indegs_u75_enwiki	5644	0.45	indegs_u75_combined	7075	0.43	closeness_u75_frwiki	4109	0.58
indegs_u_combined	7088	0.42	indegs_u_combined	7088	0.45	indegs_u75_enwiki	5644	0.42	closeness_u75_cawiki	1008	0.57
indegs_u75_frwiki	4109	0.41	indegs_u75_frwiki	4109	0.42	indegs_u75_frwiki	4109	0.39	closeness_u75_eswiki	2456	0.55
indegs_u_eswiki	2482	0.39	indegs_u75_eswiki	2456	0.39	indegs_u75_eswiki	2456	0.39	closeness_u75_nlwiki	1909	0.52
indegs_u75_plwiki	1716	0.38	indegs_u_dewiki	3552	0.39	indegs_u_dewiki	3552	0.37	closeness_u75_plwiki	1716	0.51
indegs_u_dewiki	3552	0.36	indegs_u75_plwiki	1716	0.37	indegs_u75_plwiki	1716	0.36	closeness_u75_itwiki	2530	0.49
indegs_u_ruwiki	2399	0.36	indegs_u_ruwiki	2399	0.37	indegs_u_nlwiki	1917	0.36	closeness_u75_svwiki	1889	0.49
indegs_u_nlwiki	1917	0.35	indegs_u_nlwiki	1917	0.36	indegs_u75_itwiki	2530	0.35	closeness_u75_ruwiki	2378	0.47
indegs_u75_itwiki	2530	0.33	indegs_u75_itwiki	2530	0.36	indegs_u_ruwiki	2399	0.35	closeness_u75_ptwiki	1348	0.46
indegs_u75_svwiki	1889	0.33	indegs_u75_svwiki	1889	0.34	indegs_u_cawiki	1021	0.33	closeness_u75_combined	7075	0.43
indegs_u75_ptwiki	1348	0.32	indegs_u75_ptwiki	1348	0.33	indegs_u75_svwiki	1889	0.31	closeness_u75_enwiki	5644	0.42
indegs_u_cawiki	1021	0.31	indegs_u_cawiki	1021	0.32	indegs_u_ptwiki	1369	0.29	closeness_u75_dewiki	3533	0.41

Table 5.13: Rank comparisons between Wikipedia and ULAN centralities

The alignment with the scholarly rankings resulted in different observations, shown in Table 5.14. Especially for Dilly's and Jensens' rankings, the global PageRank centrality rankings yielded higher correlation between the different language versions and the scholarly rankings than other centrality types. It was moreover interesting to see that in contrast to its correlation with the ranking from the original ULAN network, Jensen's ranking of 19th century artists was in quite high correlation with its global Wikipedia PageRank counterpart. Elkin's ranking in turn better agreed with ULAN-only PageRank rankings, but only with the ones calculated on the full set of links including those having birth date differences beyond 75 years.

Compared to the correlations between the centrality rankings from the original ULAN network and the scholarly rankings, the Wikipedia-based ranks agreed much better with their scholarly counterparts. This appeared quite reasonable, since both the bibliographic subject index count based rankings by Dilly and Elkins and the illustration counts collected by Jensen in fact represented a measure of artist popularity in art history literature, which appeared to be better approximated by Wikipedia references than by the more social-network-like structure of the ULAN. It was interesting to see, however, that PageRank, which takes the underlying network structure more into account than pure in-degree, in most cases performed better than the latter.

ext_elkinsrank			ext_jensenrank			ext_dillyrank		
pagerank_u_combined	50	0.71	pagerank_f_ptwiki	50	0.82	pagerank_f_ptwiki	135	0.61
degs_f_ptwiki	50	0.68	pagerank_f_eswiki	50	0.82	indegs_f_cswiki	130	0.6
pagerank_u_dewiki	50	0.67	pagerank_f_combined	50	0.82	pagerank_f_combined	144	0.6
pagerank_u_enwiki	50	0.66	pagerank_f_frwiki	50	0.81	pagerank_f_enwiki	143	0.6
indegs_u_itwiki	50	0.66	pagerank_f_ruwiki	50	0.8	pagerank_f_frwiki	140	0.6
pagerank_u_eswiki	50	0.65	pagerank_f_plwiki	50	0.79	indegs_u_svwiki	131	0.59
indegs_u_jawiki	50	0.64	pagerank_f_itwiki	50	0.77	indegs_f_plwiki	136	0.58
pagerank_f_ruwiki	50	0.62	pagerank_f_nlwiki	50	0.77	pagerank_f_nlwiki	132	0.57
pagerank_f_plwiki	50	0.62	indegs_f_cswiki	50	0.76	pagerank_f_eswiki	141	0.56
indegs_f_svwiki	50	0.62	pagerank_f_svwiki	50	0.76	pagerank_f_ruwiki	141	0.56
indegs_u_cawiki	50	0.61	pagerank_f_enwiki	50	0.76	pagerank_f_dewiki	142	0.56
indegs_f_cswiki	50	0.6	pagerank_f_dewiki	50	0.74	betweenness_u_itwiki	139	0.5
pagerank_f_nlwiki	50	0.58						
pagerank_f_ukwiki	50	0.57						
pagerank_u_frwiki	50	0.56						

Table 5.14: Rank comparisons between Wikipedia centralities and the scholarly rankings

5.6 Visualization of the ULAN-mapped Wikipedia network

Figure 5.18 shows a visualization of the giant component of the combined Wikipedia network, composed of 40,564 persons and 283,042 links between them. Persons are colored by nationality and their size reflects their degree within the network of ULAN biographies, while the links are colored with a gradient between the source and target colors. The tendency of persons of similar nationality to be more interlinked than those of differing origin became clearly visible through distinguishable, relatively homogeneously colored sections. Looking at the selection of labeled persons, derived from the previously used scholarly rankings and some arbitrarily chosen famous art history figures, revealed that there was a very coarse chronology present in the layout of the person nodes, where the lower left part appeared to represent persons from the Renaissance and the Baroque, the center occupied by representatives of (neo-)classicism and the upper right leading from the early modern artists to contemporary art. This sequence was closely related to the position of the different nationalities, mainly representing European countries reaching some form of "golden age" at various periods in history.

Despite of its overall rather tight packing, the cloud of interlinked persons relevant to art history featured some structures in its periphery looking like solar prominences. One of the most notable of these "eruptions" is located on the top of the Figure and represents two loosely interconnected groups of mainly Chinese and Japanese artists. A closer look at the persons there revealed that they represented artists from various periods, starting as early as in the 6th century A.D. and reaching well into the 20th century. The Chinese group contained two thirds of all Chinese persons in the Figure, while the Japanese group represented slightly more than a third of the Japanese persons, respectively. It was interesting to see these two "snapshots" of independent cultural histories and how they eventually merged with the grand narrative of art history. There were, however, also relatively detached local clusters of Western artists, such as a small group of Australian

20th century artists immediately to the right of the Chinese-Japanese group. Another interesting example was the tight triangle-like cluster mainly composed of British persons further to the left, elevated above the main group of British persons. As it turned out, this group was composed of female artists from the 18th and the 19th centuries, some of which actively sought to provide art education to women, such as the painter Henrietta Ward. In some cases, one person served as main link between otherwise mostly detached groups of persons and the main body of the network. This was the case for the prominent extension on the leftmost part, representing a group of ancient Greek painters linked via the British archaeologist Sir John D. Beazley, who was described to be the most important scholar on ancient Greek vase paintings.

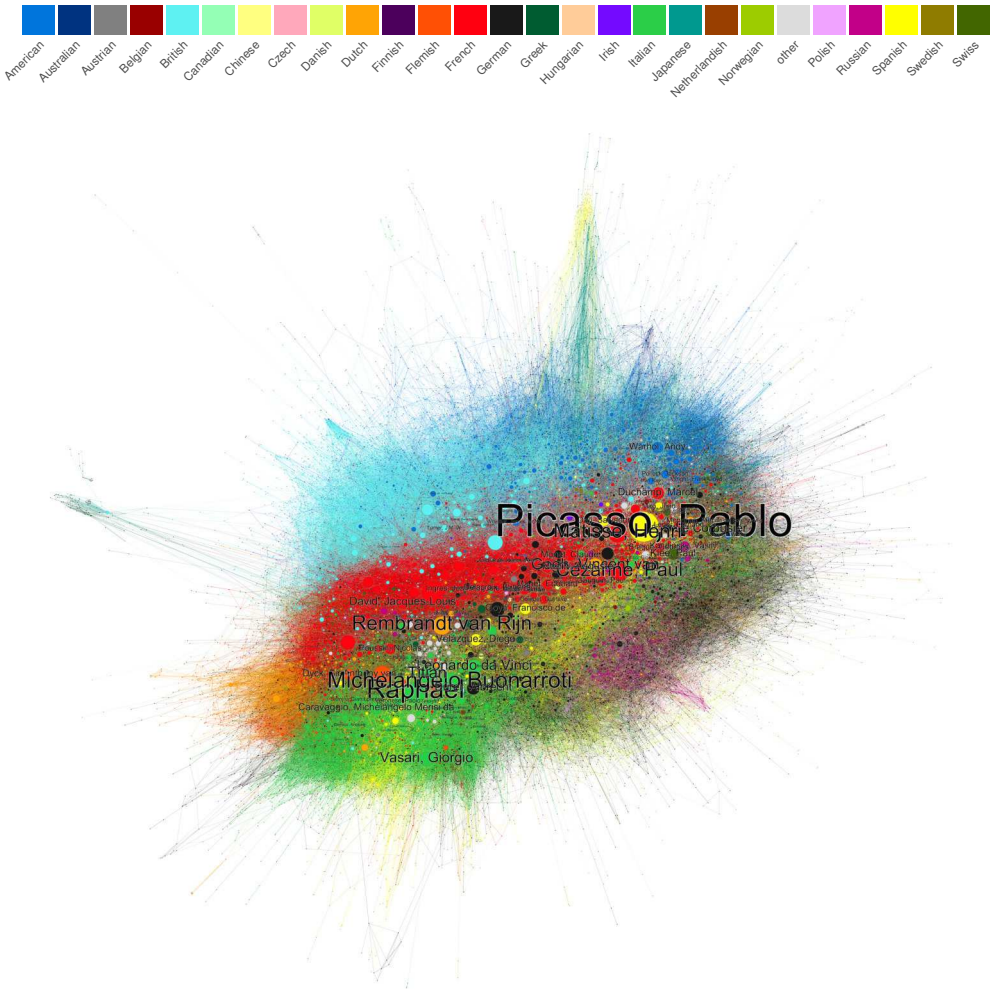


Figure 5.18: Visualization of the giant component of the combined Wikipedia network

5.6.1 Filtering the network by birth date difference

Since many Wikipedia biographies contained "long-distance" references to famous historical persons spread throughout the ages, it could be expected that the chronology embedded into the network of ULAN persons in Wikipedia would become more visible when pruning such links. As also shown for the correlations between the centrality rankings of the original ULAN and the Wikipedia hyperlink networks , the structure of the pruned Wikipedia network appeared more similar to the one of the original ULAN network. Using the previously introduced threshold to roughly separate between "long-distance" and "contemporary" relationships, all links exceeding a birth date difference of 75 years were thus removed from the network. The giant component of the resulting subset consisted of 39,316 persons and 239,793 links. Although 43,249 links (15.3%) were removed, only 1,248 persons (3.1%) from the giant component of the full network were not present in the giant component of the reduced version. As expected from the temporal distribution of ULAN biographies in general, featuring almost zero coverage for persons after Roman antiquity and before the late middle ages, the removal of long-distance links separated persons from antiquity from persons from the modern age, as it was also the case in the original ULAN network. This became evident when looking at the second-largest component of the pruned network, encompassing 275 of the 1,248 now detached persons and composed of 195 representatives from Greek antiquity and 64 from antique Rome, the remaining 16 persons being of various other cultures of that age.

The visual structure shown in Figure 5.19 represents the giant component of the pruned Wikipedia network of all language versions combined, with person nodes colored by their respective preferred nationality value. Similar to the visualization of the giant component of the original ULAN network in subsection 4.3.3, the Force Atlas 2 network algorithm was applied for creating its layout. The effect of removing long-distance links became immediately visible in that the resulting representation featured a much more exposed chronological structure. As discussed in Section 4.3.3, it was a direct result of the topology of the reduced network and made visible solely through the force-based layout algorithm without any assignment of node coordinates based on temporal attributes. Figure 5.20 shows the temporal structure of the layout through colors assigned to the person nodes based on their birth dates rounded to half centuries, revealing their relatively accurate chronological succession. It was interesting to observe, however, that the shown temporal sequence was highly non-linear and dependent on the density of persons born at specific points in time, which for example became visible in that the period from 1,100 to 1,500 A.D. occupied only about a third of the horizontal space in comparison to the years between 1,500 and 1,900 A.D.

Comparing the overall appearance of the reduced Wikipedia network in Figure 5.19 with the representation of giant component of the original ULAN network shown in Figure 4.18 revealed that the overall appearances of these two networks were relatively similar regarding the succession of national clusters therein, even though the two networks overlapped by only slightly more than 7,000 persons. The original ULAN network, however, was much less dense and thus seemed much more detailed in its representation of, for example, families, which appeared as small, highly interconnected groups. As

shown previously, however, these ties were to a large extent also covered in Wikipedia and the main difference between these two networks was thus, besides their differing person coverage, the many more hyperlinks present in the various Wikipedia networks, extending beyond immediate contact between persons and additionally covering influential or simply contemporaneous relationships.

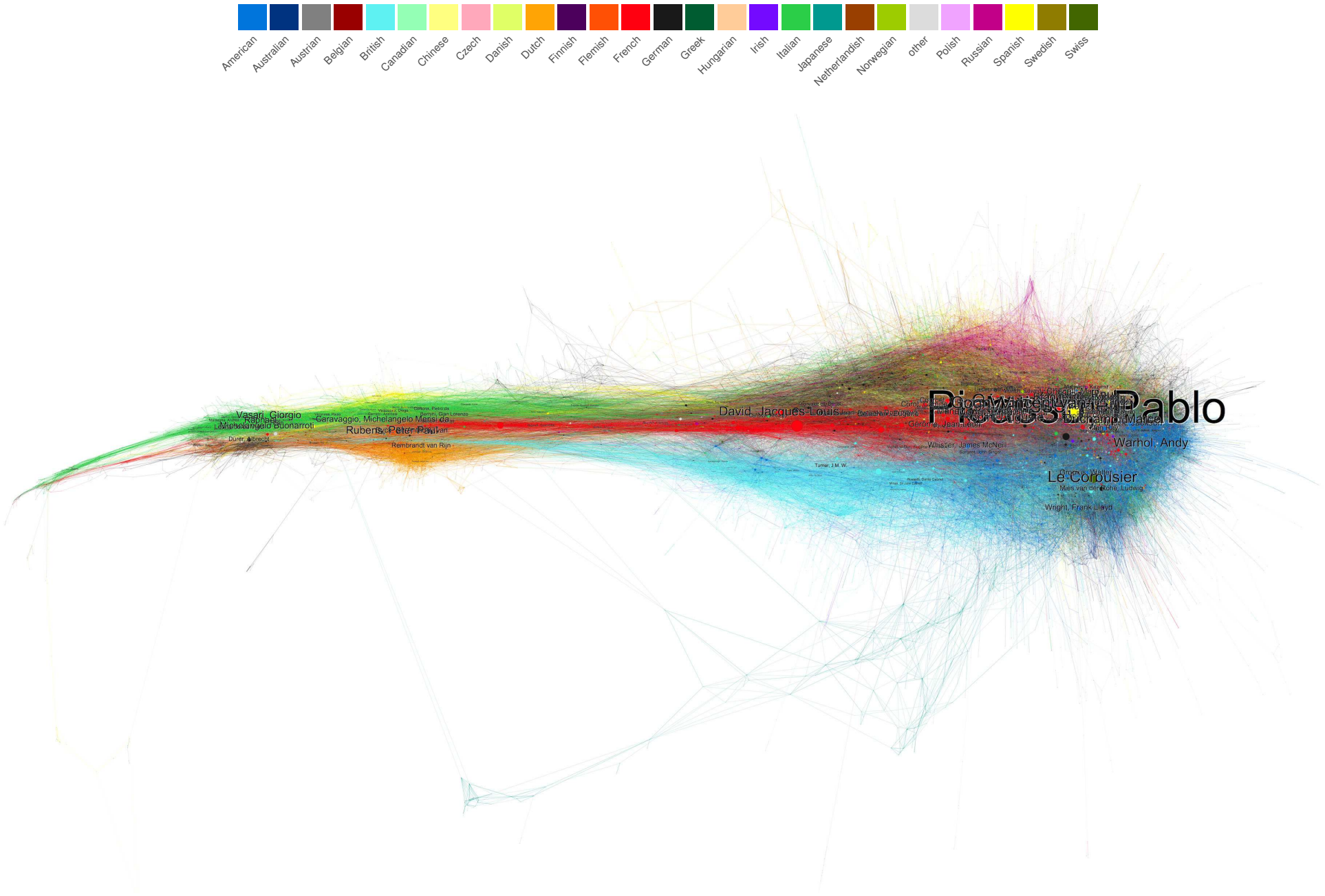


Figure 5.19: The giant component of the combined Wikipedia network without links exceeding 75 years birth date difference (**A4**)

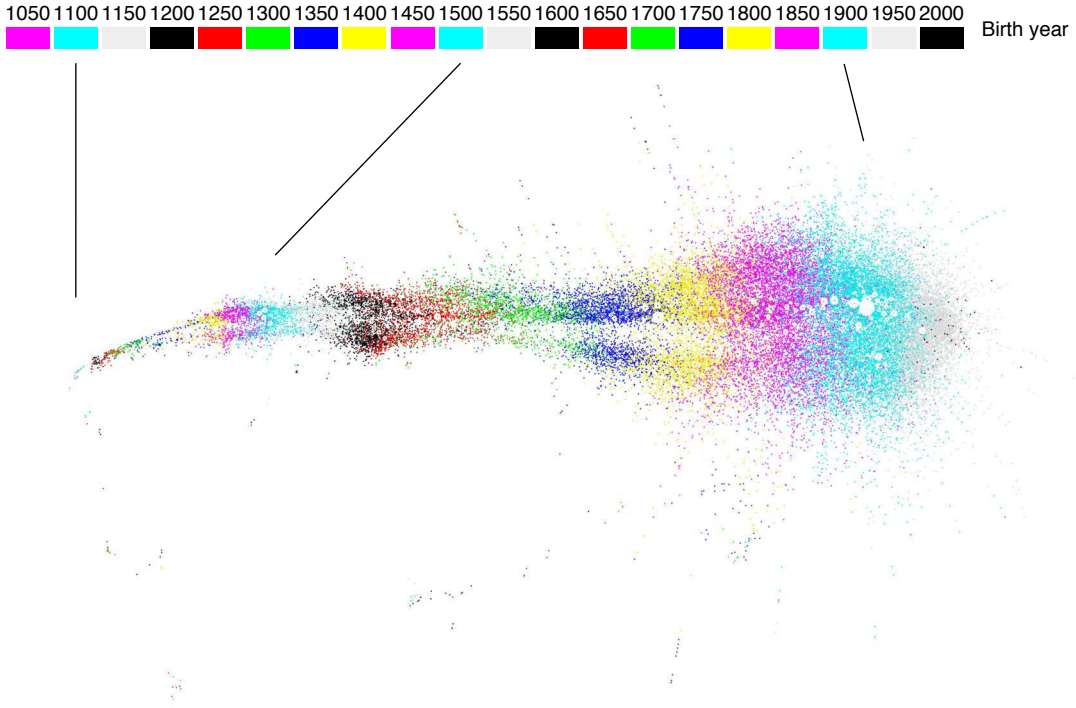


Figure 5.20: Pruned network colored by birth date rounded to half centuries

The plot starts on the leftmost side with a small group of persons born in the 11th and 12th centuries, representing clerics such Roman popes and a succession of German Holy Roman Emperors. The network was soon dominated by persons from 13th century Italy, mainly centered around clerical persons such as the Saint Francis of Assisi. He was connected to a number of different popes, saints and artists, the latter including some of the founding fathers of Western art such as Duccio and Cimabue, who both created depictions of him. Together with the painter Giotto, the latter marked the beginning of a far stretching strand of Italian artists which contained many of the famous representatives of the various stages of the Renaissance, such as Leonardo da Vinci, Michelangelo, Raphael and Titian, and the succeeding Baroque period, including Caravaggio and Guido Reni, until it appeared to fade out at around the age of the Rococo. From there it nevertheless continued in less dense form until the present day, with many well-known Italian representatives from modern art such as Amedeo Modigliani or the Futurists around the theorist and later Fascist politician Filippo Tommaso Marinetti, as well as post-WWII contemporary artists such as Michelangelo Pistoletto.

It was interesting to see a French strand of succeeding persons emerging at about the same time as the pioneers of Italian Renaissance art, running along in parallel beneath the Italians. In its beginning, however, it was mainly held together by the succeeding representatives of the Avignon Papacy and various French rulers, but soon also included French Gothic church architects, book illuminators and painters. The French strand of

art history too continued throughout the Renaissance and the Baroque periods, after which it visibly increased in coverage and link density, at about the same time as the Italian strand appeared to loose its dominance. This period of classic French art started with Nicolas Poussin and later included famous representatives such as Jacques-Louis David or Jean-Auguste-Dominique Ingres and appeared to be the most central cluster in the network of ULAN biographies mapped to Wikipedia, with many of its proponents having numerous national and international links. It continued until the second half of the 19th century, where the famous pioneers of Impressionism marked the onset of modern art whose break with the past was eventually sealed in the early 20th century by artists and theorists such as Marcel Duchamp and Andre Bréton.

Also artists from other nations or regions began to appear alongside the representatives from the French Gothic era. Most notably these were Netherlandish/Flemish, German and Spanish artists, each forming barely visible lineages which gained momentum around the beginning of the 15th century. Here, a relatively tight interlinked group of artists from the countries north of the alps marked the beginning of the Northern Renaissance, starting with representatives such as Jan van Eyck and Rogier van der Weyden and continuing into a dense cluster around the German artist Albrecht Dürer. The biographies of these three artists featured many in- and outgoing links to Italian Renaissance artists, reflecting the influence of the latter on the former. As previously stated for Dürer, none of these ties were present in the original ULAN network. Other important cross-cultural links concentrated around the important early artist biographers such as Giorgio Vasari with his already mentioned "Lives of the artists" and Karel van Mander with his "Schilder-Boeck", the first art history compendium created north of the alps.

Many international interlinkages of this era, however, also reflected the complex political interplay of the different powers struggling for hegemony throughout Europe which also took place in Italy, such as for example between France and Spain. This was also present in the network through many important sovereigns of that time serving as bridges between persons of different origin, one famous example for which being Francis I, King of France, the founder of the school of Fontainebleau and patron to many important artists such as Leonardo da Vinci. Another emerging struggle of that time were the events associated with the onset of the Reformation, eventually leading to the Thirty Years War but also being one of the root causes for the establishment of the independent Dutch republic at the height of the so-called Dutch Golden Age. In the Figure, this historical period became visible as densely linked cluster of Dutch artists during the Baroque era, including important figures such as Frans Hals, Rembrandt van Rijn or the artist-biographer Arnold Houbraken, whose "Schouburgh" represented an important source about the Dutch artists of these days. The confessional and political rift of this age also became evident in the visible separation between Dutch and Flemish artists, where the latter also tended to be more closely linked to Spanish and French persons.

Recalling Belting's statement from [Belting, 1992] that the reformation had profound impact on visual arts in Germany, it was also interesting to see that after Martin Luther, the dense cluster of German artists around Dürer, Cranach and Holbein quickly fainted, followed by a century long sparse presence of Germans. Only around the time of

the enlightenment, coinciding with the increasing ambition to establish a concept of a German nation and a national art, German artists reappeared as a densely interconnected group of persons, emerging from important early scholarly contributors to German art history, such as Johann Joachim Winckelmann and Johann Gottfried Herder. The artists of this time included Anton Raphael Mengs, Peter Cornelius, Wilhelm von Schadow and Caspar David Friedrich, representatives of German classicism and romanticism. Similar to the developments in France, however, German artists too increasingly turned towards modern art with Impressionists such as Max Liebermann and Lovis Corinth or Expressionists such as Franc Marc and August Macke, eventually turning to Surrealism and Dada with artist such as Max Ernst, but also to the specific German style of the "Neue Sachlichkeit", including proponents such as George Grosz and Otto Dix. The development of German art, however, appeared to come to a rather abrupt halt with the onset of the National-Socialists who declared many of the recent German developments in the arts to be "degenerate". It was interesting to see that the main figure behind these developments, Adolf Hitler, appeared in a very central position in the network and that many "original" Germans such as Walter Gropius, founder of the Bauhaus style, were featured as Americans in the ULAN due to their forced emigration. A cluster of German artists nevertheless reappeared after the Second World War, including famous representatives such as Joseph Beuys, Gerhard Richter and Sigmar Polke.

Also the Spanish group of persons appeared to feature different alternating periods of higher and lower density. A quite dense cluster formed during the time of the Spanish Golden Age at about the beginning of the Baroque period, including famous representatives such as El Greco (Actually having "Greek" as preferred nationality but being tightly interlinked with Spanish persons), Diego Velázquez, Jusepe de Ribera or Bartolomé Esteban Murillo. This period faded out with the demise of the Habsburg reign in Spain and continued with artists appearing around the onset of the reign of the House of Bourbon at the beginning of which a relatively dense group of artists was concluded by the painter Francisco de Goya. During the 19th century, another dense cluster formed, eventually leading to the famous Spanish modern artists such as Pablo Picasso, Salvador Dalí and Joan Miró. The network around the latter artists contained many other representatives of Modernism and it was especially interesting to observe that it appeared to be far more international than any other region in the plot.

Stepping back in time, another group of persons gained presence during the Dutch Golden Age. British persons began to emerge around the English King Henry VIII, who was connected to many important political figures of his time, including many Italian popes, the latter especially due to his abolishment of their authority in England. His biography was also interlinked with a number of international artists, most notable of which is the German painter Hans Holbein, who created a portrait of him, but also the Italian sculptor Jacopo Sansovino, who designed his tomb. Similar to the development of the French strand, relatively few artists were present amongst the "early" British persons, only at about the onset of the Baroque period they started to appear in higher numbers. It was interesting to observe that many of the British Baroque artists were connected to the Flemish painter Anthonis van Dyck, who almost appeared as another

root figure for the development of British art history as visible through the Wikipedia hyperlinks between the mapped ULAN biographies. Van Dyck's strong influence was indeed no coincidence, since he moved to London later in his life and became an important portraitist to the English court. In general, the development of British art appeared to be quite detached from other nations, especially from the French, although some highly interconnected individuals — court members and aristocrats, but also well known artists such as J.M.W. Turner or John Constable — served as hubs in this regard.

One newly emerging nation was nevertheless highly interconnected with British persons. Appearing during the 18th century, Americans had strong ties to their former colonizers and maintained them throughout the network. Not surprisingly, the American strand of art history started with the founding fathers of the American nation such as Benjamin Franklin, George Washington and Thomas Jefferson. They were connected to a number of early American artists who painted portraits of them, such as Benjamin West, Gilbert Stuart or Charles Willson Peale, the former two of which also spent periods of their lives in London and were thus connected with many British persons in the network. Especially during the second half of the 19th century, however, there were increasing links between American and French artists which soon outnumbered those to British persons, while as shown in the right of the plot, a relatively contiguous group of American artists confirmed that this nation played a significant role in the arts after the redrawing of the cultural landscape due to the aftermath of the Second World War, including famous artists such as Jackson Pollock and Andy Warhol.

On top of the plot, a group of Russian persons started to emerge around the first emperor of the Russian empire, Peter I the Great. Especially during the 19th century, this group formed a relatively tight and contingent cluster, containing famous representatives such as the Romanticist Karl Pavlovich Bryullov, the Realist Ilja Jefimowitsch Repin and the Impressionist Konstantin Alexejewitsch Korowin. It was interesting to observe how two different strands appeared to emerge from this cluster. One was well connected to the other representatives from different cultures, containing many of the famous modern artists such as Wassily Kandinsky, Kasimir Malevitch, El Lissitzky and Alexander Rodtschenko. The other strand seemed to follow a completely different track and also broke away from the overall chronological ordering in the plot, in which it was visible as a triangle like peak moving away to the top of the Figure. This group mainly featured 20th century representatives from Soviet Social Realism, such as the proponents of the Leningrad school of painting including Nikolai Efimovich Timkov and Yuri Mikhailovich Neprintsev and Evgenia Petrovna Antipova and it was interesting to see how detached it was from the remaining persons.

Another group of persons was even more detached than the proponents of 20th century Soviet art. Japanese persons appeared as early as during the 16th century and were initially linked to European cultures via only one person, namely pope Gregory XIII, who was known to have had diplomatic relations to Japanese officials. He was linked to the Japanese artist Kanō Eitoku via a hyperlink only present in the Japanese Wikipedia, stating that one of his works was presented to the pope as gift and subsequently lost

after his death.[23] Kanō Eitoku was a well known representative of the so-called Kanō school of painting[24], to which also many of the other featured Japanese artists of this time belonged. Additional ties to Western nations appeared only later, for members of the following Ukiyo-e school, famous representatives of which included the painter Katsushika Hokusai, an important influence to many Western artists. Closer contact to representatives of other nations, however, did not take place before the Meiji period beginning in the second half of the 19th century, during which two distinct branches of Japanese arts emerged, namely the Western oriented Yōga and the rather traditional Nihonga styles. It is interesting to note that the latter were actively proposed by the American art historian Ernest Francisco Fenollosa in order to preserve the original artistic Japanese styles, working as scholar and museum director for the Japanese government and serving as bridge between Japanese and Americans in the Wikipedia network.

With the subnetwork of Japanese artists and related persons as most detached example, all of the described nation-specific subsets featured much more links between members of the same culture than to representatives of any other culture. Figure 5.21 provides an exploded view on these subnetworks as they appeared in the combined layout from Figure 5.19, which is again included for reference. Due to its already exposed position at the bottom of the combined layout, the Japanese subnetwork is not redrawn individually in the Figure. Drawing the network in this form provided a better illustration for the underlying nation-specific "art histories", since the combined layout contained many overlapping regions which were not easily distinguishable from each other. Especially for the German, Italian and Dutch subnetworks it moreover became visible that at the right side of the plot, in more recent times, their individual structures appeared to separate into two relatively distinct groups. As it turned out, the lower branches each represented dense clusters of mainly architects, while the upper branches contained more artists. In the combined view of Figure 5.19, the different architect clusters merged together into a dense, internationally interlinked architect cluster at the bottom right hand side of the structure, close to the position where the Japanese group of artists merged with the main body of the network.

In the immediate vicinity of the architect group, other professions were present as well, especially designers and photographers, but also filmmakers and theater directors were notable in this regard. Interestingly, they in many cases appeared to be located right between the artists and the architects which could be rooted in the simultaneous role of photography as fine arts medium, means of architecture documentation and medium for general advertising, or the role of stage design in the performing arts, which includes both elements from fine arts and architecture.

[23] https://ja.wikipedia.org/wiki/%E7%8B%A9%E9%87%8E%E6%B0%B8%E5%BE%B3, retrieved Sept. 14th, 2020, translated via Google translate. No other source for this information could be identified via an Internet search.

[24] http://www.metmuseum.org/toah/hd/kano/hd_kano.htm, retrieved Sept. 14th, 2020

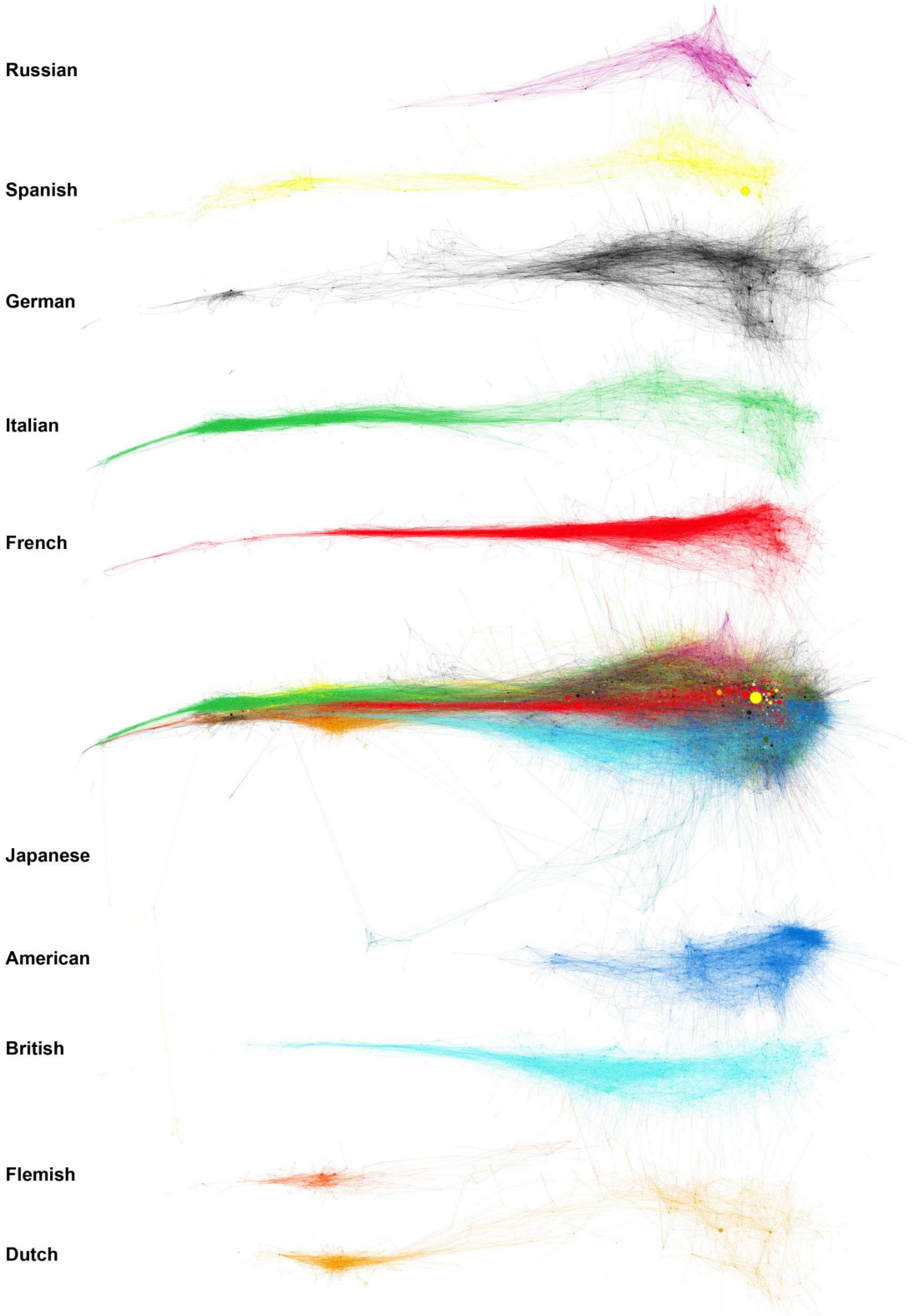


Figure 5.21: Sections by nation in the GCC of the combined, pruned Wikipedia network

The shape of the reduced network of hyperlinks between ULAN biographies in Wikipedia visualized the concurrent development of the arts in different, mainly European, cultures and also provided insight on their mutual interactions and dependencies. It allowed a bird's eye view on a specific version of art history distilled from many different sources, edited by a multitude of Internet users drawing from textbooks, newspaper articles and other repositories. It was particularly interesting to see how different European nations appeared to play interchanging, sometimes more, sometimes less important roles in the arts and how these developments appeared to be closely linked to the political and cultural events that took place throughout history.

An additional aspect of the reduced layout was its tapered form, growing thinner the further one looked into the past. This was mainly due to the decreasing coverage of historical persons in the ULAN and in Wikipedia, but also representative for the overall notion that only a limited number of historical entities — using Aleida Assman's terminology from [Assmann, 2008] — made it into the archive or even into the canon of (art) history, while the remaining entities increasingly became forgotten with time. It was interesting to compare the reduced shape to the shape of the full network in Figure 5.18 in this regard, which, although exhibiting some minor chronological structure, did not appear to feature this "historical reduction". As shown previously, however, this was based on the relatively small proportion of links with "long temporal distance", connecting a limited number of timeless "heroes" of art history up to the present which, again in Assman's words, highlighted the function of the canon for preserving the past as present.

A striking analogy became evident when comparing the pruned shape and the composition of national clusters therein with Alfred H. Barr's Torpedo Diagrams from [Barr, 1933], originally conceived to visualize the ideal composition of the collection of the Museum of Modern Art in New York City at various points in time, shown in Figure 5.22. Based on a visual metaphor of a torpedo-like shape floating through time, these diagrams featured temporally changing subdivisions containing different sections dedicated to the artistic contributions of different world regions or nations. While the tip of the torpedo was thought to represent the contemporary spearhead of the current development of Modern art, its tapered tail was conceived as the point of transition during which previously Modern art became "old art" and therefore should be sorted out from the collection and moved to other museums with a different focus instead. It was interesting to see how the different versions of the diagram increasingly represented American art to cover the "avant-garde" section of the torpedo and that the remaining space was prominently dedicated to French art, which appeared as the driving influence for its overseas "successor", while only little space and influence was granted to other nations. The diagram could thus be seen as representing a specific sub-canon of Modern art as it was conceived by an American museum dedicated to that period, highlighting specific aspects of art history deemed most important to the specific agenda by selecting appropriate content from the archive of art history. It was particularly interesting to see that in the visualization of the pruned Wikipedia network, it was again Americans who appeared to represent the largest contiguous cluster of persons at contemporary times, tightly interconnected with French persons preceding it.

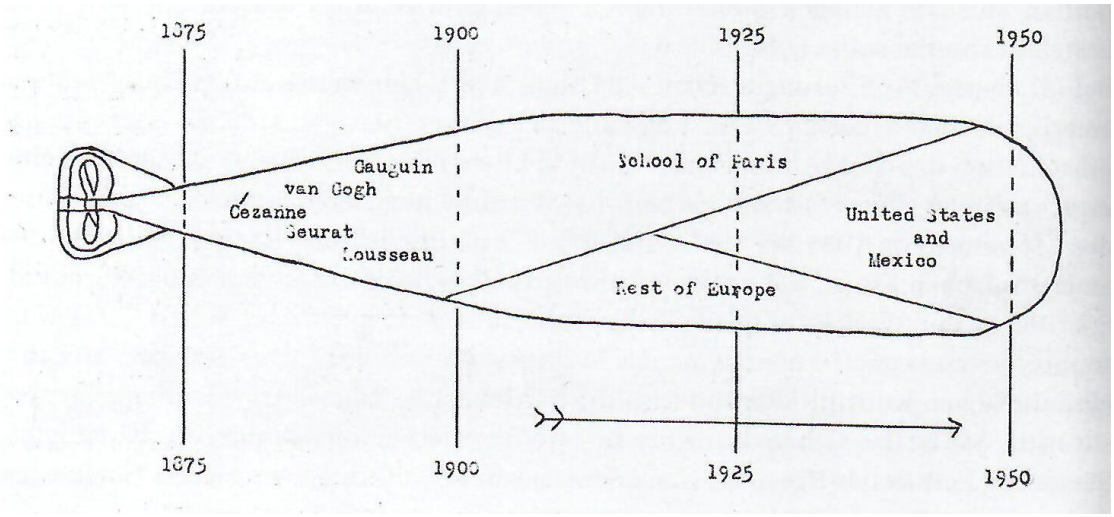


Figure 5.22: Torpedo Diagram, Alfred H. Barr, Jr.

The analogy between Barr's torpedo and the reduced Wikipedia network of ULAN biographies was likely based on the high proportion of British and American persons in the ULAN and the many links between Americans and French people. Although the focus of the Wikipedia visualization was much broader than the Modern art centered view in the torpedo diagram, it nevertheless represented another form of art history sub-canon, based on the collections of the many institutions participating in the Getty Vocabulary Program, which to a large extent came from the English-speaking world.[25].

5.6.2 Visualizing international interactions across time

As far as international relationships across time were concerned, it was possible to create a time-based visualization of the nation to nation interactions in the pruned Wikipedia network similar to the one created for the original ULAN network in Section 4.4, shown there in Figure 4.23. For the Wikipedia version, it was extended for comparing the ULAN network with its combined Wikipeda counterpart as well as the individual Wikipedia languages with each other. Figure 5.23 shows the resulting matrix of nation to nation interactions each represented as time series whose amplitudes at each time step represent aggregations of the individual links between persons of the respective

[25] As seen during the analysis of the original ULAN network, the ULAN person records featuring associative links appeared to be deemed as historically more important than those without, based on the comparison with the three scholarly rankings. This separation could again be interpreted based on Aleida Assmann's framework by conceiving the complete ULAN as the archive and its subset for which additional information existed in form of associative links as the canon, respectively. In this sense, persons not present in the ULAN would belong to the "forgotten" group which would strongly differ when considering other data sources from different cultural contexts.

nationalities by the mean birth date of each pair of interlinked persons, rounded to decades. The upper triangle of the matrix features the interactions in the original ULAN network alongside those in the pruned Wikipedia network combined from all language versions, while the lower triangle juxtaposes the pruned networks for a selection of individual Wikipedia languages. In order to make them better comparable, the counts for the Wikipedia networks were based on the versions with mirrored hyperlinks, thus ignoring their direction. The number shown for each nationality combination represents the total link count for the time-series having the highest total number of links for that combination. In the upper triangle, it is accompanied by the Pearson correlation between the Wikipedia and the ULAN time-series, while color-codes in the lower triangle identify the language version having the most hyperlinks.

Considering the comparison between the ULAN and the combined Wikipedia networks in the upper triangle, the Pearson correlation measures between the time-series for each nationality combination suggested that the two link-sets corresponded with each other to quite an extent. This especially became visible for the same-nation interactions between American, British, French, Flemish and Russian persons, where the Wikipedia and the ULAN link counts appeared to be quite similarly distributed across time. In other cases, however, the shapes of the time-series deviated from each other especially with respect to the counts for more recent links, such as for the same-nation links between persons from Italy, Germany, the Netherlands and Spain, for which the ULAN and Wikipedia corresponded quite well for the Renaissance or Baroque periods, but Wikipedia featured many more recent same-nation links. As visible in Figure 5.24 comparing the counts of biographies with links to others across time between the ULAN and the combined Wikipedia networks, both datasets appeared to feature relatively similar counts until about 1750, from when on many more Wikipedia biographies featured links to others.

Looking at the different time series for the combined Wikipedia network revealed that they, at least as far as interactions within the individual nations were concerned, quite accurately reflected some of the observations previously described on network level and shown in Figure 5.21. Interactions between Germans for example featured a clear peak around the time of Dürer, declined afterwards and reached new heights from about 1700 on. Similarly visible was the steady decline of Italian-Italian interactions after the Baroque period, which also gained presence again during the 19$^{\text{th}}$ century, a development even stronger visible for Dutch interactions after the Golden Age. Links between Americans and also between Japanese persons on the other hand showed the most recent peaks, located in the 20$^{\text{th}}$ century.

The comparison of the individual Wikipedia language versions in the lower triangle of Figure 5.23 in turn revealed some interesting insight into their individual contributions to the combined Wikipedia network. As shown before, the different Wikipedia languages in the Figure featured clearly higher counts for links between culturally related nationalities, which was the case for all but the Flemish same-nationality interactions and for many of the international ties as well. The temporal distributions, however, revealed that these higher counts were again often based on links between persons of rather recent birth. Taking the time-series of Dutch-Dutch interactions as an example, it became visible that

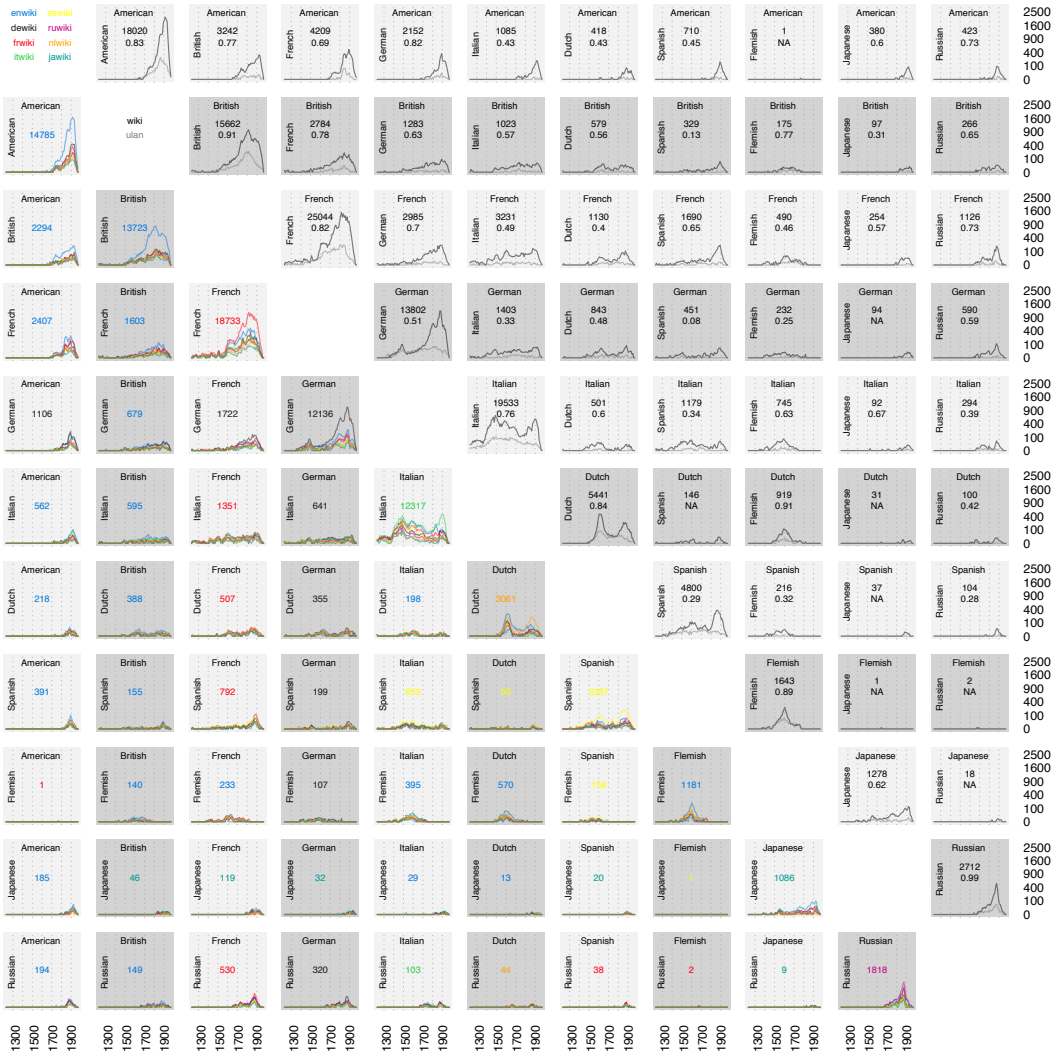


Figure 5.23: Time-Series of hyperlink counts based on mean birth dates of links

while the Dutch Wikipedia clearly featured the highest link counts for interactions taking place after about 1800, those during the Dutch Golden Age had higher coverage in the English Wikipedia. This observation also applied to other combinations of nationalities, where ties in the "early days" often had much higher overall coverage in multiple language versions than those happening at more recent times. An interesting exception in this regard appeared for Spanish-Spanish ties, for which the Spanish Wikipedia appeared to have the highest coverage throughout the ages. This observation nevertheless suggested some form of global agreement on the canonical status of many of the featured Renaissance and Baroque artists across the different Wikipedia versions, while more recent persons seemed to be less acknowledged beyond their immediate cultural sphere.

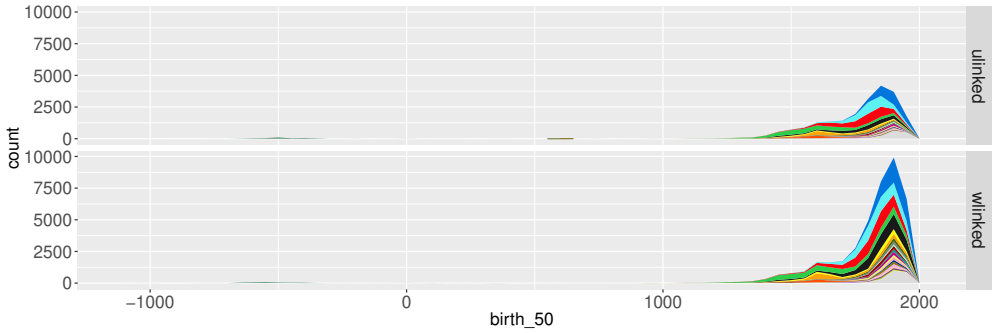


Figure 5.24: Linked biographies by nationality and time in the ULAN and in Wikipedia

5.6.3 Visualization of contributions from individual Wikipedia editions

The observations regarding the temporal differences in nationality coverage between Wikipedia language versions encountered in Figure 5.23 suggested to create and compare language specific visualizations of the pruned Wikipedia network. Figure 5.25 shows these visualizations for the Wikipedia language versions considered in Figure 5.23, each language version using on the same layout calculated for the combined Wikipedia network but omitting all nodes and links not present there. This way, the different coverages could be effectively compared with each other. Not surprisingly, the English Wikipedia appeared to have the highest visual similarity with the representation of the combined network, which was on the one hand due to its higher coverage of British and American persons who represented a significant proportion of the overall ULAN biographies, but on the other hand also based on the high international coverage of the English version. The latter was less expressed in the other language versions, which highlighted their specific focus on culturally related individuals even more. As already indicated by the time-series representations from Figure 5.23, this was especially the case for more recent sections of the respective networks, which, starting at about the time of the enlightenment, increasingly appeared to be quite different from each other, while the earlier periods, although also differing in coverage, showed much higher agreement. Comparing the individual network visualizations to the national clusters in the exploded view from Figure 5.21 again highlighted each language version's individual, culturally specific contribution to the combined network.

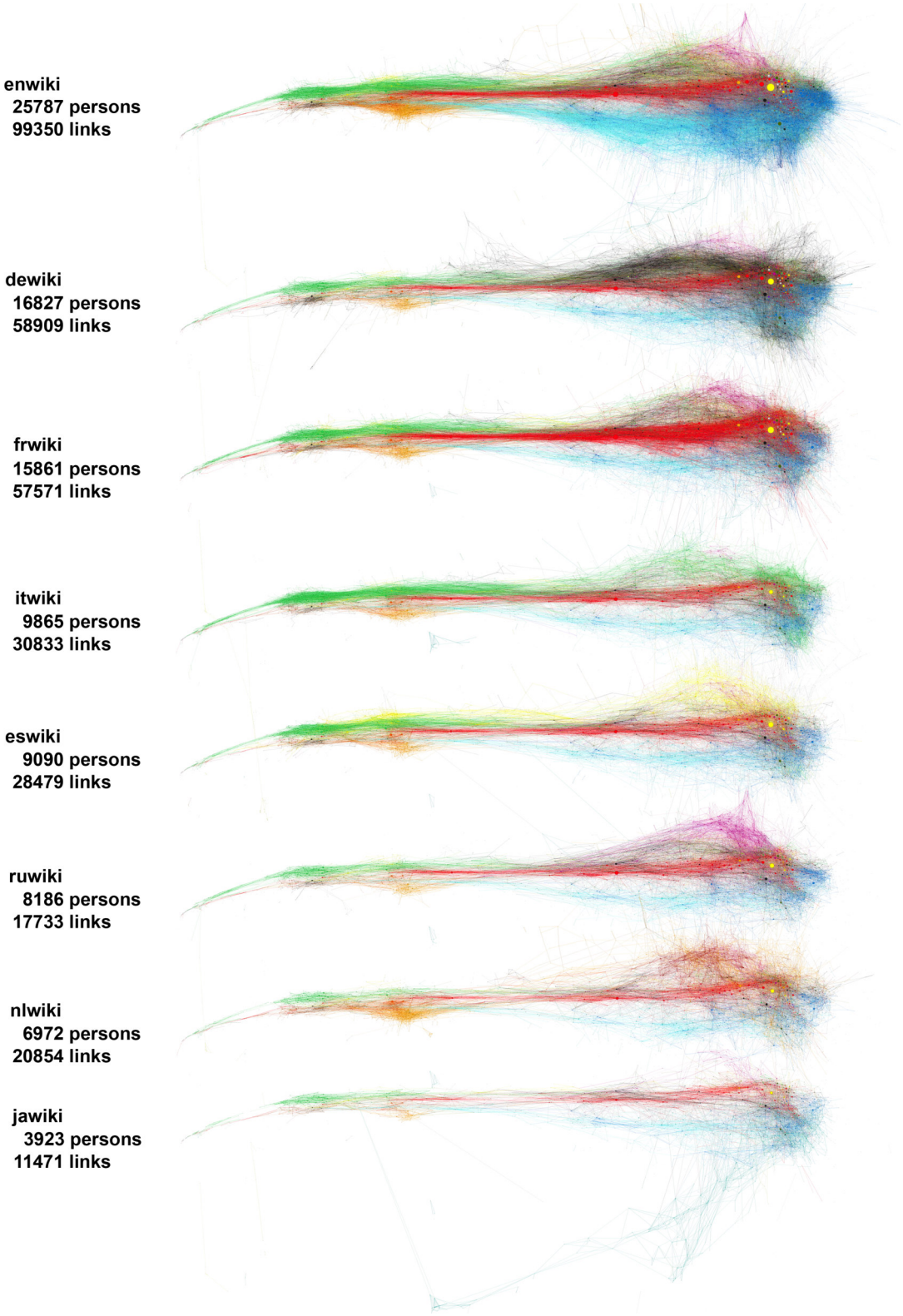


Figure 5.25: ULAN biography networks in different Wikipedia languages (**A5**)

5.7 Network of aggregated links between nationalities

The visualization of the individual person ties in the pruned Wikipedia hyperlink network shown in Figure 5.19 revealed strong interactions between different nationalities across time. The temporal aspects could be even more highlighted by plotting the aggregated mean-birth dates of pairs of interlinked persons for selected combinations of nationalities across time as shown in Figure 5.23. Another potentially interesting view was to consider the totality of interactions between nationalities regardless of the specific time they took place in history. This could be done in form of an aggregated network where the different nationalities represented the nodes which were connected via links weighted by the respective number of individual person links that existed for each linked pair of nationalities, discarding nation-nation links with zero weight. Figure 5.26 shows the resulting network derived from the combined Wikipedia hyperlink network. In line with their extraordinary presence in the ULAN, the few "top-nationalities" accumulated the highest number of hyperlinks, which became visible in their central position in the network and the high weight of the aggregated ties between them. This was especially the case for the French nationality which occupied the most central position of all, corresponding with the observed position of French persons in the network, ranging from the Baroque to Modernity. "Self-loops" had the highest weight, since most of the hyperlinks connected persons of the same nationality.

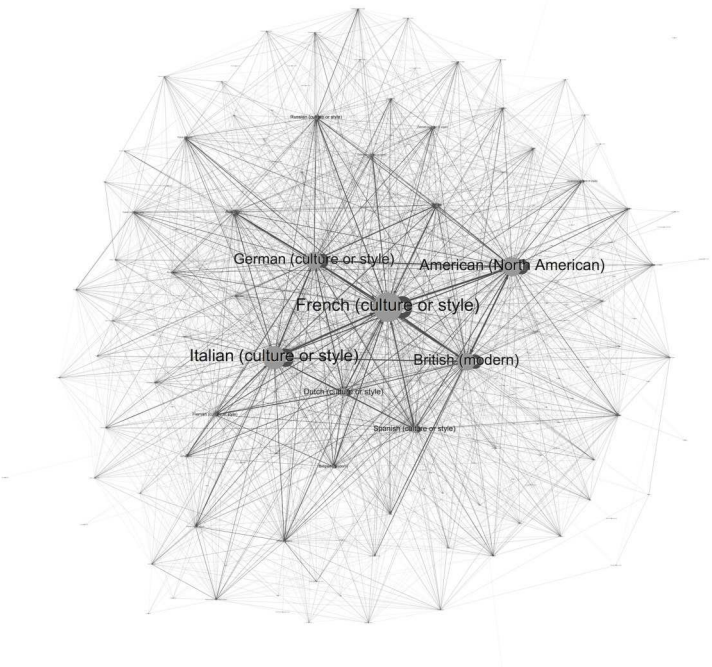


Figure 5.26: Network of aggregated nation-to-nation links

5.7.1 Deriving a relative measure for nation-to-nation ties

Up to this point, the direct and derived visualizations of the Wikipedia network had in common that they were based on absolute person and link counts, which, when considering nationalities, privileged those with many representatives in the ULAN, while the relationships with or between persons of less prominent origin were often "masked" by these large-scale interactions. It was thus desirable to seek for alternative representations which were able to emphasize "interesting" relationships regardless of the underlying absolute counts. As far as nationalities were concerned, this could be achieved by deriving a relative measure for quantifying the aggregated linkage for any pair of them.

One immediate, straightforward approach was closely related to the Chi-Square test of independence of two categorical variables, described in detail by Agresti in [Agresti, 2007], and based on the arrangement of the raw counts of each nation-nation combination in a two-dimensional $I \times J$ contingency table, such as shown in the toy example in Table 5.15. The top rows of the table's cells represent the observed aggregated counts of links between persons belonging to the respective nation pairs, their row/column sums are shown at the rightmost column and at the bottom row of the table, yielding marginal distributions for the frequencies of nationalities occurring as source or target of a Wikipedia hyperlink between two ULAN persons. As described in [Agresti, 2007], these marginal sums were used to calculate expected link counts for each row/column nationality combination, shown on the left side of the bottom row in each cell. Similar to the individual cell contributions to Chi-Square tests, the relative measure for the importance of each nationality pair was then based on the difference between observed and expected values, but used here in standardized form according to Agresti's definition of "standardized residuals" from [Agresti, 2007], essentially representing a z-score expressing deviation between observed and expected values in terms of standard deviations, shown on the right side of the bottom row of each cell. Negative values suggested that the particular nation combination appeared less than it would have been expected based on the overall occurrences of the involved nations, positive values the opposite, respectively.

In the example Table 5.15, deviations above two are considered to indicate significant importance for the specific nation pair and thus highlighted there. In line with the observations for both the ULAN and the Wikipedia networks, "same nation" links have the highest share, which becomes visible in the example table's diagonal cells. Besides these rather obvious cases, the highlighted interaction between "nation 3" and "nation 4" shows an example for how the transformation also enabled the detection of significant deviations for low aggregated link counts.

The procedure outlined for the example table was applied to the combined Wikipedia language network of mapped ULAN biographies, resulting in a nation to nation contingency table containing 34,710 (195 source and 178 target nationalities) cells. In many cases, nationality combinations had only very low or zero links counts and the presence of even single ties was thus sometimes shown as very significant standardized deviation. In order to filter out such potential noise, only 12,932 ties between 123 source and 107 target nation values having observed counts of at least five links were considered. Moreover, 12,420 of them had negative residuals and could be discarded as well, since only

	nation 1	**nation 2**	**nation 3**	**nation 4**	**nation 5**	
nation 1	2000 1112.08 **52.09**	100 565.57 -31.59	50 175.98 -13.83	25 108.52 -11.49	100 312.85 -18.12	2275
nation 2	100 565.57 -31.59	1000 287.63 **55.90**	25 89.50 -8.19	12 55.19 -6.87	20 159.11 -13.70	1157
nation 3	50 175.98 -13.83	25 89.50 -8.19	250 27.85 **45.63**	25 17.17 **2.02**	10 49.51 -6.29	360
nation 4	25 108.52 -11.49	12 55.19 -6.87	25 17.17 **2.02**	150 10.59 **44.99**	10 30.53 -4.10	222
nation 5	100 312.85 -18.12	20 159.11 -13.70	10 49.51 -6.29	10 30.53 -4.10	500 88.01 **50.92**	640
	2275	1157	360	222	640	4654

Table 5.15: Contingency table for Wikipedia links aggregated by source and target nation

nation-nation interactions having higher counts than expected were of interest. Applying a threshold of three positive standardized deviations from expectation and removing self-loops, the remaining 512 ties were further reduced to 299 interactions between 123 distinct nationalities.

Visualizing the remaining 299 interactions with significant positive residuals as a network, shown in Figure 5.27, revealed that it had a clustered structure featuring highly interconnected groups of nationalities, which could be confirmed by applying the Louvain community detection method described in [Blondel et al., 2008] as implemented in the Gephi[26] network visualization platform. Based on the maximization of the modularity value, defined in [Newman, 2006] on a minus one to plus one range for indicating the clustering present in a given partition of an observed network, the community detection yielded a partition of the 123 interlinked nationalities into 11 groups, colored accordingly in Figure 5.27. The identified partition had a very high modularity value of 0.736, its 11 groups thus featuring members highly interlinked within each group, but not across different groups. Table 5.16 complements Figure 5.27 by grouping the names of clustered nationalities for better readability.

The identified groups represented relatively distinct clusters based on varying combinations of geographical, cultural and temporal ties. Common language appeared to be a strong factor in this regard, especially visible for the French (cluster #7, bottom center, purple), Spanish (cluster #5, lower right, blue), and English (cluster #2 top center, black) dominated clusters, which also highlighted the underlying colonial histories of these nations, spanning different geographical regions of the world. An interesting art historically relevant tie between the French and the Spanish groups was related to the Kingdom of Naples, referred to as "Neapolitan" in the ULAN, which was identified as part of the Spanish cluster due to many documented ties based on its Spanish-dominated rule, but also linked to the French in this regard.

Other European centered regional ties appeared to be constitutive for the Scandinavian (cluster #10, center, grey), Dutch (cluster #8, left, dark grey), German (cluster #11, lower left center, orange), Southern Slavic (cluster #9, lower center, light blue) and Russian/Soviet (cluster #1, lower center, green) groups and it was interesting to see

[26] `http://www.gephi.org`, retrieved Sept. 14th, 2020

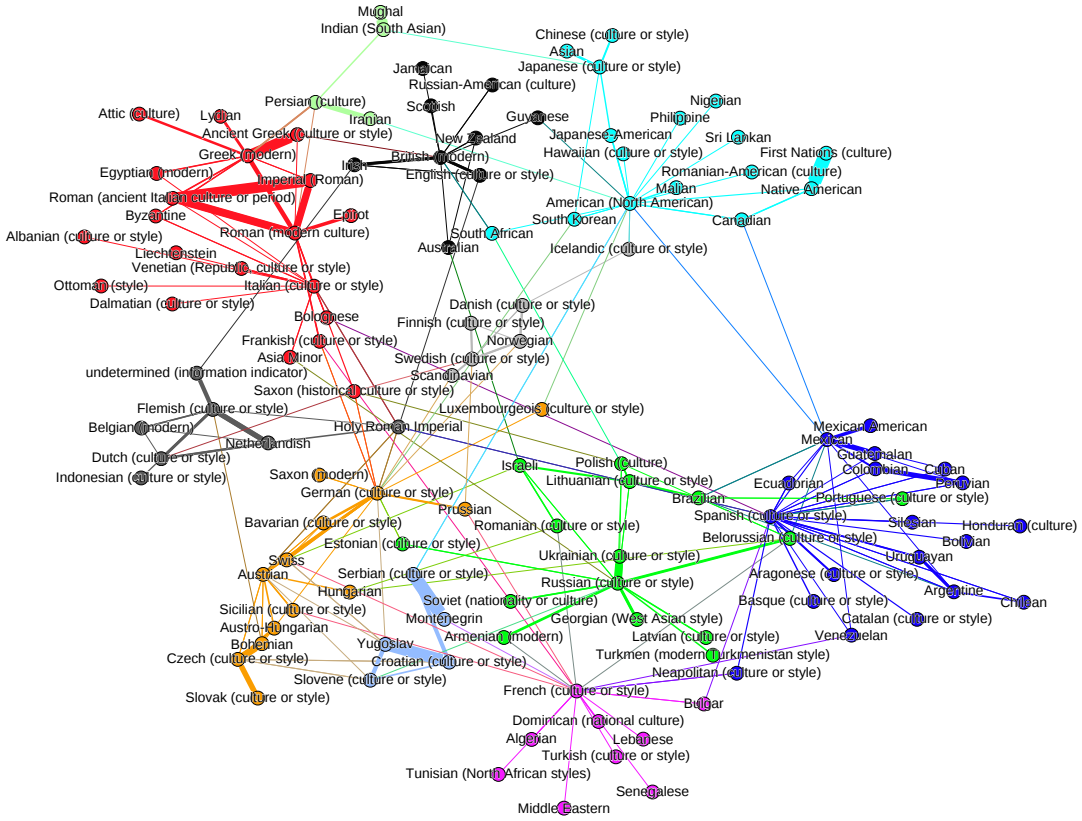


Figure 5.27: Residual network of aggregated nation-to-nation links (**A6**)

especially the latter three to include nationalities formerly incorporated into larger state organizations, such as the successor nations to the Habsburg empire, those to the Yugoslavian state and former USSR/Warsaw pact countries. Spanning even longer durations, some of the clusters also linked ancient and middle age "nationalities" to their modern counterparts, such as the group including the Italian and the Greek nationalities (cluster #6, top left, red) closely interlinked with ancient Greek, Roman and Egyptian cultures, or the small group consisting of Persian/Iranian and Indian/Mughal nationalities (cluster #3, top center, light green). Having at least one endpoint with low counts in the ULAN, these relationships tended to have particularly high residual values, visible as thick lines in Figure 5.16 which also appeared between other less represented entities such as "Native American" and "First Nations", highlighting the sensitivity of the filtering approach towards such low counting cases.

Overall, the composition of members in the encountered clusters revealed quite high levels of homogeneity with respect to common traits such as language, location or culture. Although many clusters featured a few members whose presence appeared less appropriate,

there were clearly explainable reasons for the co-occurrence of the majority of group members in each cluster and the individual "outliers" most likely due to specific ties present between the observed ULAN persons. The group of Asian and North American (cluster #4, top right, cyan) nations, however, was the least homogeneous, which on the one hand suggested that the community detection algorithm was not able to separate all the groups potentially present in the dataset, on the other hand highlighted other processes such as migration to be partially responsible for the resulting clustering.

1	Armenian	4	American	6	Albanian	8	Belgian
	Belorussian		Asian		Ancient Greek		Dutch
	Brazilian		Canadian		Asia Minor		Flemish
	Estonian		Chinese		Attic		Holy Roman Imperial
	Georgian		First Nations		Bolognese		Indonesian
	Israeli		Hawaiian		Byzantine		Netherlandish
	Latvian		Japanese		Dalmatian		undetermined
	Lithuanian		Japanese-American		Egyptian	9	Croatian
	Polish		Malian		Epirot		Montenegrin
	Portuguese		Native American		Frankish		Serbian
	Romanian		Nigerian		Greek		Slovene
	Russian		Philippine		Imperial		Yugoslav
	Soviet		South African		Italian	10	Danish
	Turkmen		South Korean		Roman		Finnish
	Ukrainian		Sri Lankan		Roman		Icelandic
2	Australian		Romanian-American		Saxon		Norwegian
	British	5	Aragonese		Venetian (Republic, culture or style)		Scandinavian
	English		Argentine		Liechtenstein		Swedish
	Guyanese		Basque		Lydian	11	Austrian
	Irish		Bolivian		Ottoman		Austro-Hungarian
	Jamaican		Chilean	7	Algerian		Bavarian
	New Zealand		Colombian		Bulgar		Bohemian
	Russian-American		Cuban		Dominican		Czech
	Scottish		Ecuadorian		French		German
3	Indian		Guatemalan		Lebanese		Hungarian
	Iranian		Honduran		Middle Eastern		Luxembourgeois
	Mughal		Mexican		Senegalese		Prussian
	Persian		Mexican American		Tunisian		Saxon
			Neapolitan		Turkish		Sicilian
			Peruvian				Slovak
			Spanish				Swiss
			Uruguayan				
			Venezuelan				
			Catalan				
			Silesian				

Table 5.16: Detected community structure in the pruned nation-to-nation residual network

The clusters identified in the residual based nationality network derived from aggregated Wikipedia hyperlinks between ULAN persons and its overall structure revealed interesting associations with existing related Wikipedia research, although performed at a very different level: Samoilenko et al. in [Samoilenko et al., 2016] analyzed a network of Wikipedia language editions based on counting edit activities for articles about the same topic in different languages. Shown in Figure 5.28, it also featured dense clusters of culturally related Wikipedia language versions, some of which, such as the groups of

Scandinavian language versions and of those from the former USSR, were remarkably similar to their counterparts in the nationality network.

Being focused on languages, however, the identified commonalities with the nationality network were limited to entities having distinct nationality and language, not featuring its many relationships between nationalities speaking the same language. Moreover, the Wikipedia language network reflected a map of relationships amongst contemporary world cultures, while the nationality network was extracted from biographical relationships spanning more than four millennia, including a number of entities which had ceased to exist many years ago. This aggregated, total view on the compound of multi-temporal ties was nevertheless based on contemporary Wikipedia content, thus subject to current modes of filtering such as the previously shown preference for own cultures in related Wikipedia languages, but also more general perceptions of the world and its history as present in the many sources used for the analyzed Wikipedia biographies. A structured analysis of the derived nationality network in light of international editing behavior would be an interesting path for future research, especially also when including an analysis of the referenced sources, which is a growing branch of contemporary Wikipedia research, such as performed by Teplitskiy et al. in [Teplitskiy et al., 2017].

Besides identifying a clustering of Wikipedia language versions, the work presented in [Samoilenko et al., 2016] also included an approach to find statistical explanations for the relationships between them. It was based on a mapping of the present languages to contemporary countries and the inclusion of related country fact data such as religion, spoken languages, common border, etc., with which the authors found the identified language ties to be mainly rooted in linguistic similarity of language, bi-linguality and shared religion. While it would have been tempting to perform a similar analysis for the nationality network, its historical dimension represented a significant caveat, since many of the featured "nation" entities had been changing across time or even existed only temporarily, making their mapping to contemporary countries infeasible. If possible at all, it would have required significant effort to identify comparable sources for historical information on religion, population, location of capital, bordering nations etc. across the ages in order to perform a comparable quantitative analysis, which was, unfortunately, out of scope for this work.

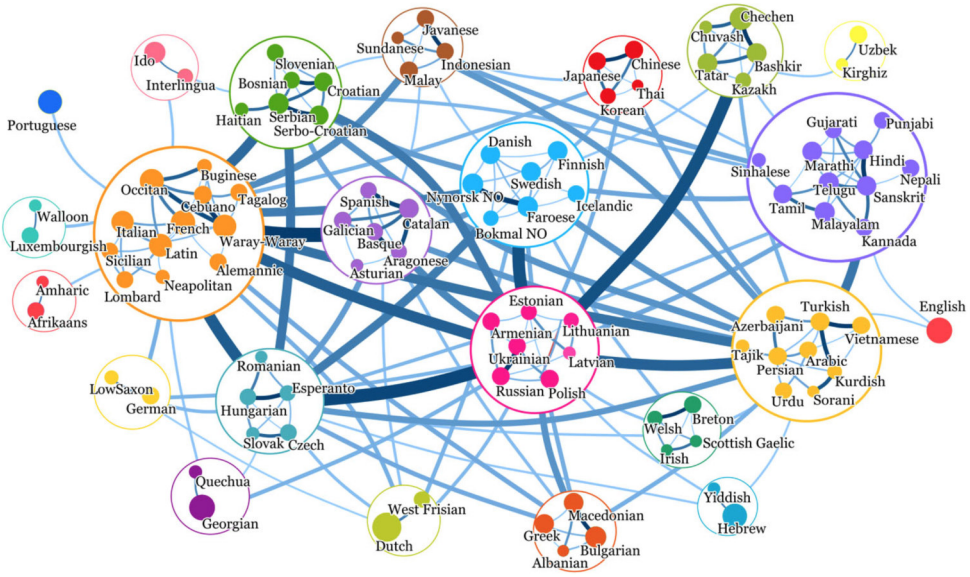


Figure 5.28: "Network of significant Wikipedia co-editing ties between language pairs" from [Samoilenko et al., 2016]

5.8 Summary

The alignment of the ULAN person records and their associative links with corresponding Wikipedia biographies and their mutual hyperlinks in multiple language editions of the public encyclopedia allowed to study the overlap between these two resources based on the data attributes available in the ULAN person records. Representing methodological framework **M2**, it allowed a rigorous analysis of the distribution of person attributes amongst the mapped Wikipedia biographies in different languages and the coverage of ULAN associative links via hyperlinks between them. This revealed that for persons covered in both datasets, the art history knowledge represented by the ULAN links was to a large extent also represented in the free encyclopedia, whose hyperlink network was moreover found to be much denser than its professional counterpart. Many of the WGA portrait relationships from artifact **A2** (Figure 3.8) which were found to be missing in the ULAN data could therefore be found in one or more Wikipedia language editions.

Comparing the distributions of ULAN person nationalities in different Wikipedia language editions and the analysis of top-ranked persons there revealed that characteristics that had already been identified for other content types, such as self-focus bias for places, were also present for biographical content, which at the same time underlined the importance of the individual knowledge contributions of the different language versions for achieving the high coverage of ULAN content. Moreover, the comparison of persons top-ranked in Wikipedia with the external scholarly rankings revealed that in addition to covering many social relationships such as teaching and family ties, the Wikipedia

hyperlinks also appeared to contain many general references between related persons. This likely contributed to the still present, but less expressed assortativity of the hyperlink network with respect to the different ULAN person attributes, although a similar tendency of biographies about persons having the same nationality and living at the same time to be connected with each other could also be identified in Wikipedia.

The analysis of the temporal characteristics of hyperlinks between persons revealed interesting patterns of connectedness across time. These findings motivated to filter Wikipedia hyperlinks by birth date difference for uncovering the temporal structure of the network. The resulting method **M3** enabled the approximately chronological force-based visualization of the filtered network which was used in artifact **A4** (Figure 5.19), which was compared with the previously created visual representation of the ULAN network, artifact **A3** (Figure 4.24). This revealed basic correspondence of the two sources especially regarding the clustering and the sequence of different nationalities there, but also showed the much higher density of the Wikipedia network and its higher coverage of many nationalities. The filtered network visualization approach was also used for comparing the network representations of the individual Wikipedia language versions in artifact **A5** (Figure 5.25), which provided useful views on the cultural preferences present in the different editions. The similarity of the generated visual representations with Barr's Torpedo Diagrams, such as the one shown in Figure 5.22, represented another remarkable agreement with a scholarly diagram related to art history, although in this case less concerning the representation of developments in the arts but rather providing a view on the underlying institutional policies, which remarkably appeared to be based on a similar form of self-focus bias as it was also evident in the different Wikipedia representations. Another outcome of filtering the network using method **M3** was the plot of (inter)national link densities across time shown in Figure 5.23, providing further insight on the differences between the ULAN network and its multilingual Wikipedia counterpart.

The overall interconnectedness of different nationalities in the Wikipedia network could be highlighted by transforming it into an aggregated form, resulting in the network of nationalities weighted by the number of underlying interperson links. In order to overcome the strongly skewed distribution of person nationalities, method **M4** was conceived to provide a relative measure for the strength of ties between them. The resulting visualization, artifact **A6** (Figure 5.27), revealed remarkable insight into the clustered structure of the international ties between the observed Wikipedia biographies which reflected the totality of cultural, colonial and political developments across the ages. The correspondence of the clustering of nationalities with similar attempts to find groups of related Wikipedia language versions raised questions on the interplay between processes of content creation and their actual results and provided strong incentive for future research.

The analysis of the network of ULAN biographies on Wikipedia showed that this resource was a valuable dataset for data-driven explorations in art history and potentially also other domains. One important step into this direction was to go beyond the ULAN as means to identify relevant biographies in the encyclopedia.

CHAPTER 6

Art History on Wikipedia beyond the ULAN

The analysis of the biographical networks present in the ULAN and its mappings to Wikipedia revealed that the many interactions between the featured persons provided a bottom-up view on large-scale developments in art history across the ages, which could be quantified and visualized using various means. It revealed specific interactions between higher-level entities, such as nations, across time, which also highlighted an obviously Western and particularly English-speaking-hemisphere oriented view on art history present in the selection of biographies identified via the ULAN. This was on the one hand based on the origin of the ULAN itself, on the other hand also on general historical circumstances of the discipline with its European origin and its, according to Joyeux-Prunel in [Joyeux-Prunel, 2014], focus "[...] on Paris and New York [...]".

The latter referred to a discussion about the "center" and the "peripheries" of art history, a common concept in contemporary scholarship reflecting ambitions to include previously marginalized groups into the canon of art history. Although, as pointed out by Joyeux-Prunel, this conception is not without problems and raises a lot of additional questions, it stands to reason that there has been, at least until well after WWII, an undeniable "cultural bias" in general art history with its main focus on European developments after Greek antiquity and notions of "center" and "periphery" thus nevertheless applied quite well to a number of observations made in the data. As far as the ULAN network was concerned, it for example became visible for the Native Americans in the dataset for which obvious effort had been made to include them and who were highly interconnected amongst each other via associative relationships, but not to the main network of historical persons. For the Wikipedia network it appeared in the aggregated view on interactions between nations from Figure 5.16 whose partition was in part obviously rooted in historical colonial ties, suggesting that the relational data present in the two collections were quite similar in this regard.

Since the selection of Wikipedia biographies was based on the ULAN, an important question was how a selection of Wikipedia biographies relevant to art history would be like if identified via other means than via ULAN mappings. It was of particular interest to find out if the chronological structure of the network would reappear in the set of alternatively identified biographies as well and if it there were additional aspects not present in the ULAN mapped version, or vice-versa. Another aspect of interest was the potential of extending the purely biography based approach to include additional types of entities and how the structure of a combined network of Wikipedia articles would differ from the biographical one. This chapter presents two related experiments which were performed in order to address these questions. The first one was based on the analysis and visualization of a bi-partite network of biographies and articles about art and architecture styles and is presented in Section 6.1. Its main outcomes were design artifacts **A7** (Figure 6.8) and **A8** (Figure 6.10) and the respective methods **M5** and **M7**. The second experiment, presented in Section 6.2, was dedicated to find an alternative way to identify Wikipedia biographies relevant to art history by using Wikidata, which could also be generalized to other domains. It resulted in design artifacts **A9** (Figure 6.12), **A10** (Figure 6.14), **A11** (Figures 6.16, 6.17 and 6.18) and method **M8**.

6.1 Network of biographies and articles about art styles in Wikipedia

The visualizations of the ULAN network and its Wikipedia counterpart corresponded with existing scholarly diagrams, showing that the aggregation of individual "bits and pieces" of information present in single biographical records or articles, edited by a multitude of persons, tended to approximate individual scholarly viewpoints to some degree. Only a subset of the identified scholarly diagrams, however, focused on individuals and/or national affiliations, which was found to be in line with ongoing debates in the field of art history which, as outlined by Hatt and Klonk in [Hatt and Klonk, 2006], have often been dealing with the extent of the influence of individuals and the overall social context on the developments of artistic form and content, with extreme variants focusing only on individual geniuses, others only on stylistic developments derived from comparing formal aspects of artworks.

A more formalist approach was Alfred Barr's diagram for his "Cubism and Abstract Art" exhibition of 1936, featuring the different Modern art styles represented in the show and the influential relationships between them, already introduced in Chapter 2 and shown again in Figure 6.1. As pointed out by Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005], the diagram was based on Barr's thorough analysis of stylistic changes in works of Modern art created between 1890 and 1935 and featured only few individuals conceived as the founding fathers of the sketched developments, taking the interrelated Modern art styles center-stage instead. The high popularity of this diagram and its impact well into the 21th century provided motivation to create data-driven versions of it which included additonal content and thus embedded it in a larger historical and contemporary context.

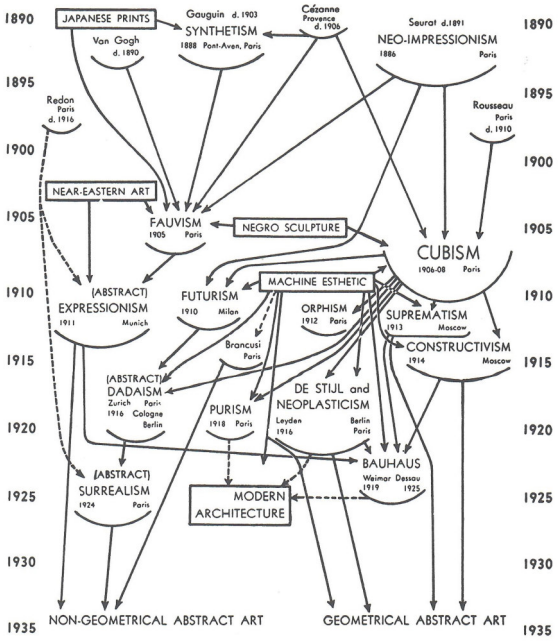


Figure 6.1: Diagram of Stylistic Evolution from 1890 until 1935, [Barr, 1936]

Considering computational versions of such formalist conceptions of developments of style, it rather stood to reason to use methods such as image analysis to directly extract patterns of style and their mutual interactions from the artworks themselves. An example for recent work in this regard was presented by Elgammal et al. in [Elgammal et al., 2018], based on formal principles established more than 100 years earlier by Wölfflin[1] in [Wölfflin, 1915]. Given the results obtained for the networks of interlinked biographies treated in this work, however, an analysis of the existing historiography about mutual relationships between art styles and the potential for obtaining similar bird's eye visualizations of their high-level structure was of interest as well, especially for considering social context such as nationality in this regard. One approach to such a view was to use Wikipedia to relate art and architecture styles with each other via affiliated biographies in form of a bi-partite hyperlink network, following the underlying assumption that many related styles would be "bridged" by mutually connected biographies which would provide valuable structured information for analyzing them in a socio-chronological context.

[1] Wölfflin coined the conception of an "art history without names".

6.1.1 Data retrieval & analysis

The data infrastructure used for the analysis and visualization of the Wikipedia biography network could be extended for the bi-partite case by incorporating additional information from Wikidata and DBpedia. Using the classes "art movement"[2] and "architecture style"[3], Wikidata could be queried for instances of them via dedicated "P31:instance of"[4] properties. This returned records such as "Baroque", "Art Noveau" or "Cubism" which again contained references to Wikipedia articles in different language versions. Their extracted multilingual hyperlinks to and from Wikipedia biographies which were in turn identified via Wikidata records having "P31:instance of" references to the class human[5] could again be found in DBpedia. Figure 6.2 illustrates an example for the envisioned multilingual bipartite network of styles connected via persons extracted from Wikidata and DBpedia. As shown for one of the three person records there, not all of the biographies connected to art and architecture styles were expected to be linked to the ULAN, potentially allowing the identification of persons relevant to art history in Wikidata/Wikipedia without having to rely on a single name authority in this regard.

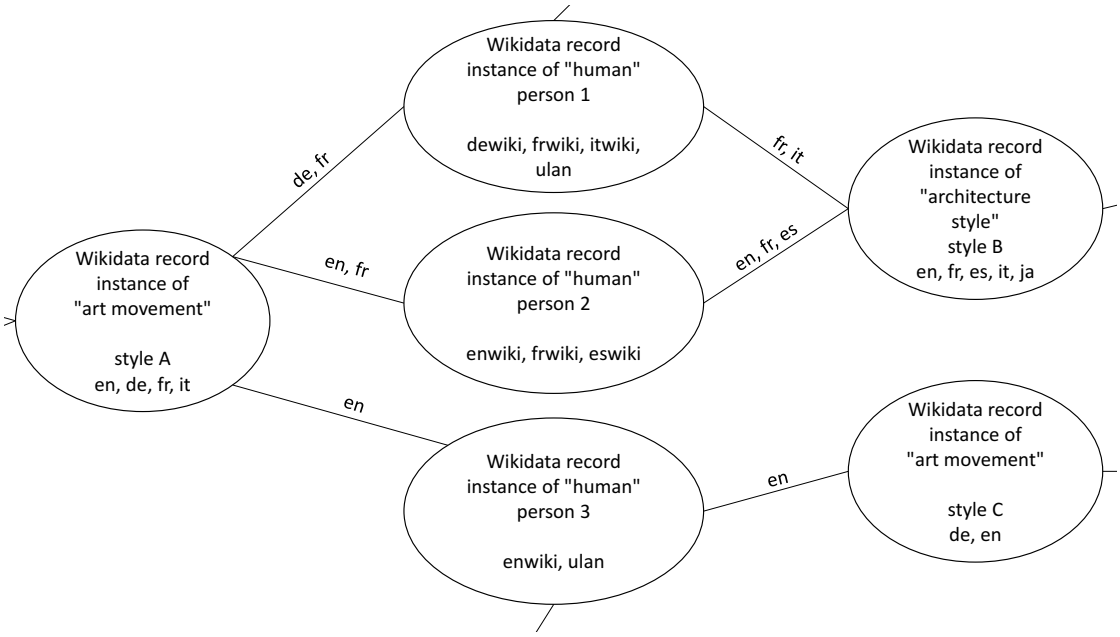


Figure 6.2: Example styles connected via persons in Wikipedia

The data retrieval procedure yielded 426 distinct styles, 406 of which were found to be connected with Wikipedia biographies about 69,908 persons via 145,079 unique hyperlinks (similar links occurring in multiple language versions counted as one) identified

[2] https://www.wikidata.org/wiki/Q968159, retrieved Sept. 14th, 2020
[3] https://www.wikidata.org/wiki/Q32880, retrieved Sept. 14th, 2020
[4] https://www.wikidata.org/wiki/Property:P31, retrieved Sept. 14th, 2020
[5] https://www.wikidata.org/wiki/Q5, retrieved Sept. 14th, 2020

in DBpedia page-link datasets for the same 50 language versions used for the analysis of the biography-only networks. Where available, additional attributes were retrieved for the identified person records, in case of multiple assignments of occupations or nationalities, the assignments of the more prominent value were used, i.e. a person record stating both "Austrian" and "German" as nationality and both "painter" and "sculptor" as occupation would be filtered to "German" and "painter", since these values appeared more often than the co-occurring ones. Table 6.1 lists the relevant Wikidata properties used in the context of data extraction.

Property ID	Property Name	Person	Style
P31	Instance of	human (Q5)	art movement (Q968159) architecture style (Q32880)
P27	Country of citizenship	X	
P106	Occupation	X	
P569	Date of birth	X	
P570	Date of death	X	

Table 6.1: Wikidata properties used for gathering records about styles and persons

The resulting network revealed a rich landscape of historical and contemporary styles, some of which highly interlinked with thousands of persons. Also the latter featured varying degrees of connectedness to one or more styles, sometimes linked to dozens of them. Table 6.2 shows the top-30 styles ranked by the number of interlinked persons and the top-30 persons ranked by the number of interlinked styles. While, as expected, broad periods such as Baroque or Renaissance were ranked highest, it was interesting to see a number of styles tied to relatively short timespans, such as Surrealism, Impressionism or even Socialist Realism, to be amongst the group of very highly interlinked styles. As far as the 30 persons top-ranked by number of linked styles were concerned, it appeared that especially artists and architects from the onset of the era of modern art/architecture were interlinked with the highest number of different styles.

Only a quarter of the almost 70,000 persons featured in the derived bi-partite person-style network could be found amongst the 40,000+ ULAN biographies mapped to Wikipedia via Wikidata, which suggested the identification via art styles to be a complementary approach to find persons relevant to art history there. In order to check how likely it was that a person identified this way was really missing in the ULAN or just no mapping present between its Wikidata record and an already existing ULAN dataset, a subset of 100 person records having the occupation "painter" was randomly chosen from the encountered Wikidata records not mapped to the ULAN, expecting that the restriction to painters would yield a more realistic estimate in this regard. Out of the 100 manually checked records about painters, 22 were found to have an existing but unmapped ULAN record, while the remainder could not be located there, this majority of almost 80% was interpreted as supporting evidence for the potential of this approach. The newly identified person records were thus analyzed for their distribution of role

style 1-15	count	style 16-30	count	person 1-15	count	person 16-30	count
baroque	7657	contemporary art	2233	Pablo Picasso	78	Walter Gropius	44
renaissance	7403	postmodernism	2089	Le Corbusier	71	Antoni Gaudí	43
romanticism	6607	mannerism	1865	Wassily Kandinski	69	El Lissitzky	43
surrealism	4909	rococo	1816	Piet Mondrian	64	Eugène Viollet-le-Duc	43
expressionism	4763	conceptual art	1644	Frank Lloyd Wright	61	Joan Miró	43
avant-garde	4612	futurism	1587	Kazimir Malevich	61	Michelangelo	43
impressionism	4394	gothic revival architecture	1523	Ludwig Mies van der Rohe	61	Raphael	42
modernism	3985	bauhaus	1516	Marcel Duchamp	59	John Ruskin	40
art nouveau	3600	gothic art	1465	Henri Matisse	57	Andrea Palladio	39
realism	3485	pop art	1464	Vincent van Gogh	57	Karl Friedrich Schinkel	39
neoclassicism	3358	art deco	1383	Paul Klee	52	Robert Delaunay	39
socialist realism	2892	abstract expressionism	1353	Paul Gauguin	50	Yves Klein	39
symbolism	2881	dadaism	1345	Salvador Dali	49	Leonardo da Vinci	38
classicism	2570	minimalism	1283	Claude Monet	46	Philip Johnson	38
cubism	2289	academic art	1269	Paul Cézanne	45	El Greco	37

Table 6.2: Top 30 Wikipedia styles/persons ranked by #links from opposite type

and nationality. The top 30 of the respective attribute values are shown in Table 6.3, with the top-30 roles accounting for 40,194 (81.00%), the top-30 nationalities for 37,054 (74.68%) of the 49,620 persons without ULAN mappings. Significant fractions of the records without ULAN references, however, had no role (4,159) and/or nationality (7,253) assigned at all.

As far as the roles were concerned, eight of the top-30 roles represented professions of immediate relevance to art history, highlighted in bold in Table 6.3 and assigned to 17,414 of the person records without ULAN mappings. At the same time, however, it also became visible that many other persons of less art history related roles were connected to articles about art and architecture styles, most notable were 9,783 writers and 3,777 composers in this regard, two professions representing different art forms that nevertheless have always been connected to the fine arts in terms of mutual influence and overall artistic "Zeitgeist". Many of these persons were interlinked with articles about rather broad styles, such as Romanticism, Modernism, Baroque or Renaissance, but also to more distinct ones such as Symbolism or Surrealism which had expressions in different artistic domains. Some styles retrieved via the Wikidata "art movement" category were even dedicated to other art forms, such as "20th-century classical music". While more broad professions such as university teacher, journalist or historian potentially included art history related persons, many additional designations appeared to be from completely different domains. This was nevertheless also the case for many ULAN records about important patrons or clients, but the overall fraction of "non-artists" was much higher in the group of persons interlinked to art and architecture styles.

Looking at the nationalities, the top ranking values in Table 6.3 were again dominated by the same major western countries as it was the case for the ULAN, although not without some notable exceptions. The 2,718 identified persons assigned with nationality "Soviet Union" for example clearly represented an extension to the rather western-oriented distribution of person nationalities in the ULAN, 1,939 of them were moreover unique contributions from the Russian Wikipedia. This was also the case for persons of other

nationalities from the "periphery", such as the 1,243 of the 1,427 Japanese persons not mapped to the ULAN contributed by the Japanese Wikipedia version or the 503 out of 596 Brazilians sourced from the Portuguese language version. Together with the insights into the distribution of person roles, these findings underscored the value of using a multilingual resource fed by global contributions, such as Wikipedia, for identifying persons relevant to art history beyond those present in sources such as ULAN with its specific institutional background.

writer	9783	**art historian**	**433**	Germany	6371	Brazil	596
painter	**9218**	philosopher	307	United States of America	4653	Finland	569
architect	**4489**	film director	260	France	3976	Switzerland	541
composer	3777	translator	226	Soviet Union	2718	Hungary	522
sculptor	**1594**	priest	209	Spain	2458	Czech Republic	496
university teacher	1460	**art collector**	**194**	United Kingdom	1913	Norway	488
poet	1436	conductor	183	Italy	1448	Kingdom of Italy	405
politician	1186	musician	158	Japan	1427	Portugal	391
artist	**863**	Catholic priest	155	Poland	1040	Romania	381
journalist	721	novelist	151	Kingdom of the Netherlands	980	Canada	346
actor	719	pianist	144	Austria	964	Argentina	333
screenwriter	545	diplomat	139	Belgium	784	Israel	266
photographer	**504**	literary critic	132	Russian Empire	769	Australia	264
historian	500	entrepreneur	127	Denmark	740	Mexico	250
singer	462	**illustrator**	**119**	Sweden	716	Russia	249

Table 6.3: Top 30 roles and nationalities of WDA records not linked to the ULAN, identified via interlinked art and architecture styles

Centered upon relatively distinct historical periods and world regions, many of the identified styles featured specific distributions of nationalities of interlinked persons. Figure 6.3 shows them as derived from the the Wikidata person records linked to the 40 most interlinked styles, highlighting some styles with very regional focus, including Socialist Realism, with about two thirds of interlinked persons from the Soviet Union (mapped to "Russian" in the legend due to the limited set of visually distinguishable colors), School of Paris with a comparable fraction of French persons, Abstract Expressionism with about 60% Americans, Futurism with about 50% Italians and Bauhaus with about the same fraction of Germans. Another "style" used in a very negative connotation by the National Socialist dictatorship during the 1930's and 1940's was Degenerate Art, having a very high share of German artists, ironically being one of the "purest" styles with respect to the variety of nationality values linked to it.

Other styles in turn featured high fractions of persons with unassigned nationality, especially visible in the Figure for Gothic Art, Renaissance, Mannerism, Baroque and Rococo. At the same time, some of them featured unexepectedly low fractions of Italian persons which appeared to be suspicious when considering the seminal importance of artworks from that region for the "early days" of Western art history. Since 2,800 of the 10,053 Wikidata records with unassigned nationality had corresponding ULAN records,

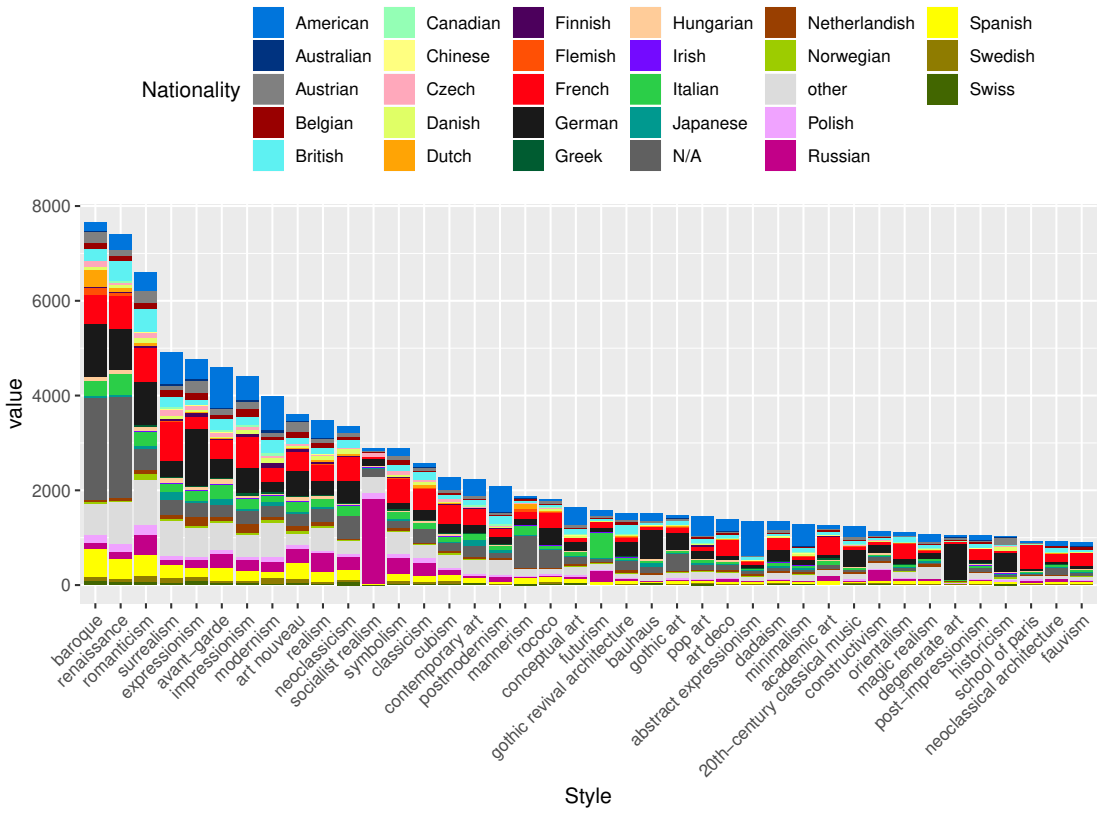


Figure 6.3: Nationality distributions for the 40 most interlinked styles

it was possible to consult the nationality attributes of the latter, which revealed 1,820 "missing" Italians.

Besides those of nationalities, the distributions of roles of persons interlinked with the different styles were of interest as well. Figure 6.4 shows a similar comparison of the 40 most interlinked styles in this regard. Some of the styles were interlinked with persons of very similar roles, especially visible for a number of architectural styles almost exclusively featuring interlinked architects, such as Gothic Revival Architecture, Historicism and Neoclassical Architecture, or art styles rather exclusive to painters, such as Academic Art, Abstract Expressionism, Post-Impressionism, School of Paris and Fauvism. Styles mainly belonging to other art forms than fine arts and architecture featured quite unique distributions of linked roles reflecting the underlying domain, such as 20th-century classical music mentioned before, a number of "shared" styles such as Magic Realism, Postmodernism, Symbolism, Surrealism or Dadaism, divided mainly between painters and writers, stood out as well. As expected, broad styles such as Baroque or Renaissance in turn represented a more balanced mix of different art forms.

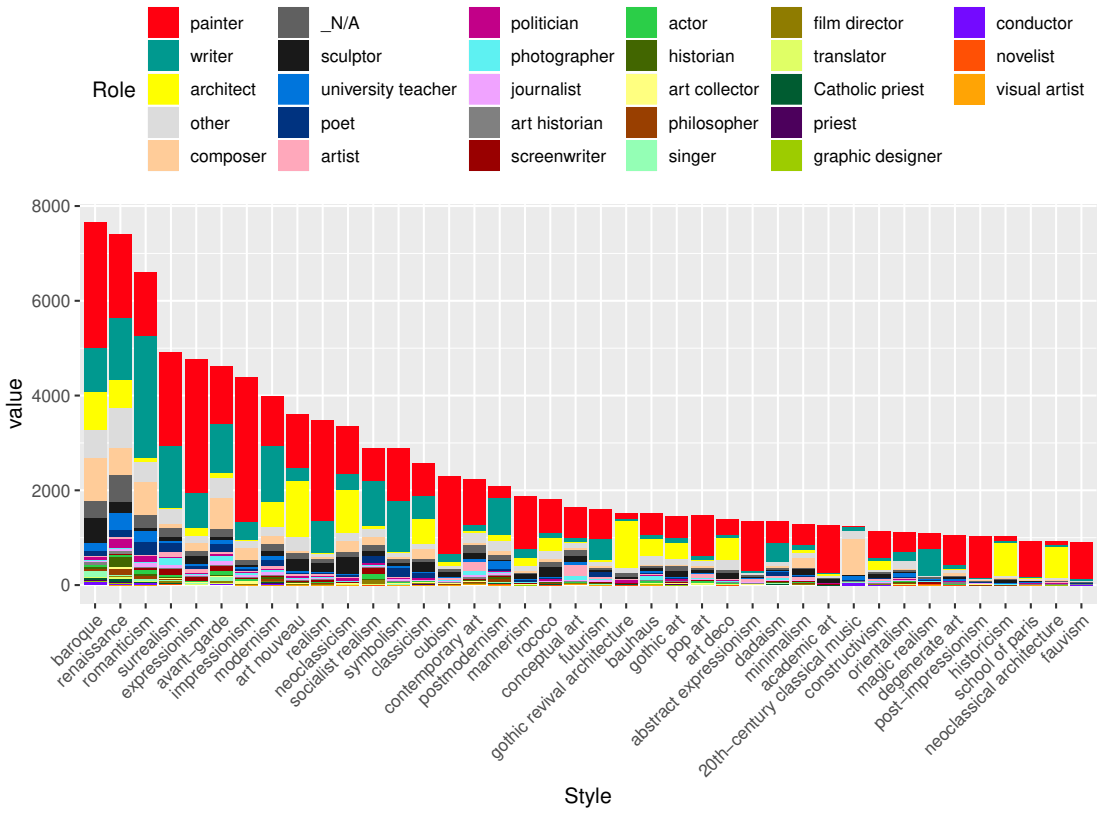


Figure 6.4: Role distributions for the 40 most interlinked styles

6.1.2 Visualization of the person-style network

As far as the visual form of the person-style network was concerned, its rendering "as is", shown in Figure 6.5, did not suggest the presence of clear chronological succession of styles, similar to the unfiltered representation of the Wikipedia biography network. It nevertheless appeared to feature distinct regions, with the present styles arranged according to mayor epochs such as Renaissance/Baroque and Modern art. Moreover, developments at the "periphery" of the art world were likewise located at rather remote sections of the plot, such as the group of interconnected Japanese styles at the bottom of the Figure. These observations raised the question if a more chronological structure could be unraveled there in a similar fashion to the biography network via a filtering approach.

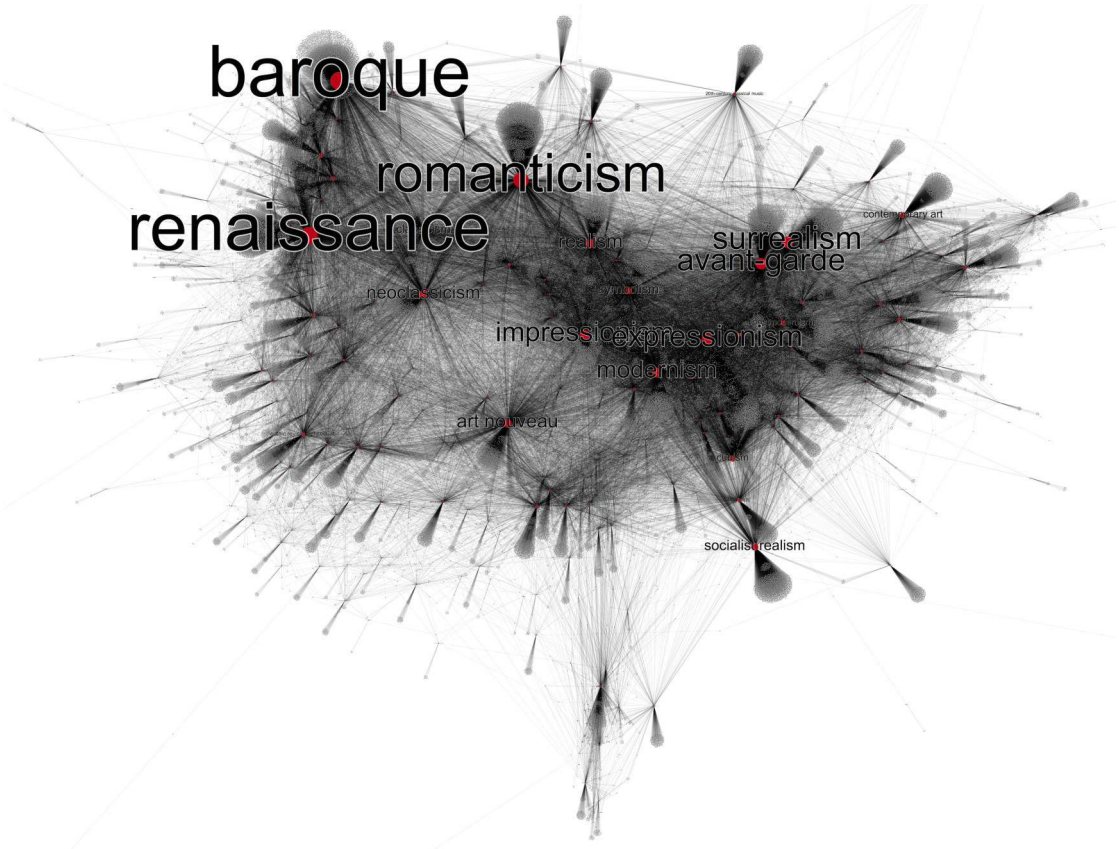


Figure 6.5: Full network connecting styles with persons in Wikipedia

In order to analyze the network for its temporal characteristics, this aspect of the person-style hyperlinks could be made visible in form of scatterplots, one for each style, featuring all interlinked persons positioned by the time they were active, based on the average year ("mid-life-year") between each person's birth and death year as simple measure. It was expected that each of the different plots would reveal a specific historical period featuring most of the interlinked persons, coinciding with the actual historical extent of the style and thus also pointing out any "outliers", i.e. interlinked persons belonging to different periods. A subset of the created plots is shown in Figure 6.6, revealing that especially for well known stylistic periods such as Baroque or Renaissance, the temporal distribution of linked persons often spanned multiple centuries, not at all limited to only persons alive during the actual duration of the period. In the case of the two mentioned ones, each of them featured two relatively distinct temporal accumulations of linked persons, one of which relatively well-aligned with the actual duration, the other one mainly comprised of more recent persons from the 19th and 20th centuries. Similar to the case for the biography network, such "long distance" references had strong impact on the visual result of the force-based layout algorithm since they often lead to interrelated

styles that actually belonged to different epochs, which, although of interest in its own regard, "masked" the potentially embedded chronology.

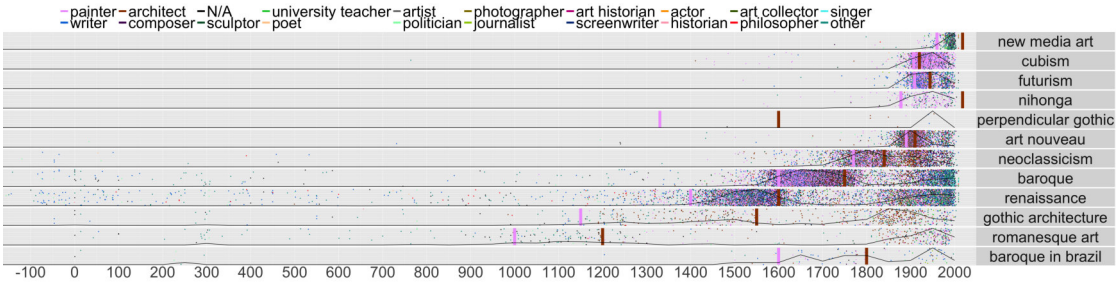


Figure 6.6: Jittered scatterplots showing the temporal distribution of persons linked to art styles in Wikipedia

For unraveling the chronological structure in the visualization, it was therefore decided to filter the data by removing links from persons who were not active during the art style period. As it turned out, however, this could not be achieved easily, since many of the Wikidata records about styles only featured incomplete or missing information about approximate beginnings and endings of the respective periods. Considering a separate approach to deriving the desired start and end dates, the identified temporal distributions of interlinked persons suggested to be a potential solution to that problem, allowing to calculate approximate values accordingly. As visible in Figure 6.6 via the density of the scatterplots and the respective black line graphs, however, not all styles featured unique temporal accumulations of interlinked persons, which in case of the presence of multiple "peaks" required to decide on the correct one. A straight-forward approach was to aggregate the full temporal range of all person life dates present in the data for each style into distinct 50 year sections and to identify the one featuring the highest number of interlinked persons, using it together with adjacent sections above a certain threshold to approximate the actual duration of each style. The results obtained this way, however, were often misleading, since especially "older" periods such as "Romanesque art" or "Gothic architecture" often had more links from and to persons from the 19th and 20th centuries than contemporaries. Figure 6.6 highlights this issue by showing the actual durations of the featured styles via pink and brown bars. It suggested that an alternative, potentially better performing method would take the overall skewed temporal distribution of person records into account, weighting obtained link counts differently for different points in time based on the varying number of persons across time.

In principle, the requirement to find such a relative measure was similar to the problem encountered for analyzing the aggregated links between nationalities, where the interactions between smaller groups could not be identified due to high volume ties between large ones. It thus stood to reason to apply a contingency table approach similar to the nationality case sketched in Table 5.15, this time featuring link counts in cells whose rows represented the individual styles and the columns the birth-death average year of the interlinked persons rounded to the next half-century. This contingency

table was used to calculate standardized residuals as described in [Agresti, 2007, p. 38], similar to the nationality case, highlighting for each style those cells whose counts were significantly above what would have been expected based on the overall number of persons active in the respective half-century. For each row, the half-century with the highest positive standardized residual was then used together with its immediate neighboring half-centuries above a certain threshold to determine the style's duration.

The procedure outlined above is illustrated in Table 6.4, featuring four hypothetical cases for Wikipedia articles about styles with different temporal distributions of interlinked Wikipedia person biographies. In each cell, the upper half shows the observed count of persons interlinked with the style represented by the current row having their average birth-death year falling within the respective half-century represented by the current column, while the lower half shows the expected count calculated from the marginal distributions (shown in the rightmost column and the bottom row) and the resulting standardized residual derived from the deviation between observed and expected counts. Using a standard residual threshold of zero for this example, each row features bold entries for half-centuries with the highest residuals and their neighbors having values above threshold, used together to determine the duration of the respective style. An example for a style having two distinct peaks for observed counts is shown in the second row, with the more recent peak having a slightly higher absolute count which, as mentioned above, was often the case for styles which nevertheless were located in earlier times. Specifically targeted at such cases, the residual-based approach detects significant deviation from expectation and thus suggests the lower counting peak to represent the style's duration.

	1700	**1750**	**1800**	**1850**	**1900**	**1950**	
style 1	0 3.46 -2.35	0 5.35 -3	5 3.46 **1.04**	10 5.35 **2.61**	20 12.91 **2.9**	5 9.45 -2	40
style 2	10 3.38 **4.53**	7 5.22 **1.01**	5 3.38 **1.11**	0 5.22 -2.95	12 12.59 -0.24	5 9.21 -1.91	39
style 3	1 1.91 -0.75	10 2.94 **4.86**	0 1.91 -1.59	5 2.94 1.42	6 7.1 -0.55	0 5.2 -2.87	22
style 4	0 2.25 -1.76	0 3.48 -2.25	1 2.25 -0.98	2 3.48 -0.96	3 8.39 -2.54	20 6.14 **7.17**	26
	11	17	11	17	41	30	127

Table 6.4: Determining stylistic periods via standardized residuals

Especially for cases where more recent links outnumbered those from the actual historical period, the residual-based peak selection in the temporal distribution of persons interlinked with each style suggested to allow a better identification of their temporal extent. This raised the question if following this approach would introduce errors on the "other side" of the spectrum, i.e. wrongly identifying more recent styles as being from earlier periods in history and thus leading to comparable or even worse results as obtained via counting absolute numbers of links. A number of metrics were thus calculated in order to compare the results obtained with the two methods with each

other, which required reference data in form of manually annotated temporal extents for the styles. Dealing with a relatively small set, it was thus decided to annotate all of the styles with beginning and end dates, but not without filtering out those having relatively sparse data: Out of the 406 styles having at least one link to a person in one of the Wikipedia language version, 31 were removed for not having at least one half-century with more than two person links, reducing the observed set of styles to 375 of them.

The temporal distributions of person links from and to the annotated set of 375 styles were subsequently analyzed for temporal peaks using different runs of the count and the residual approaches with varying thresholds. While the latter were directly applied to the actual standardized residual value for the residual-based approach, the absolute counts in the other approach were first normalized for each style so that the value of the highest counting half-century equalled one, enabling the use of the same threshold value across all different styles with their differing absolute link counts. Each run was described using a number of different measures, which included counts of cases where the duration of a style was determined to be clearly before or after its actual extent (detected start/end half-century after/before actual end/start half-century), counts of cases for which the calculated beginning and end dates overlapped with, wrapped or fell within the actual ones, and measures of precision and recall on link-level. The latter were based on the standard notion of true/false positive/negative cases, counting links from persons whose birth-death average year rounded to the next half-century fell within both the annotated and the calculated style duration as "true positives", those within the actual but not within the derived duration as "false negatives", those only within the latter but not in the former as "false positives", those outside both ranges as "true negatives". The obtained positive/negative counts were subsequently used to calculate F1 and Matthews correlation coefficient (MCC) measures, which were used to determine the best-performing setting.

Table 6.5 shows the results of the top-20 runs ordered by decreasing F1 measure, its columns representing the used threshold (thresh), the applied method ("z" or "count"), the number of styles for which no peaks could be detected for the current threshold (NA), the number of styles for which the detected and the actual start/end dates at least overlapped (match), more detailed counts for styles where start and end dates agreed (same), where detected dates wrapped the actual dates (wrap), where they fell within the actual dates (within) or where either only the detected start or the detected end date fell within the actual duration (overlap), the count of wrongly detected durations (false) including more detailed counts if the detected style duration began after the actual one ended (late) or if it ended before the actual duration began (early), the MCC, F1, precision (prec) and recall measures (rec) as well as the counts for the true/false positives/negatives (tp/fp/tn/fn). The F1 measure was used for choosing the best approach and the respective threshold because it weighted precision and recall equally, supporting the most precise detection for as many styles as possible. As shown in the Table, the residual-based approach performed better in terms of precision, while the count based approach yielded better recall. The latter resulted in a higher number of wrongly detected style durations which, as anticipated, mainly included cases where,

due to higher numbers of references from more recent persons, the actual style duration ended before the detected one. The residual-based approach in turn detected a number of style durations as earlier than they actually were, which was anticipated as well, but overall resulted in clearly less mis-detections than the count-based method. While the differences between the count- and the residual-based methods appeared relatively small at first sight, the gap between the two widened when taking the actual deviation of wrongly detected style durations from the actual ones into account. Considering the mean deviations of detected start/end dates from the actual ones, the count-based approach consistently featured higher deviations than the residual-based one.

thresh	method	NA	match	same	wrap	within	overlap	false	late	early	mcc	f1	prec	rec	tp	fp	tn	fn
0.33	z	0	357	4	201	32	120	18	12	6	0.6131	0.8948	0.8760	0.9144	93147	13183	27601	8722
0.67	z	0	356	4	195	33	124	19	12	7	0.6131	0.8946	0.8766	0.9133	93041	13101	27683	8828
1.00	z	0	355	4	190	33	128	20	13	7	0.6076	0.8924	0.8769	0.9086	92556	12998	27786	9313
1.33	z	1	352	4	180	34	134	22	14	8	0.6008	0.8898	0.8770	0.9031	91994	12907	27877	9875
1.67	z	1	352	4	172	36	140	22	14	8	0.5983	0.8878	0.8793	0.8966	91334	12540	28244	10535
2.00	z	4	347	4	163	37	143	24	15	9	0.5948	0.8864	0.8793	0.8936	91035	12496	28288	10834
2.33	z	6	347	4	160	40	143	22	13	9	0.5989	0.8863	0.8834	0.8892	90578	11958	28826	11291
2.67	z	10	345	4	155	41	145	20	12	8	0.5920	0.8834	0.8832	0.8836	90009	11900	28884	11860
0.33	count	0	347	2	182	27	136	28	28	0	0.5377	0.8828	0.8362	0.9348	95225	18649	22135	6644
3.00	z	16	342	4	150	40	148	17	8	9	0.5907	0.8821	0.8847	0.8796	89609	11684	29100	12260
0.17	count	0	347	2	229	18	98	28	28	0	0.5033	0.8809	0.8023	0.9766	99484	24512	16272	2385
0.43	count	0	347	3	159	38	147	28	28	0	0.5547	0.8806	0.8572	0.9052	92216	15358	25426	9653
0.40	count	0	347	3	166	33	145	28	28	0	0.5418	0.8801	0.8473	0.9154	93255	16803	23981	8614
0.20	count	0	347	2	217	18	110	28	28	0	0.4993	0.8793	0.8083	0.9639	98195	23292	17492	3674
3.33	z	22	337	4	144	44	145	16	7	9	0.5824	0.8787	0.8842	0.8734	88970	11656	29128	12899
0.23	count	0	346	2	207	21	116	29	29	0	0.5005	0.8785	0.8132	0.9553	97311	22347	18437	4558
3.67	z	25	335	4	141	43	147	15	6	9	0.5819	0.8782	0.8847	0.8719	88815	11576	29208	13054
0.27	count	0	347	2	199	22	124	28	28	0	0.4986	0.8776	0.8149	0.9508	96852	21994	18790	5017
0.13	count	0	347	2	236	18	91	28	28	0	0.4832	0.8775	0.7933	0.9817	100008	26058	14726	1861
4.00	z	32	328	4	134	48	142	15	6	9	0.5787	0.8765	0.8852	0.8680	88422	11471	29313	13447

Table 6.5: Style duration detection runs (threshold, method) ordered by decreasing F1

Extending the marks delimiting the correct style durations shown in Figure 6.6, Figure 6.7 adds annotations representing the start and end dates detected via the standard residual-based approach using the identified best fitting threshold 0.33 as light and regular green bars, the location of the peak is represented by an olive green bar in between. In addition, red line graphs now represent the normalized standard residual values for the neighborhood around the highest standard residual peak (olive green bar) above threshold, set to zero outside that range. Comparing them with the black lines representing the count distributions reveals how the residual-based approach amplified durations with low absolute but — with respect to the overall number of persons active during that period — significant counts. Two examples in Figure 6.7 nevertheless represent typical cases where the residual-based approach yielded wrong results. The wrongly identified beginning and the end dates for "Perpendicular Gothic" represent the situation where data for the actual period was missing but enough information present for other periods so that the style passed the initial noise filter. The mismatch for "Baroque in Brazil" was in turn due to a very small number of person links from antiquity which, because of the sparsity of persons from that time in the dataset, lead

to a peak of higher significance there than detected for the actual period. As shown in Table 6.5, however, only 18 out of the 375 style periods were wrongly detected at the best-performing standard residual threshold of 0.33, while the identified beginning and end dates at least overlapped with the actual durations for the remaining 357 periods.

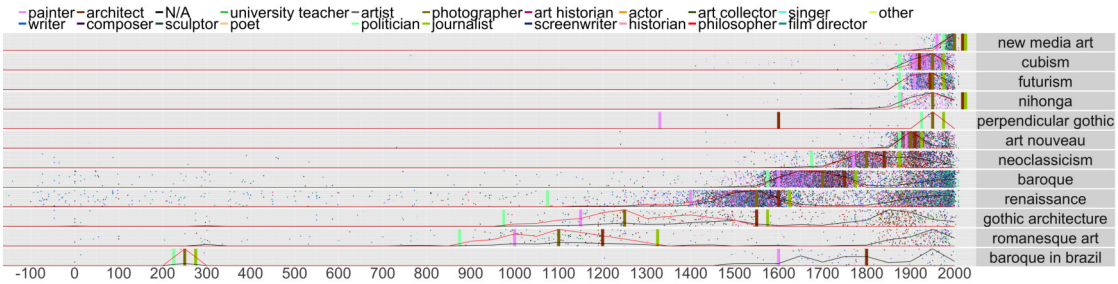


Figure 6.7: Jittered scatterplots showing the temporal distribution of persons linked to art styles in Wikipedia, including the results of the start/end date detection

Pruning all the links from persons whose average year between birth-death laid outside the detected temporal ranges for each style (all rounded to the next half-century) and consolidating the remaining 106,330 links (93,147 True and 13,183 False Positives, 82,920 pointing from persons to styles and 23,410 from styles to persons) left 93,197 unique, strictly unidirectional ties from 55,499 persons to the 375 styles. After applying the same force-based layout algorithm (ForceAtlas2) as for the pruned person-to-person network from Figure 5.19, the visualized topology of the reduced person-style network, shown in Figure 6.8, revealed a clearly visible succession of styles across the 2nd Millenium AD, beginning at around the time of (Pre-)Romanesque art and architecture and ending with contemporary late 20th/early 21th century styles/genres such as New Media Art.

Until about the Baroque era, extending across Gothic, Renaissance and Mannerist epochs, the succession appeared to feature arts and architecture styles in a relatively intermingled manner. After Baroque, however, this single strand began to seperate into two quite distinct branches, the top part representing architectural styles, the lower part mainly showing fine art styles. Although clearly seperated throughout the following centuries, these two strands were nevertheless continuously interrelated with each other, in particular due to styles such as Art Nouveau, Art Deco and Bauhaus which, as also visible in Figure 6.4, had relatively equal shares of links coming from artists and architects.

Comparing the topology-based Wikipedia visualization with Barr's chronologically ordered diagram showed that although not in a similarly strict chronological order, Modern art styles still occupied a relatively clearly-outlined temporal region of the data-driven version, whose larger coverage put them into a much broader art historical context than it was the case in the original. This included all the developments taking place before the end of the 19th century and after World War II and moreover, especially by coloring the person nodes by nationality using the same color coding as in Figure

6.3, again highlighted the "regional" character of many entries such as "Arts and Crafts Movement", "Futurism", "Modernisme", "School of Paris" or "Abstract Expressionism" which was not directly visible in Barr's visualization. Non-Western developments were prominently visible as well, such as the separate strand of Japanese[6] styles and schools shown at the top of Figure 6.8, merging with the main strand at around 1900 or the satellite-like cluster of persons arranged around Socialist Realism. Especially the latter appeared clearly separated from the main body and, as mentioned before, also contained many artists not featured in the ULAN, underlining the specific focus of the biographies collected there. A similarly detached satellite was "20th-century classical music" whose connection with the main body nevertheless reflected the mutual influence between different artistic disciplines, inviting for deeper exploration in future work.

[6] The influence of Japanese art styles on Modern art was also acknowledged by Barr via a dedicated entry for "Japanese prints" in his diagram

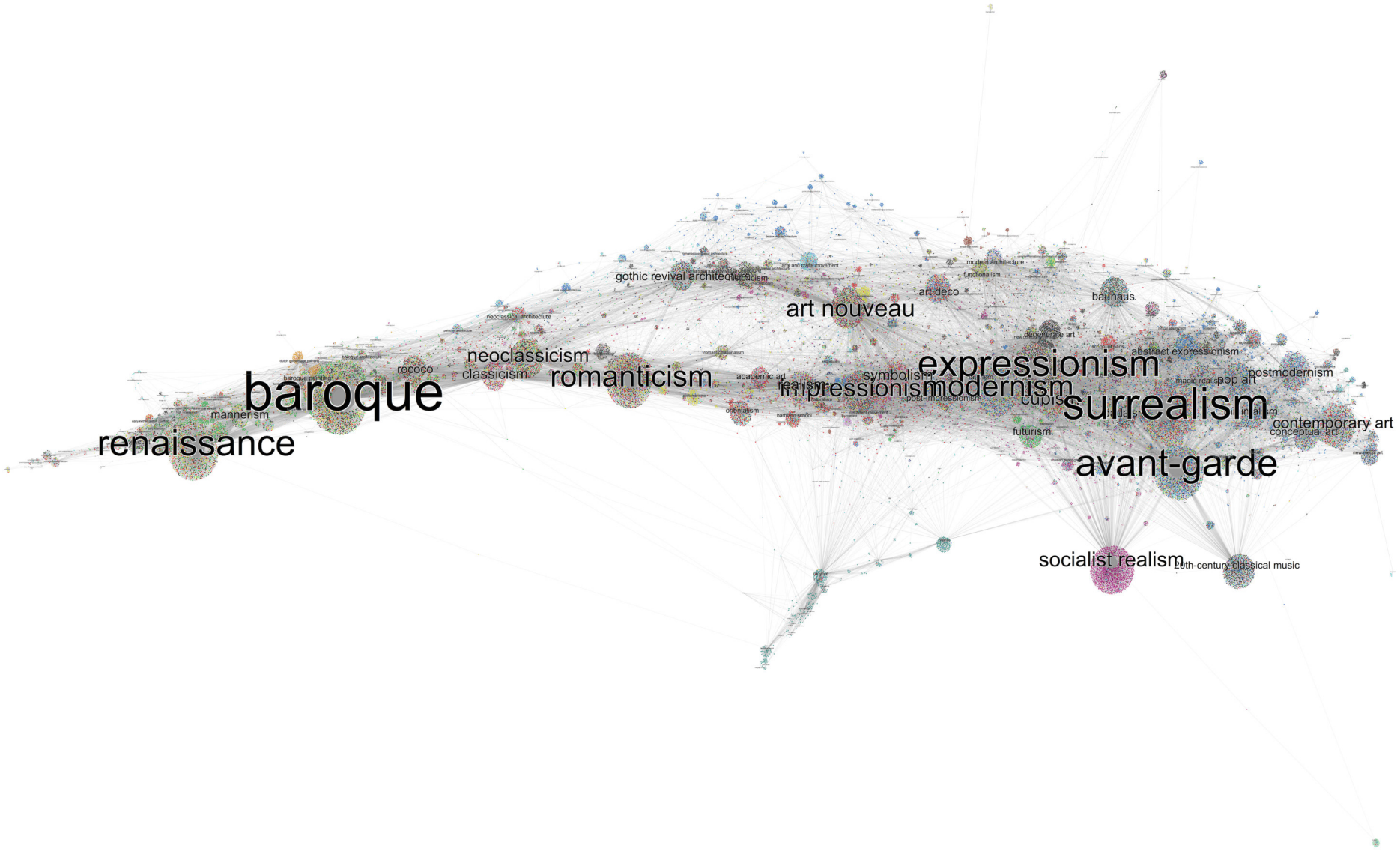


Figure 6.8: Pruned network of persons and art & architecture styles (**A7**)

6.1.3 Bi-partite projections

As described in [Newman, 2010, p. 123 ff.], bi-partite networks can be reduced to two different weighted uni-partite networks via a process called bi-partite network projection, in which any pair of two entities of one partition would be assigned a link weighted by the number of their commonly linked entities from the other partition. In the case of the person-style network this would result in one network of persons connected via links weighted by the number of commonly linked styles, and in another network connecting pairs of styles weighted by commonly linked persons. Assuming the bi-partite person-style network to be represented by a 0-1 adjacency matrix M with rows representing persons and columns representing styles, the person-person projection could be calculated as MM^T, the style-style projection as M^TM, respectively.

Given the focus of the Barr diagram from Figure 6.1 on relationships between art styles, especially the projected style-style network was of interest to be studied in more detail. The immediate outcome of projecting the person-style network to a weighted network of styles, however, had a similar distribution of style-style link weights as encountered in the aggregated nationality network shown in Figure 5.26. It featured the most prominent styles with many incoming person links as densely interconnected with each other via strongly weighted ties, while potential clusters of styles having less person links remained barely visible due to mutual ties with very small weights.

Since the filtering of the nationality network using the residual-based approach described in Section 5.7 resulted in the increased visibility of ties having low absolute counts, the application of a comparable method to identify significant relationships between styles in the style-style projection of the bi-partite person-style network seemed appropriate. Although it would in principle have been possible to directly apply the same method as used for the nationality network to the projected style-style network as well, it had been shown by Zweig and Kaufmann in [Zweig and Kaufmann, 2011], who referred to this method as *Statistical Indepence Model (SIM)*, that for bi-partite networks, better results could be achieved by applying the so-called *Fixed Degree Sequence Model (FDSM)* instead. This approach is based on the comparison of the projection of an observed bi-partite network with the average of the projections of a large number (5,000-10,000) of randomly generated bi-partite networks which must be structurally similar to the observed one by preserving its original node-degrees. A dedicated R package "BiRewire", described by Gobbi et al. in [Gobbi et al., 2014], could be used for the efficient generation of such random bi-partite networks, each of them based on a sufficient number of degree-preserving random "rewirings" of the links present in the original, observed network. The authors of the package suggested an analytically derived lower bound for the number of such rewirings required to ensure that each randomly generated bi-partite network represented an independent random sample out of the space of all possible degree-preserving versions of the original network, which was followed accordingly.

The overall procedure is illustrated in Figure 6.9: An example matrix $OBS_{bipartite}$ for an observed bipartite network of persons (rows) and styles (columns) is shown in the top left of Subfigure a). Standing as representatives for the much larger set R of n

random networks which must be generated according to the FDSM, two degree preserving random rewirings of this example network, R_1 and R_2, are shown further to the right, row/column and total sums are provided for comparison, showing that they keep the degree structure of the original network intact. The bottom row of Subfigure a) shows the respective $M^T M$ column projections of the three matrices on top, the projection OBS_{proj} of the observed matrix $OBS_{bipartite}$ on the left, the two projections RP_1 and RP_2 as representatives for the set of the n random projections RP derived from the set R on the right. The diagonals of the projections are discarded since they would represent self-loops of the projected nodes weighted by their original degree, which are not of interest. Due to the degree preserving rewiring of the source bipartite network, the projected random style-style networks RP_1 and RP_2 all feature the same sum of link weights as OBS_{proj}, although with varying individual link weights for each randomly generated rewiring. The central assumption of the FDSM based approach according to [Zweig and Kaufmann, 2011] is that, given a large number of such generated random degree-preserving networks, the mean of the projected link weights for each pair of projected nodes represents a good estimate for their expected link weight given the degree distribution of the original bi-partite network. The left side of Subfigure b) shows the mean matrix $\mathbb{E}(RP)$ of expected values for the link weights obtained from each projection in RP, calculated via

$$\mathbb{E}(RP) = \sum_{i=1}^{n} RP_i/n$$

The right side of Subfigure b) shows the matrix $\mathbb{SD}(RP)$ of standard deviations of projected link weights from RP, calculated via

$$\mathbb{SD}(RP) = \sqrt{\mathbb{E}(RP \odot RP) - \mathbb{E}(RP) \odot \mathbb{E}(RP)}$$

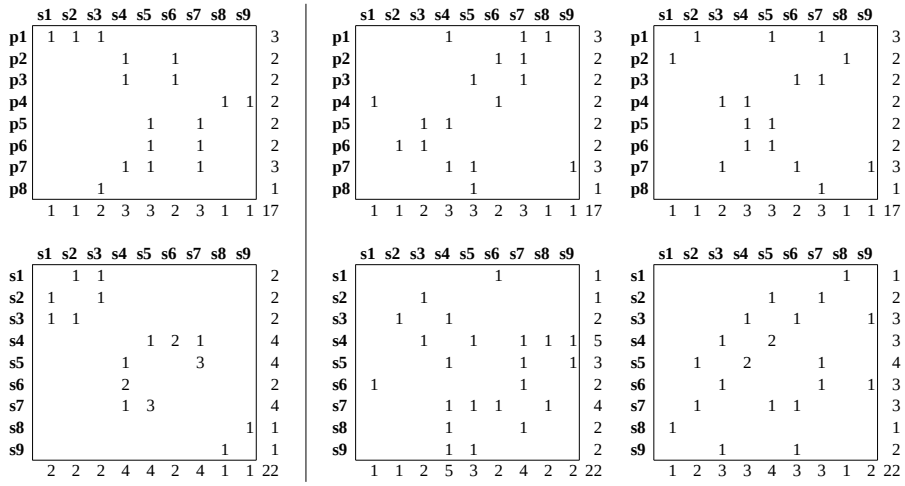
where $\odot$ refers to element wise multiplication of two matrices and

$$\mathbb{E}(RP \odot RP) = \sum_{i=1}^{n} (RP_i \odot RP_i)/n$$

The matrix $STDRES$ of standardized residuals between the observed projected matrix OBS_{proj} and the mean matrix $\mathbb{E}(RP)$, shown in Subfigure c), is then calculated via

$$STDRES = \frac{OBS_{proj} - \mathbb{E}(RP)}{\mathbb{SD}(RP)}$$

Assuming the range of two standard deviations above or below zero to be "within expecation", Subfigure c) shows all positive residuals beyond that range in bold, suggesting that the respective ties between the involved styles are stronger than expected.



	s1	s2	s3	s4	s5	s6	s7	s8	s9	
p1	1	1	1							3
p2				1		1				2
p3				1		1				2
p4								1	1	2
p5					1		1			2
p6					1		1			2
p7				1	1		1			3
p8			1							1
	1	1	2	3	3	2	3	1	1	17

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
p1				1			1	1		3
p2						1	1			2
p3					1		1			2
p4	1					1				2
p5			1	1						2
p6		1	1							2
p7				1	1				1	3
p8					1					1
	1	1	2	3	3	2	3	1	1	17

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
p1		1			1		1			3
p2	1							1		2
p3						1	1			2
p4			1	1						2
p5				1	1					2
p6				1	1					2
p7			1			1			1	3
p8							1			1
	1	1	2	3	3	2	3	1	1	17

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
s1		1	1							2
s2	1		1							2
s3	1	1								2
s4				1	2	1				4
s5				1			3			4
s6				2						2
s7				1	3					4
s8									1	1
s9								1		1
	2	2	2	4	4	2	4	1	1	22

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
s1						1				1
s2			1							1
s3		1		1						2
s4			1		1		1	1	1	5
s5				1			1		1	3
s6	1						1			2
s7				1	1	1		1		4
s8				1			1			2
s9				1	1					2
	1	1	2	5	3	2	4	2	2	22

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
s1								1		1
s2					1		1			2
s3				1		1			1	3
s4			1		2					3
s5		1		2			1			4
s6			1				1		1	3
s7		1			1	1				3
s8	1									1
s9			1			1				2
	1	2	3	3	4	3	3	1	2	22

(a) Bipartite network $OBS_{bipartite}$ (top left) with two rnd. rewirings R_1 and R_2 (top mid/right) and M^TM projections OBS_{proj}, RP_1 and RP_2 (bottom)

	s1	s2	s3	s4	s5	s6	s7	s8	s9	
s1		0,07	0,17	0,26	0,25	0,17	0,26	0,08	0,08	1,34
s2	0,07		0,16	0,26	0,26	0,16	0,26	0,08	0,08	1,33
s3	0,17	0,16		0,54	0,55	0,34	0,52	0,16	0,16	2,62
s4	0,26	0,26	0,54		0,84	0,53	0,85	0,26	0,25	3,80
s5	0,25	0,26	0,55	0,84		0,54	0,85	0,25	0,26	3,81
s6	0,17	0,16	0,34	0,53	0,54		0,54	0,16	0,16	2,62
s7	0,26	0,26	0,52	0,85	0,85	0,54		0,26	0,27	3,82
s8	0,08	0,08	0,16	0,26	0,25	0,16	0,26		0,08	1,33
s9	0,08	0,08	0,16	0,25	0,26	0,16	0,27	0,08		1,34
	1,34	1,33	2,62	3,80	3,81	2,62	3,82	1,33	1,34	22

	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1		0,07	0,14	0,19	0,19	0,14	0,19	0,07	0,07
s2	0,07		0,14	0,19	0,19	0,13	0,19	0,07	0,07
s3	0,14	0,14		0,35	0,35	0,25	0,34	0,14	0,14
s4	0,19	0,19	0,35		0,46	0,36	0,46	0,19	0,19
s5	0,19	0,19	0,35	0,46		0,35	0,46	0,19	0,19
s6	0,14	0,13	0,25	0,36	0,35		0,35	0,14	0,14
s7	0,19	0,19	0,34	0,46	0,46	0,35		0,19	0,19
s8	0,07	0,07	0,14	0,19	0,19	0,14	0,19		0,07
s9	0,07	0,07	0,14	0,19	0,19	0,14	0,19	0,07	

(b) $\mathbb{E}(RP)$ and $\mathbb{SD}(RP)$ derived from n rewirings of example matrix

	s1	s2	s3	s4	s5	s6	s7	s8	s9
s1		**3,55**	**2,23**	-0,59	-0,58	-0,46	-0,59	-0,29	-0,29
s2	**3,55**		**2,27**	-0,59	-0,59	-0,44	-0,59	-0,29	-0,29
s3	**2,23**	**2,27**		-0,92	-0,93	-0,67	-0,90	-0,44	-0,44
s4	-0,59	-0,59	-0,92		0,23	**2,46**	0,21	-0,59	-0,58
s5	-0,58	-0,59	-0,93	0,23		-0,91	**3,16**	-0,58	-0,59
s6	-0,46	-0,44	-0,67	**2,46**	-0,91		-0,92	-0,44	-0,44
s7	-0,59	-0,59	-0,90	0,21	**3,16**	-0,92		-0,59	-0,60
s8	-0,29	-0,29	-0,44	-0,59	-0,58	-0,44	-0,59		**3,45**
s9	-0,29	-0,29	-0,44	-0,58	-0,59	-0,44	-0,60	**3,45**	

(c) Matrix $STDRES$ of standardized residuals $OBS_{proj} - \mathbb{E}(RP)$, normalized by $\mathbb{SD}(RP)$

Figure 6.9: Example for FDSM-based derivation of significant ties between art styles

Following the approach illustrated in Figure 6.9, a set of 10,000 random networks each featuring the same entities as the original pruned bi-partite network described in Subsection 6.1.2, consisting of 55,499 persons and 375 styles interlinked via 93,197 links placed at random was calculated under the constraint that each person and style had exactly the same degree as the original pruned network. A threshold of three standard deviations was used to filter the standardized style-style residuals obtained from the procedure outlined above, which left a network of 355 styles interlinked via 7,862 ties. 20 styles were discarded because all their ties to other styles were below the applied threshold.

The visualization of the resulting network is presented in Figure 6.10. In a similar manner as observed for the network of nationalities shown in Figure 5.27, quite clearly delineated groups of densely interrelated styles could be identified using Louvain clustering, highlighted in the plot via dedicated node colors and also shown explicitly in Table 6.6. Generally following the structural observations made for the immediate visualization of the pruned bipartite network from Figure 6.8, the identified clusters on the one hand appeared to group the present styles based on the broad seperation between fine arts and architecture, on the other hand also according to their chronological succession.

Shown on the leftmost side of the plot, the largest identified cluster represented a compound group of styles ranging across multiple centuries from Romanesque to Baroque arts and architecture styles (cluster #1, red). It was mainly connected to a group of styles around Classicism (cluster #11, green) representing the beginning of the observed separation between art and architecure styles into relatively distinct strands. Following this bifurcaton, the bottom of the Figure is dominated by three consecutive clusters of architecture styles, the first (cluster #6, light blue) of which featuring many "revival styles" popular during the 19^{th} century period of Historicism, followed by a smaller group (cluster #4, black) representing more modern late 19^{th}/early 20^{th} century styles including Industrial Architecture as well as Art Nouveau/Art Deco. A larger group of $20^{\text{th}}/21^{\text{st}}$ century architecture styles (cluster #3, yellow) marks the contemporary end of this view on architectural developments. Representing the fine art strand, the top of the Figure is in turn dominated by clusters of 19^{th}-21^{st} century art styles which were also found to follow a general chronological order in parallel to the architecture styles. Starting with Romanticism, a cluster of 19^{th} century art styles (cluster #10, orange) covered most developments of that century, including Academic Art and various movements that subsequently developed out of opposition to it, leading to a number of turn-of-the century styles around Post-Impressionism at the onset of the Avant-Garde. The latter was found to be strongly represented in the following clusters of pre-WWII modern art movements shown further to the right (cluster #12, light gray; cluster #13, dark green; cluster #2, dark orange; cluster #8, middle blue) which also included the majority of the styles featured in the Barr diagram, the latter additionally highlighted in bold font in Table 6.6 for easier comparison. Their concentration in a specific, chronologically sound region of the visualization again suggested the overall correctness of the underlying relationships derived from Wikidata and Wikipedia. Further to the right, the Figure features a large cluster of post-WWII styles (cluster #7, blue), including Abstract Expressionism and

other styles mainly emerging from North-America, but also European postwar movements. The tight interconnection of the styles featured in this group, resulting in its relatively clear cut visual separation from the pre-war styles in the Figure, appeared to underscore the radical changes to art and culture going hand in hand with the global conflict. Grouping a handful of late 20th/early 21st century styles (cluster #14, dark gray) which to a large extent covered new media forms gradually shifting towards the realm of the digital, the cluster shown at the outer right of the Figure marked the contemporary end of the fine art strand, at the same time suggesting the beginning of a new era of artistic development to be carried on by future generations.

Two clusters showed additional developments beyond the general picture of modern-age Western art history presented in the Figure. As also present on individual level in both the biographical and the bi-partite network of persons and styles and shown in their respective visualizations, a dedicated cluster of Japanese styles (cluster #5, pink) was mainly connected to late 19th and early 20th century Western art styles, revealing mutual stylistic influence emerging from the gradual approach between the different cultures. A cluster of ancient Greek styles (cluster #9, light pink), shown on the left of the Figure, was in turn completely detached from the other styles featured in the scene, which again underscored the century-long gap caused by limited availability of traces of artistic activity between antiquity and the late middle ages.

Due to the bi-partite projection and the subsequent residual-based filtering of style-style links, the two branches of art and architecture styles emerging after Classicism appeared to be more separated from each other in the projected and filtered network shown in Figure 6.10 than in the direct visualization of the bi-partite person-style network in Figure 6.8, where many art and architecture styles were still often connected via individual persons forming weak ties which were filtered out in the projected and filtered version. Some notable exceptions to the separation of the two strands, in turn, stood even more out: While Art Nouveau and its siblings such as Jugendstil or Modernisme were located between art and architecture in both versions of the network, the latter were shown much less prominent in the bi-partite visualization due to their lower absolute number of connections. Taking this difference into account, the residual-based approach put them much more center-stage instead. Similarly, the mixed domains of the members of some WWI and interwar period styles beyond Bauhaus, such as especially Constructivism, Purism, New Objectivity or De Stijl, were very barely noticeable in the bi-partite version although these styles appeared as very significant bridges between the two domains in the residual-based network. Interestingly, besides the very general designation of Post-Modernism, none of the featured post-WWII styles featured such a strong interaction between the two domains. This again also became especially visible in the residual-based version, which suggested the continuation of the two separate streams into the future.

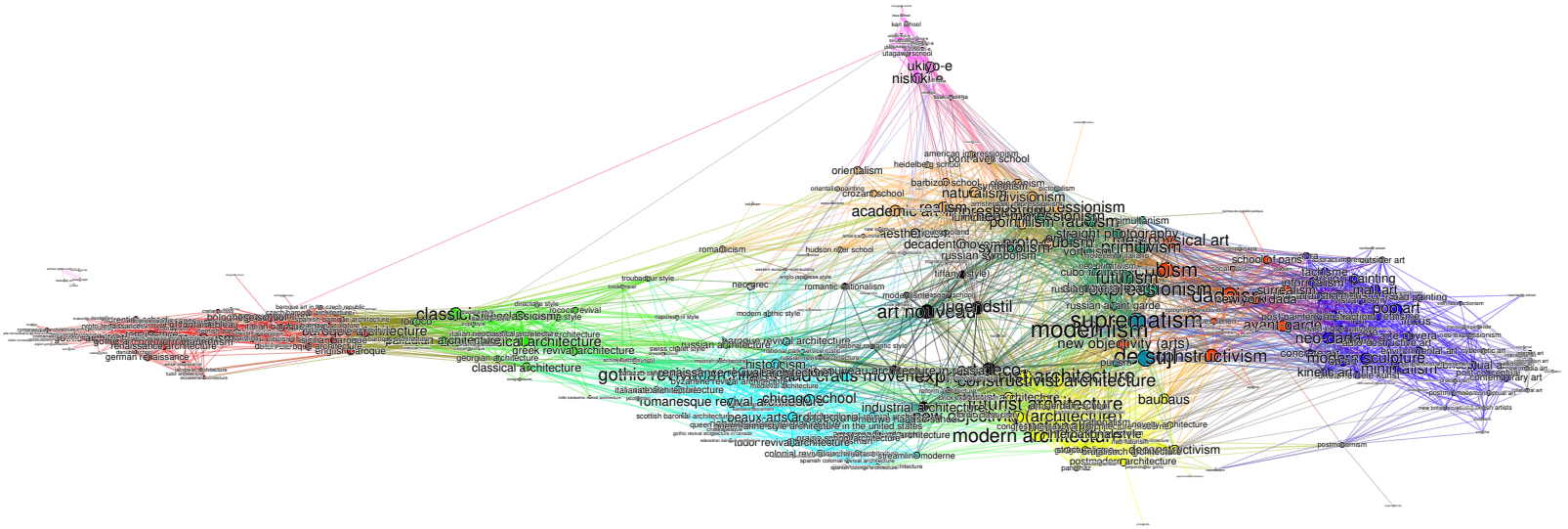


Figure 6.10: Projection of pruned person-style network to art & architecture styles (**A8**)

1
mannerism
baroque architecture
baroque
gothic architecture
northern mannerism
english baroque
baroque painting
school of fontainebleau
gothic art
renaissance
italian baroque art
high renaissance
renaissance architecture
dutch baroque architecture
florentine school
antwerp school
bolognese school
german renaissance
sicilian baroque
early netherlandish painting
louis xiii style
czech baroque architecture
spanish baroque architecture
proto-renaissance
baroque sculpture
italian baroque architecture
flemish baroque painting
romanism
haarlem school of painting
baroque art in the czech republic
french renaissance architecture
dutch golden age painting
utrecht caravaggism
plateresque
danube school
fantastic art
romanesque art
cretan school
churrigueresque
jacobean architecture
early renaissance
romanesque architecture
flamboyant
spanish baroque
weser renaissance
italian gothic architecture
tudor architecture
elizabethan architecture
french gothic architecture
brabantine gothic
pre-romanesque art/architecture
rayonnant
norman architecture
baroque architecture in portugal
cuzco school
catalan gothic
english gothic architecture
first romanesque
italian romanesque architecture
mozarab
brick gothic
mudéjar
cartilage baroque
stroganov school

2
cubism
dadaism
constructivism
metaphysical art
avant-garde
new objectivity (arts)
new york dada
surrealism
school of paris
cologne progressives
neues sehen
social realism
antropophagia
magic realism
poetism
socialist realism
fantastic realism
agitprop
peintres de la réalité poétique

3
futurist architecture
modern architecture
expressionist architecture
constructivist architecture
new objectivity (architecture)
functionalism
bauhaus
international style
deconstructivism
stalinist architecture
rationalism
postmodern architecture
amsterdam school
brutalism
high-tech architecture
congrès intl. d'archit. moderne
structuralisme
novelty architecture
googie
nordic classicism
new formalism
panelház
neo-futurism
paulista school
anglo-saxon architecture
perpendicular gothic
traditionalist school (architecture)
experimental architecture
pittura colta

4
art nouveau
art deco
jugendstil
industrial architecture
art nouveau architecture in russia
tiffany (style)
modernisme
glasgow school
romantic nationalism
brick expressionism
national romantic style
reform architecture
russian revival architecture
postconstructivism
moorish architecture
École de nancy

5
ukiyo-e
nishiki-e
utagawa school
kanō school
nikuhitsuga
surimono
uki-e
sōsaku-hanga
aizuri-e
ishizuri-e
uchiwa-e
torii school
yokohama-e
benizuri-e
katsukawa school
urushi-e
yōga
tosa school
nihonga
shin-hanga
hasegawa school

6
gothic revival architecture
arts and crafts movement
chicago school
eclecticism
beaux-arts architecture
romanesque revival architecture
renaissance revival architecture
historicism
richardsonian romanesque
nieuwe haagse school
tudor revival architecture
victorian architecture
american foursquare
dutch colonial revival architecture
american craftsman
mayan revival architecture
queen anne style archit. in the USA
baroque revival architecture
colonial revival architecture
prairie school architecture
italianate architecture
streamline moderne
queen anne style architecture
scottish baronial architecture
modern gothic style
palazzo style architecture
byzantine revival architecture
mission revival architecture
american renaissance
spanish colonial revival architecture
medieval architecture
mediterranean revival architecture
national park service rustic
châteauesque
eastlake movement
gothic revival archit. in canada
spanish colonial architecture
jacobethan
edwardian architecture
edwardian baroque architecture
western-european-style-building
cottage style

7	neo-dada	10	proto-cubism	12	modernism
	modern sculpture		impressionism		aestheticism
	abstract expressionism		pointillism		russian symbolism
	minimalism		post-impressionism		decadent movement
	pop art		academic art		neo-grec
	mail art		realism		young poland
	color field		divisionism		anglo-japanese style
	action painting		naturalism		migration period art
	kinetic art		pont-aven school		post-romanticism
	arte povera		**synthetism**		norse art
	conceptual art		cloisonnism		angry penguins
	informalism		luminism		belorusian national architecture
	concrete art		barbizon school		verism
	lyrical abstraction		orientalism	13	**expressionism**
	nouveau réalisme		crozant school		**fauvism**
	nul		american impressionism		**neo-impressionism**
	op art		romanticism		**futurism**
	tachisme		heidelberg school		primitivism
	auto-destructive art		hudson river school		symbolism
	fluxus		amsterdam impressionism		**orphism**
	zero		orientalist painting		degenerate art
	cobra		american luminism		straight photography
	fundamentele kunst		new sculpture		vorticism
	post-painterly abstraction		costumbrismo		cubo-futurism
	funk art		rustic architecture		simultanism
	bad painting		volcano school		russian futurism
	environmental art		indianism		russian avant-garde
	contemporary art		hoosier group		neo-primitivism
	visual poetry		australian tonalism		novecento italiano
	outsider art	11	classicism		pictorialism
	hard-edge painting		neoclassical architecture	14	new media art
	post-conceptual		palladian architecture		cybernetic art
	postminimalism		classical architecture		postmodernism
	abstract impressionism		rococo		electronic art
	neo-expressionism		neoclassicism		information art
	young british artists		greek revival architecture		relational art
	neo-conceptual art		russian architecture		interactive art
	light and space		empire style		internet art
	new british sculpture		georgian architecture		neomodern
	transavantgarde		egyptian revival architecture		altermodern
	viennese actionism		italian neoclassical architecture		interstitial art
	process art		louis xvi style		postinternet
	feminist art movement		rundbogenstil		supermodernity
	neue wilde		rococo revival		
	stuckism		troubadour style		
	superflat		adam style		
	photorealism		directoire style		
	chicago imagists		napoleon iii style		
	20th-century classical music		french baroque architecture		
	northwest school		zopf style		
	new leipzig school		schinkel school		
	anti-poetry		gustavian style		
	maya architecture		swiss chalet style		
8	**suprematism**		biedermeier		
	de stijl		russian baroque		
	purism		ukrainian baroque		
9	hellenistic art		dresden school		
	archaic greek sculpture		frederician rococo		
	severe style		indo-saracenic revival architecture		
	pioneer group		octagon house		
	neo-attic				

Table 6.6: Detected community structure in the style to style residual network

6.1.4 Summary

The framework **M2** which was used for the analysis of the biographical Wikipedia network was extended to enable the incorporation and representation of art and architecture styles as second type of entity in the network. This included the increased use of Wikidata to source descriptive metadata about styles and persons and resulted in framework **M5** which allowed another bottom-up large-scale view on how art history is represented in the free encyclopedia. The analysis of the nationalities of the persons interlinked with the different styles revealed the "regionality" of some styles which were almost exclusively linked to persons of one or a few culturally closely related nationalities, while others featured more even distributions in this regard. As it especially became visible for some of the more regional styles, the language-specific preference of culturally related biographies in different Wikipedia language versions also appeared in the context of articles about artistic styles, seemingly following comparable modes of preference.

Local preferences also contributed to the identification of many non-ULAN artists, including many persons having nationalities considered to be from the periphery of the art world, such as Japan or the former Soviet Union. Since some articles about styles also covered other artistic domains and were also interlinked with a significant number of "non-fine-artists", however, the group of newly identified persons contained many other professions as well. Although this was also observed in the ULAN which included important patrons but also other persons of public interest which would, if at all, rather be considered as "hobby-artists", the proportion of persons unrelated to fine art was much higher in the non-ULAN group identified via Wikipedia links to art and architecture styles. This left room for finding additional methods to identify and narrow down on selecting a more clearly delineated group of fine arts and architecture related persons on Wikipedia.

Similar to the varying distributions of nationalities of persons linked to the individual styles, the roles of the retrieved persons were sometimes more evenly distributed for some of the styles they were linked to, and sometimes their distributions were much more specific. Overall it became visible that broad period designations such as Baroque or Renaissance had much more even distributions of both role and nationality. Comparable to the observed regionality of some styles with respect to nationality, some styles were in turn more clearly delineated with respect to role. While the separation was quite obvious for many fine art and architecture styles which were almost exclusively linked to either painters or architects, it was also interesting to see "mixtures" such as for Dadaism dominated by painters and writers, or Minimalism to be "shared" between painters and composers.

As far as the visualizations were concerned, the pruning of "long-distance" links via method **M6** again revealed a clear chronological structure embedded in the person-style links, resulting in artifacts **A7** (Figure 6.8) and **A8** (Figure 6.10), comprehensive maps of temporally interlinked styles which could potentially serve as navigational aids for online collections of artworks but also as data-driven extensions of related scholarly diagrams: The comparison of the created visualizations with Barr's manually created diagram from the 1930's for example revealed that the Wikipedia content broadly matched the

condensed view presented in its scholarly counterpart, but in addition provided a much broader historical context for the specific period covered there which also included related styles not deemed relevant by the art historian.

Another interesting aspect of the chronological visualizations was the quite strict separation of fine arts and architecture styles which generally appeared to follow the increasing distinction between the two fields after the end of the Baroque era, coinciding with a number of societal developments especially in Europe which included the diminishing influence of religion, the rise of the bourgeous class and the onset of art history as scientific discipline. The visualization of the projected style-style network which was statistically filtered using method **M7** suggested an even greater seperation between the two domains but at the same time also highlighted some recent cases where they appeared to be quite closely connected again.

The observed visual separation between arts and architecture called for a more rigorous analysis in this regard. One straight-forward approach was to compare for each century the fraction of persons whose Wikipedia biographies were only linking to style articles tagged as either art movement or architecture style in Wikidata with the fraction of biographies interlinked with a mixture of articles about arts and architecture styles. Seen from this view, shown in Figure 6.11 for all persons on top and for architects, painters and sculptors in particular at the bottom, especially the times of the (early) Renaissance appeared to be the days of the polymaths, where artists also served as architects and vice versa, a notion which strongly declined after that period and only slightly returned at the beginning of the 20^{th} century. Looking at the individual roles, however, revealed an interesting difference between architects and artists, suggesting that a notable fraction of the architects remained connected with other fine art forms also in later times, which appeared to be clearly less the case for artists.

Figure 6.11: Fractions of persons connected to either only or both art or/and architecture styles over the centuries

6.2 Identifying domain specific persons via co-occurring occupations in Wikidata

The analysis of the bi-partite person style network in Section 6.1 revealed that multilingual Wikipedia featured many articles about artists not featured in the ULAN and that Wikidata provided structured metadata for many of them. This observation highlighted the limitations of using only the ULAN as a single authority for identifying persons relevant to art history and the approach to find them via links to art and architecture styles offered itself as an alternative in the context of Wikidata/Wikipedia. Since this procedure also returned many non-artists, however, it left lots of room for improvement. Ideally, a good solution would be able to identify a more clearly delineated group of persons relevant to art history and could also be generalized to those from any other domain.

Many of the Wikidata person records linked to art and architecture styles featured one or more entries about their occupation, which included professions such as "painter", "sculptor" or "architect" for representatives from the domain of art history, but also many others for those related to other domains. Using this occupation information thus appeared as a promising general approach to identify more domain specific groups of persons in Wikidata. There were, however, many more designations for artistic professions

amongst the 664 roles found for the persons linked with the style articles and also the non-artist occupations appeared to be similarly diverse in this regard. It was thus of interest to find a way to automatically partition the different occupations present in Wikidata person records by their domain, e.g. to group all fine-arts occupations together and separate them from those related to other domains.

An immediate approach to find groups of related occupations was to consult their existing Wikidata concept hierarchy based on subclass relationships, featuring the entries "painter" and "sculptor" as subclasses of the entry "visual artist" for example, suggesting to use such superclasses for the identification of a person's domain. As of early 2015, however, the available hierarchy information was found to be very sparse and included only a few professions besides the mentioned ones. One alternative to using the incomplete concept hierarchy was to relate occupations with each other via person records with co-occurring occupation references, i.e. to count all persons tagged as both painter and illustrator and use this information as a basis for a quantitative measure of relatedness between the occupations "painter" and "illustrator". This was motivated by the results obtained from the analysis of the projected style network described in Subsection 6.1.3, which represented a similar bipartite case where the relatedness of styles was derived from mutually interlinked person records.

Following these considerations, this section therefore describes the results of a related experiment which used the occupation co-occurrence information present in Wikidata person records to identify domain specific groups of persons there. The main aim was to identify a group of art history related persons and to check if the resulting Wikipedia article link structure differed considerably from the one found between Wikipedia articles identified via ULAN mappings. Moreover, link structures between persons belonging to other domains should be considered as proof of concept that the co-occurrence based identification approach could be generalized beyond art history.

6.2.1 Data retrieval and filtering

A set of 1.363.032 Wikidata person records were selected based on the requirement that each record had a *P31:instance_of* reference to the class "human" (*Q5*) and featured at least one entry for the property *P106:occupation*. Altogether these records featured 1,706,766 references to 3,419 distinct occupations, but many of the latter were only referenced a few times. The set of persons was thus further reduced to those linking to occupations featured at least 100 times, which resulted in a set of 1,352,909 person records linking to 431 occupations via 1,685,000 references. Only about 19% these person records, however, featured at least two references to different occupations, which was a prerequisite to establishing co-occurrence between the latter, further reducing the set to 430 occupations co-occurring in 253,316 person records via 585,407 references.

The resulting co-occurence set was encoded in 0-1 matrix form similar to the person-style setting outlined in the top left of Figure 6.9 in Subsection 6.1.3, with 253,316 rows representing persons and 430 columns representing occupations, respectively. Following the FDSM procedure from [Zweig and Kaufmann, 2011], 5,000 degree preserving rewirings of this matrix were generated using the R package "BiRewire". Standardized residuals

between the occupation-occupation projection of the observed network and the mean of the projections of its randomly generated counterparts were subsequently calculated as described in Subsection 6.1.3. The resulting network of co-occurring occupations weighted by the standardized residuals was then filtered by discarding all links with weights below three standard deviations, leaving only the co-occurrences which appeared significantly more often than expected.

The filtered co-occurence network was then partitioned using the Louvain Modularity maximization approach as it was also used for partitioning the aggregated nationality and the projected style-style networks in Subsections 5.7 and 6.1.3. This was based on the expectation that the calculated partition would represent a usable clustering of occupations by domain.

Each of these occupation clusters was subsequently used to collect a domain-specific set of Wikidata person records with all the hyperlinks between their respective Wikipedia biographies. These networks were then filtered for absolute birth date difference less than or equal 75 years and visualized in the same way as it was done for the Wikipedia art history biographies identified via the ULAN in design artifact **A4** (Figure 5.19).

6.2.2 Partitioned network of occupation co-occurrencess

Figure 6.12 shows a force-based visualization of the filtered occupation co-occurrence network created using the ForceAtlas2 algorithm in Gephi. The thickness of each line represents the magnitude of the standardized residual, i.e. the strength of the association between to co-occurring occupations. The size of each occupation node reflects the number of Wikidata person records annotated with it, while the size of its label shows its degree, i.e. how many significant co-occurrences each occupation has with others. The colors of the nodes show the partitions derived via Louvain clustering.

The visualization clearly shows that the topology of the filtered co-occurrence network tended to form clusters of occupations which appeared to be quite closely related to each other, protruding from the center like arms of a starfish. The visually clustered layout corresponded well with the partitions found via Louvain clustering, which divided some regions in even more detail. A cluster of sportspersons (pink) and related occupations, shown on the lower left side, appeared to be most clearly delineated from the others. As shown via the node colors, this "sports cluster" was further divided into various subdisciplines such as, amongst others, winter sports, athletics, martial arts or motor sports. Similarly well delineated domain clusters included musicians (orange, top left), visual artists (greenish yellow, top center), clergymen (cyan, top right) and life-sciences professions (dark pink and eggplant, lower right).

Closer to the center were occupations which increasingly tended to co-occur with a variety of different domains. This included broad occupations related to writing (turquoise, upper center) and media (brown, upper center), such as blogger, autobiographer or author, and business related occupations (green, lower center), such as entrepreneur, businessperson, philanthropist or manager. A number of cross-domain co-occurrences between more specialized occupations could also be observed. In the context of sports related occupations, this for example included the close relationship between motor sports

related professions and engineers, polo players which were also stockbrokers and financiers, or professional wrestlers also being models and actors. In the context of visual arts professions, such co-occurrences appeared with business occupations such as art collector or art dealer, with related humanities disciplines such as art-, architecture- or cultural historians or with media/writing related professions such as art critics. Moreover, some specific artistic professions were closely related to other domains, including, amongst others, illuminators related to hagiographers, botanical illustrators with botanists, photographers with various media related occupations.

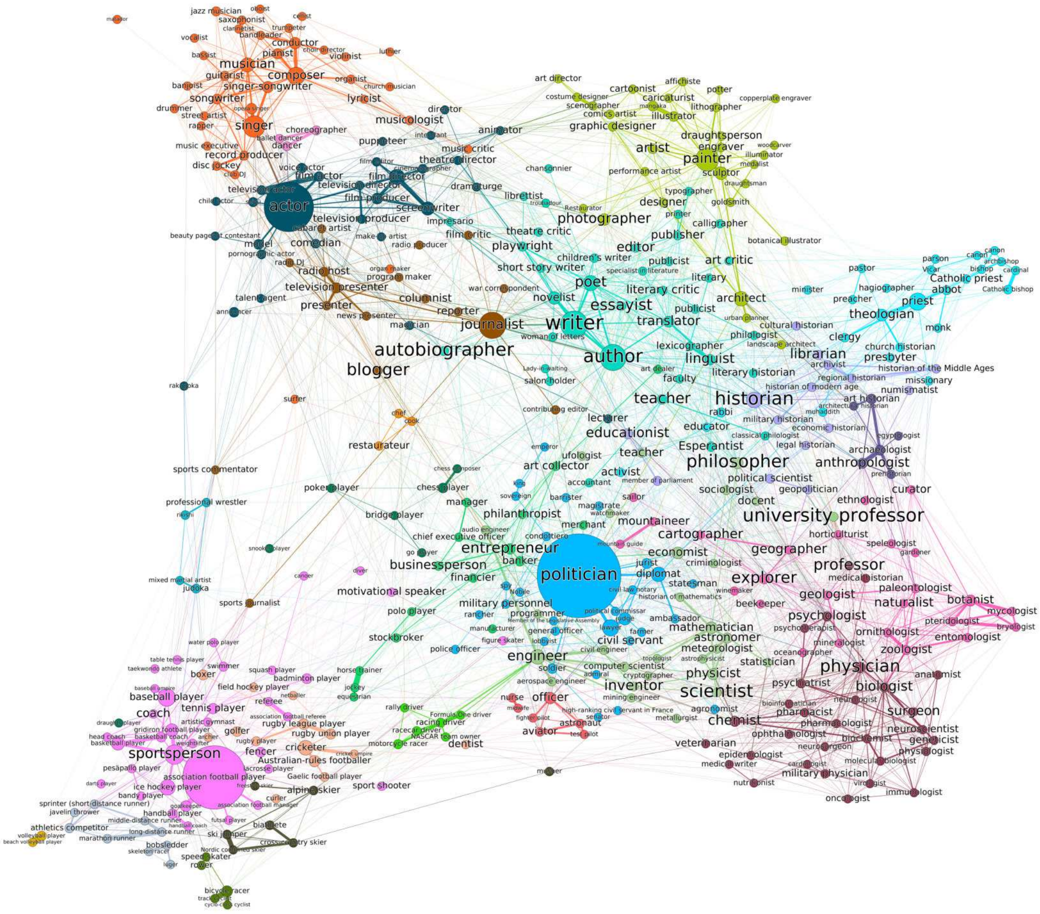


Figure 6.12: Filtered network of co-occurring occupations (**A9**)

Table 6.7 gives a full, indexed overview on the identified occupation clusters. In general, the observed partitions suggested that the co-occurrence based clustering of occupations was a feasible approach to identify domain specific groups of professions. As far as the cluster #9 of art history occupations was concerned, all the co-occurring occupations related to visual arts "creators" were grouped together with high recall, i.e. based on the inspection of the clusters related to other domains, no visual art profession was included elsewhere. Concerning precision, however, it could be argued that some of the professions were not immediately related to fine arts and should therefore rather be classified as belonging to other, related domains such as creative industries or artisanry. This included professions such as "art director", "comics artist", "goldsmith", "botanical illustrator" or "urban planner". The concept of what counts as fine art and what not, however, has never been stable throughout history and also varies significantly across different cultural regions. Many Renaissance artists were active as artisans, urban planners, botanical illustrators, while comics for example were reflected in Western post-WWII pop art and also play a significant role in e.g. Japanese contemporary culture. It could therefore equally well be argued that the inclusion of these professions reflects a broader view on the domain of fine arts in general, taking both historical as well as contemporary developments into account.

Art history related professions which did not involve "creators", however, were not detected as belonging to the visual arts cluster but as part of different ones. This especially included art and architecture historians, critics, collectors and dealers, which were found in the clusters #23, #1 and #13 instead. This stood in contrast to the domain of music represented by cluster #4, where musicologists and music critics were identified as members alongside instrumentalists, singers and composers.

Some clusters, however, featured members that were less related with each other. The "literature" cluster #1 for example consisted of writers, authors and poets, but also of language scholars (linguist, philologist, ...) and professionals related to book production (printer, publisher, editor, ...), while sports related clusters (#2, #6, #7, #8, #11, #14, #17, #21, #22) sometimes featured a mixture of different sports sub-disciplines. Some of the clusters also contained unrelated occuations, such as the "government" cluster #15 featuring politicians, kings, emperors, diplomats and other government related roles, but also some occupations related to agriculture. The cluster #18 of biology-related occupations such as botanist and zoologist did not include their "superclass" biologist — which was part of a separate cluster #12 together with rather medicine-related occupations — but in turn featured some agricultural and earth science occupations together with professions such as explorer or sailor. Especially the latter co-occurrences reminded of the age of discovery, where seafarers were often accompanied by scientists of different disciplines, providing some potential explanation for their presence in a common cluster. Many groups were moreover found to contain very broad occupation designations, such as educator or teacher, which were related to multiple domains. For each cluster shown in Table 6.7, "borderline" members that could also be assigned to other existing or new groups are shown in italic. Occupations that did not seem to fit into their cluster at all are displayed in bold instead.

1	writer
	author
	poet
	novelist
	translator
	linguist
	teacher
	Esperantist
	playwright
	autobiographer
	publisher
	essayist
	woman of letters
	philologist
	children's writer
	lexicographer
	literary critic
	editor
	literary historian
	salon-holder
	publicist
	art critic
	calligrapher
	troubadour
	specialist in literature
	Lady-in-waiting
	librettist
	printer
	theatre critic
	faculty
	typographer
	chansonnier
	literary
	classical philologist
	accountant
	publicist
	short story writer
	impresario
2	athletics competitor
	sprinter (short-distance runner)
	marathon runner
	long-distance runner
	middle-distance runner
	javelin thrower
	bobsledder
	skeleton racer
	luger
3	choreographer
	dancer
	ballet dancer

4	singer
	composer
	musician
	conductor
	singer-songwriter
	opera singer
	pianist
	songwriter
	guitarist
	jazz musician
	record producer
	disc jockey
	musicologist
	organist
	saxophonist
	bandleader
	trumpeter
	rapper
	violinist
	drummer
	clarinetist
	banjoist
	bassist
	cellist
	surfer
	organ maker
	choir director
	club DJ
	vocalist
	street artist
	oboist
	lyricist
	luthier
	music critic
	matador
	church musician
	music executive
5	*officer*
	aviator
	nurse
	astronaut
	test pilot
	fighter pilot
	midwife
6	motorcycle racer
	racing driver
	Formula One driver
	racecar driver
	NASCAR team owner
	rally driver

7	association football player
	sportsperson
	baseball player
	ice hockey player
	basketball player
	gridiron football player
	goalkeeper
	tennis player
	badminton player
	basketball coach
	fencer
	handball player
	swimmer
	association football manager
	table tennis player
	sport shooter
	head coach
	coach
	weightlifter
	association football referee
	artistic gymnast
	futsal player
	squash player
	lacrosse player
	bandy player
	pesäpallo player
	taekwondo athlete
	darts player
	water polo player
	motivational speaker
	referee
	handball coach
	baseball umpire
	diver
	figure skater
	canoer
8	cricketer
	Australian-rules footballer
	rugby league player
	rugby union player
	boxer
	golfer
	rugby player
	field hockey player
	Gaelic football player
	curler
	archer
	dentist
	cricket umpire
	netballer

9	painter
	architect
	sculptor
	photographer
	mangaka
	artist
	illustrator
	comics artist
	graphic designer
	designer
	cartoonist
	engraver
	lithographer
	caricaturist
	potter
	scenographer
	performance artist
	draughtsman
	costume designer
	art director
	urban planner
	botanical illustrator
	draughtsperson
	Restaurator
	illuminator
	goldsmith
	affichiste
	woodcarver
	landscape architect
	medalist
	copperplate engraver
10	journalist
	television presenter
	radio host
	presenter
	contributing editor
	news presenter
	columnist
	cabaret artist
	film critic
	blogger
	war correspondent
	radio DJ
	sports commentator
	sports journalist
	reporter
	program maker
	radio producer
11	volleyball player
	beach volleyball player

12	physician
	chemist
	psychologist
	biologist
	professor
	psychiatrist
	biochemist
	military physician
	neuroscientist
	surgeon
	pharmacist
	psychotherapist
	ophthalmologist
	veterinarian
	geneticist
	pharmacologist
	immunologist
	cardiologist
	medical historian
	epidemiologist
	molecular biologist
	neurosurgeon
	neurologist
	nutritionist
	anatomist
	oncologist
	virologist
	physiologist
	bioinformatician
	medical writer
13	entrepreneur
	banker
	businessperson
	jockey
	art collector
	merchant
	horse trainer
	financier
	manager
	chief executive officer
	polo player
	stockbroker
	equestrian
	philanthropist
	manufacturer
	art dealer
14	judoka
	rikishi
	professional wrestler
	mixed martial artist

15	politician
	lawyer
	diplomat
	judge
	jurist
	soldier
	farmer
	military personnel
	condottiero
	ambassador
	magistrate
	sovereign
	police officer
	general officer
	political commissar
	civil servant
	civil law notary
	rancher
	lobbyist
	barrister
	emperor
	agronomist
	statesman
	senator
	spy
	king
	Nobile
	admiral
	high-ranking civil servant in France
	Member of the Legislative Assembly
16	historian
	educationist
	librarian
	member of parliament
	political scientist
	archivist
	historian of modern age
	historian of the Middle Ages
	regional historian
	legal historian
	military historian
	numismatist
	economic historian
	geopolitician
	cultural historian
17	bicycle racer
	speed skater
	rower
	track cyclist
	cyclo-cross cyclist

18	botanist	20	actor	24	engineer
	explorer		film director		mathematician
	zoologist		screenwriter		economist
	ornithologist		film producer		*philosopher*
	paleontologist		model		physicist
	geographer		announcer		astronomer
	entomologist		voice actor		computer scientist
	geologist		film actor		*university professor*
	cartographer		seiyū		civil engineer
	mycologist		television actor		*scientist*
	sailor		cinematographer		inventor
	curator		pornographic actor		sociologist
	naturalist		comedian		audio engineer
	horticulturist		film editor		statistician
	mountaineer		theatre director		aerospace engineer
	winemaker		television producer		metallurgist
	gardener		television director		astrophysicist
	pteridologist		rakugoka		meteorologist
	speleologist		animator		historian of mathematics
	ethnologist		magician		*teacher*
	mountain guide		beauty pageant contestant		programmer
	bryologist		puppeteer		*docent*
	beekeeper		director		topologist
	mineralogist		make-up artist		criminologist
	oceanographer		**lecturer**		cryptographer
19	priest		child actor		mining engineer
	theologian		talent agent		**ufologist**
	rabbi		intendant		watchmaker
	abbot		dramaturge	25	chef
	Catholic bishop	21	chess player		cook
	presbyter		go player		restaurateur
	clergy		poker player		
	church historian		snooker player		
	Catholic priest		bridge player		
	bishop		chess composer		
	monk		draughts player		
	cardinal	22	biathlete		
	missionary		ski jumper		
	pastor		alpine skier		
	educator		musher		
	minister		cross-country skier		
	archbishop		Nordic combined skier		
	preacher		freestyle skier		
	canon	23	anthropologist		
	Vicar		archaeologist		
	parson		art historian		
	muhaddith		prehistorian		
	activist		architectural historian		
	hagiographer		egyptologist		
	canon				

Table 6.7: Detected community structure in the occupation co-occurrence residual network

6.2.3 Wikipedia artist network identified via grouped art occupations

The well-delineated selection of fine-arts-related occupations grouped in cluster #9 suggested that it could indeed be used to identify Wikidata records about persons relevant to art history. A set of person records with matching occupations was thus gathered in order to extract its corresponding Wikipedia hyperlink network for comparison with the one identified via the ULAN. 1,068,942 Wikidata person records with at least occupation and birth/death date information retrieved via the properties listed in Table 6.1 served as data-pool for this purpose.

Top nationalities and occupations

A search for entries with occupations matching those from cluster #9 returned 73,917 Wikidata records. Table 6.8 shows the top 20 nationalities and occupations encountered amongst them. For both attributes, co-occurrences were again resolved by filtering to the most prominent value, as it was done for the person records identified via art and architecture styles. In order to quantify the differences induced by this filtering, each table entry also features the total counts for each of the top-20 values including all co-occurrences.

As far as nationalities are concerned, a significant subset of person records did not have any nationality information, the ranking of the remaing values included the same major nationalities which dominated the ULAN based group of persons. It was notable, however, to see Germans in a by far leading position, since they were ranked only at fourth position behind American, British and French persons in the ULAN mapped dataset. As it already appeared to be more prominent in the person-set retrieved via style-links in Section 6.1, the Soviet Union represented another exception, suggesting a higher fraction of USSSR artists than present in the set identified via the ULAN.

Looking at the occupations revealed that while the majority of persons was already covered by the largest four occupations "painter", "architect", "sculptor" and "photographer", there were quite a few additional person records which could only be identified via other occupations. The broad class "artist", which was used as preferred role for the majority of ULAN records, played a much smaller role in the artist records identified directly in Wikidata but nevertheless ranked as fifth largest occupation there. Illustrators, comics artists, graphics designers, manganka and cartoonists appeared in relatively large numbers too. Compared to the nationalities, the larger differences between filtered and unfiltered appearance counts showed that occupations tended to co-occur much more often in person records than nationalities.

Nation	count (w. cooc)	Occupation	count (w. cooc)
Germany	13152	painter	35461
unknown	9924	architect	13551 (14068)
France	7876 (7918)	sculptor	7975 (10172)
United States of America	7034 (7149)	photographer	7237 (7646)
Italy	6077 (6113)	artist	2238 (3091)
Netherlands	5529 (5567)	illustrator	1533 (2283)
United Kingdom	4010 (4070)	comics artist	1055 (1143)
Austria	2527 (2688)	graphic designer	626 (922)
Sweden	1532 (1552)	mangaka	614 (740)
Spain	1449 (1466)	cartoonist	528 (661)
Belgium	1411 (1457)	designer	494 (778)
Switzerland	1336 (1464)	caricaturist	282 (439)
Norway	1287 (1308)	draughtsman	260 (340)
Soviet Union	876 (886)	performance artist	245 (357)
Japan	869 (875)	engraver	226 (571)
Poland	737 (856)	potter	215 (309)
Australia	640 (664)	scenographer	204 (281)
Denmark	519 (549)	costume designer	190 (231)
Russian Empire	512 (877)	art director	160 (186)
Canada	485 (503)	lithographer	159 (550)

Table 6.8: Top-20 nationalities and occupations in Wikidata records identified via artist occupations

Coverage of ULAN records

29,767 (40.27%) of the 73,917 records identified via artist occupations were found to have a mapping to a corresponding ULAN record. As it turned out, the fraction of records having ULAN links was not stable across time: Figure 6.13 shows that records describing more contemporary artists tended to have no ULAN link, while those born until about 1800 tended to have such a mapping in turn. This was not found to be surprising, since especially the "older" records referred to famous representatives of art history who were more likely to have a presence in the ULAN than contemporary artists.

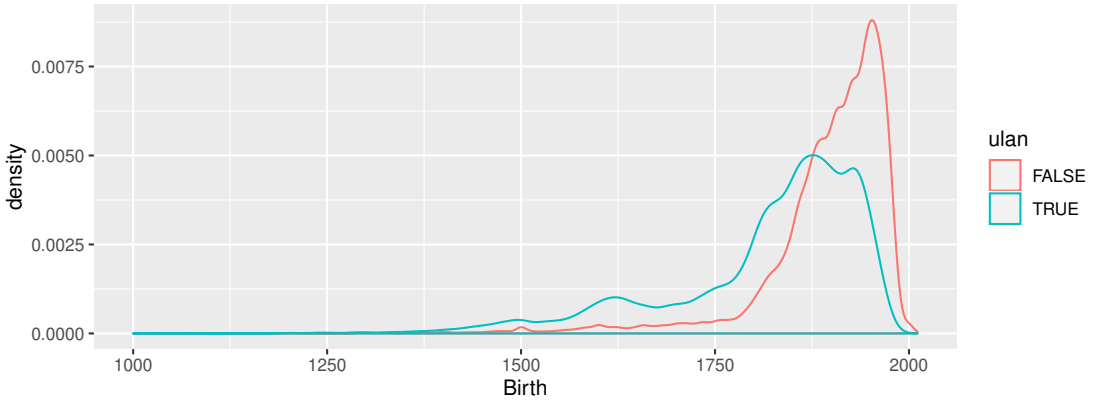


Figure 6.13: Wikidata artist records over time, with vs without ULAN mappings

Table 6.9 lists the top-25 occupation-identified artists whose Wikidata records had no ULAN mappings, ranked by their degree in the Wikipedia hyperlink network. The list

once more demonstrates the presence of many Soviet artists without corresponding ULAN records, confirmed by a manual check in the ULAN database. The lookup revealed that two of the 25 artists featured in the list actually had ULAN records but no mappings in Wikidata, which was better than expected from the findings in 6.1.1, where 22 out of 100 manually checked Wikidata painter records without ULAN mappings could nevertheless be identified there.

Sergei Ivanovich Osipov	Russia	painter	186	
Nikolai Pozdneev	Soviet Union	painter	178	
Nikolai Galakhov	Soviet Union	painter	173	
Vsevolod Bazhenov	Soviet Union	painter	169	
Alexander Semionov	Soviet Union	painter	166	
Taisia Afonina	Soviet Union	painter	165	
Irina Baldina	Soviet Union	painter	161	
Arseny Semionov	Soviet Union	painter	150	
Elena Kostenko	Soviet Union	painter	150	
Mikhail Natarevich	Soviet Union	painter	150	
Piotr Alberti	Soviet Union	painter	148	
Vladimir Gorb	Soviet Union	painter	139	
Lev Russov	Soviet Union	painter	138	
Dmitry Maevsky	Russia	painter	130	
Rudolf Frentz	Soviet Union	painter	123	X
Nadezhda Shteinmiller	Soviet Union	painter	109	
George Bernard Shaw	Ireland	artist	97	
Will Eisner	United States of America	comics artist	96	
John Byrne	United States of America	comics artist	92	
George W. Bush	United States of America	painter	84	
Valeria Larina	Soviet Union	painter	84	
Grant Morrison	Scotland	comics artist	83	
Brian Michael Bendis	United States of America	comics artist	77	
Tim Burton	United States of America	designer	75	X
André Franquin	Belgium	comics artist	73	

Table 6.9: Top 25 Wikidata records identified via art occupations, without ULAN mapping

Switching perspective, 17,696 of the 47,463 Wikidata records which had been identified via ULAN-mappings were not part of the set identified via art occupations. As it turned out, this included 2,526 (97,26%) out of the 2,597 non-artists (as defined via a separate ULAN facet) present in the ULAN-mapped set, effectively demonstrating that the occupation based mapping successfully acted as filter for non-artists. As far as the 15,170 ULAN-mapped artist records which had not been identified via art occupations were concerned, Table 6.10 shows the top-50 of them ranked by their degree in the ULAN-mapped Wikipedia network. The list contains famous writers and poets such as Breton, Baudelaire or Marinetti who were very influential for artists of their time. Moreover, it features art historians and other important writers such as Zola, who was included in the ULAN for his drawings, and persons from the domain of film. Although many of these persons had undeniable importance for art history, their ULAN categorization as "artist" could be questioned, and the occupation based selection was able to highlight these cases.

Breton, André	French	424	Humboldt, Alexander von	German	140
Baudelaire, Charles	French	328	Scorsese, Martin	American	138
Dickens, Charles	British	307	Shelley, Percy Bysshe	British	136
Cage, John	American	270	Stevenson, Robert Louis	British	136
Galilei, Galileo	Italian	256	Bergman, Ingmar	Swedish	135
Zola, Emile	French	248	Ceán Bermúdez, Juan Agustín	Spanish	134
Apollinaire, Guillaume	French	244	Wittkower, Rudolf	German	134
Vitruvius	Roman	219	Bernhardt, Sarah	French	132
Dumas, Alexandre	French	217	Rabelais, Francois	French	132
Tzara, Tristan	Romanian	217	Browning, Robert	British	126
Gautier, Théophile	French	214	Lorca, Federico García	Spanish	124
Andersen, Hans Christian	Danish	192	Dafforne, James	British	122
Welles, Orson	American	184	Pasolini, Pier Paolo	Italian	122
Godard, Jean-Luc	French	182	Pompadour, Jeanne Antoinette Poisson	French	122
Buñuel, Luis	Spanish	179	Yeats, William Butler	British	122
Borges, Jorge Luis	Argentine	175	Spielberg, Steven	American	121
Fellini, Federico	Italian	164	Valéry, Paul	French	120
Boccaccio, Giovanni	Italian	163	Eisenstein, Sergei	Russian	116
Sand, George	French	162	Allen, Woody	American	114
Éluard, Paul	French	157	Pope, Alexander	British	114
Disney, Walt	American	156	Bataille, Georges	French	113
Artaud, Antonin	French	153	Orleans, Philippe d'	French	109
Ibsen, Henrik	Norwegian	152	Eisenhower, Dwight D.	American	107
Marinetti, Filippo Tommaso	Italian	151	Musset, Alfred de	French	107
Wornum, Ralph Nicholson	British	146	Twain, Mark	American	156

Table 6.10: Top 50 Wikidata records about ULAN artists not found via art occupations

Network

54,374 of the 73,917 Wikidata artist records were found to be interlinked via 254,328 distinct links extracted from the combined page-link datasets from the 50 largest DBpedia language versions. Looking at the top-20 indegree-ranked persons in this network, shown in Table 6.11, again revealed a list of the most famous representatives of art history. Compared to the ranking of Wikipedia biographies identified via the ULAN, however, this version did not contain the various non-artists which were included in the ULAN, with the exception of persons such as Adolf Hitler who had also acted as creators of artworks but were mostly known for other reasons. Other than that, being based on the same link-set as the ULAN mapped version, the top-20 indegree ranking of the occupation based network featured all the artists that were also present in the top-20 indegree-ranking of the one identified via the ULAN.

As far as overall network characteristics were concerned, some differences were encountered regarding the tendency of persons to be connected to each other based on their attributes. While the tendency of persons of similar nationality to be connected with each other remained similar to the network of the ULAN mapped person-set (58.68% same-nationality-links/assortativity coefficient 0.54,ULAN mapped: 59.95%/assortativity coefficient 0.554), the fraction of links connecting persons of same occupation/role (75.82%/assortativity coefficient 0.58) was clearly higher than in the ULAN mapped dataset (67.65%/assortativity coefficient 0.257). Likewise, there was a higher fraction of links connecting persons born not more than 75 years apart (91.86%/birth assortativity coefficient 0.83, ULAN mapped: 85.24%/birth assortativity coefficient 0.63). The observed differences for occupation and birth assortativity appeared to be a clear

Pablo Picasso	1435
Raphael	822
Henri Matisse	782
Paul Cezanne	762
Rembrandt	762
Michelangelo	739
Vincent Willem Van Gogh	670
Peter Paul Rubens	657
Le Corbusier	602
Titian	573
Adolf Hitler	565
Leonardo da Vinci	532
Giorgio Vasari	510
Auguste Rodin	487
Johann Wolfgang von Goethe	473
Paul Klee	472
Andy Warhol	470
Wassily Kandinsky	464
Albrecht Dürer	455
Salvador Dalí	427

Table 6.11: Top 20 Wikidata artist records ranked by in-degree

consequence from the missing non-artist records, since the lack of artist-patron links for example represented a significant reduction of links between persons of different occupation and many of the "long distance" links in the ULAN mapped dataset were found to connect important non-artist figures from antiquity with persons from the modern age. The even more significant difference between the assortativity coefficients for occupation/role (0.58 vs 0.257) could additionally be explained by the fact that the roles for the ULAN mapped dataset where based on the ULAN "preferred role" attribute, for which a very high fraction (e.g. 78.75% of all persons in the GCC of the ULAN mapped Wikipedia biography network) of all ULAN records were assigned with the generic term "artist", while the person-set identified and described via the Wikidata occupation attribute featured a much higher variety of occupation assignments.

The network featured a GCC of 52,703 persons interlinked via 252,956 links which were reduced to 51,464 persons and 232,007 links after pruning relationships having a birth date difference of less than or equal 75 years. The visualization of the pruned network is shown in Figure 6.14 a), colored by nationality using the same color-code applied throughout this work. Three miniature visualizations are provided in the corners, representing b) the color-coded ULAN coverage (top left, green: with ULAN link, red: without ULAN link), c) the distribution of the roles across the network (top right) and d) the color-coded birth dates rounded to the next half-century (bottom left) showing

the chronological sequence of the visualized pruned network. The main body of the visualization resembled the one obtained from the ULAN mapped Wikipedia links from Figure 5.19 to a large degree, featuring a similar development ranging from the Italian and Northern Renaissance towards contemporary times, with quite distinct strands of closely interlinked persons of same nationality gaining and losing strength across the centuries, quite in accordance with the historical impact of the art produced there at various points in time.

In line with the different composition of nationalities compared to the ULAN mapped set of persons, however, some developments which had a very prominent presence in Figure 5.19, such as, e.g., the British strand (light blue), stood out much less in Figure 6.14 while others, such as Germans (black) or persons related to Russia, the Russian Empire or the Soviet Union (all colored in purple for convenience) had much more space there in turn. Moreover, since the occupation-based network did not feature non-artists, rather sparse strands such as the Japanese for example which had appeared as parallel development beneath the representation of the "core" developments in Figure 5.19 and whose contiguity was in part only enabled via links with important non-artists such as Pope Gregory XIII, were only represented as thin, feeler like extensions in Figure 6.14 which protruded there from the main body into various directions. Other extensions, however, were almost solely visible in Figure 6.14 in turn: As it was already found in the person-set identified via links to art styles, many artists from the Soviet Union, such as those from the so-called Leningrad School of Painting[7], had biographies in Wikipedia and person records in Wikidata, but no corresponding record in the ULAN. Shown as a satellite like extension at the bottom of the main body, this quite separate cluster of mainly painters appeared as a visual representation for literally being located "outside the canon of art history". Additional clusters with similar appearance could be found further to the right of Figure 6.14. As shown in the legend of the top right miniature plot colored by role, they represented illustrators and comic artists who were also rather shown as outsiders compared to the main developments, which also became visible in their similarly marginal ULAN coverage shown in the top left miniature. An interesting aspect of these clusters was their quite clear cut homogeneity with regards to the nationality of their members, mainly consisting of American, Italian, French, Belgian and Japanese artists.

Besides highlighting the clusters of illustrators and comic artists, the miniature visualization colored by occupation/role shown in the top right of Figure 6.14 also reflected the overall high assortativity of the Wikidata based network in this regard. Also quite well-delineated, but much more integrated with the "main body" were the rather international occupation clusters of architects and photographers shown at the contemporary end of the chronological visualization. Besides that, also many of the "national strands" that emerged in the layout appeared to feature further subdivisions according to occupation, which became visible for example for the central French strand or the Russian/Soviet cluster at the bottom, which both featured mutually interconnected but also clearly parallel sub-streams of architects, painters and sculptors.

[7] `http://www.leningradschool.com/`, retrieved Sept. 14th, 2020

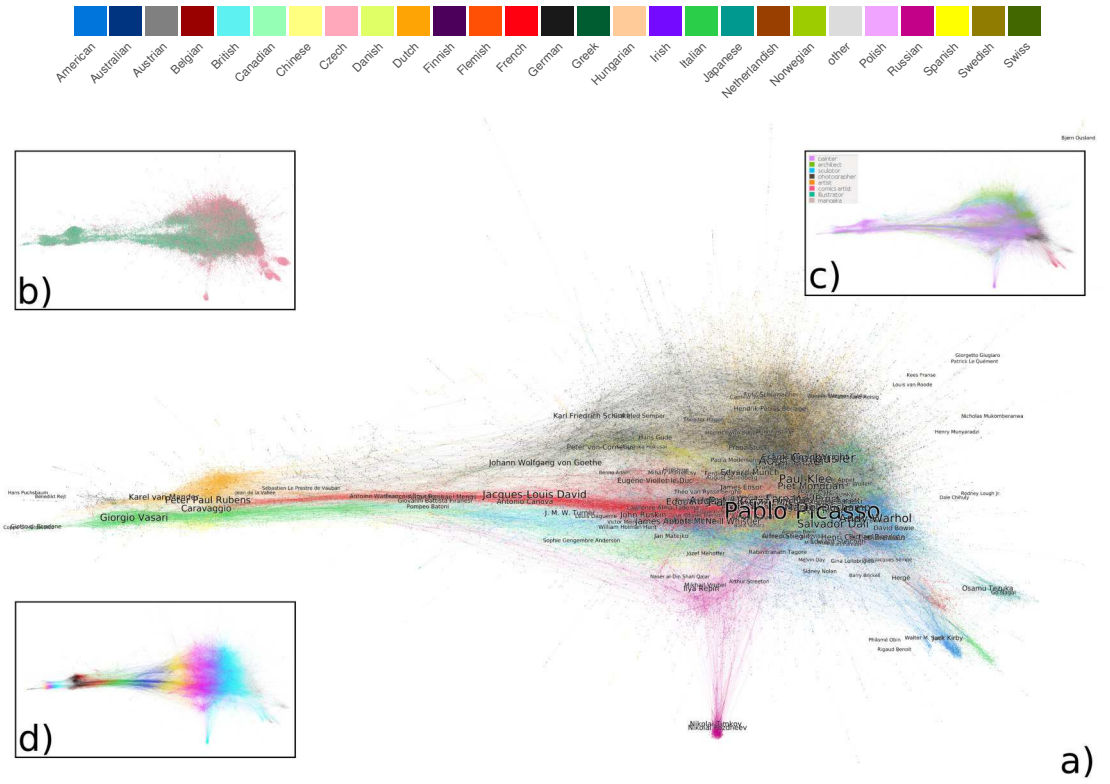


Figure 6.14: Artist network identified via clustered artistic Wikidata occupations, 51,464 persons, 232,007 links (**A10**)

Figure 6.15 provides a view on the individual contributions to the combined occupation-based artist network from the seven largest and the Japanese Wikipedia versions. It once more demonstrated that the combined network shown in Figure 6.14 was composed of a multitude of contributions from different Wikipedia editions each having specific cultural preferences. Although the combined network contained more persons than the one composed of ULAN-mapped persons, it was interesting to see that the six of the seven occupation-based artist networks identified in the largest language versions and also the one found in the Japanese Wikipedia featured less persons than their ULAN-mapped counterparts shown in Figure 5.25, which was to a large extent due to the now missing non-artists. Only the occupation-based artist network in the German Wikipedia featured a higher number of persons there, mainly due to painters and architects of German origin. Together with unique contributions about relevant persons featured in other language versions, these biographies contributed to the overall larger set of artists identified via occupations.

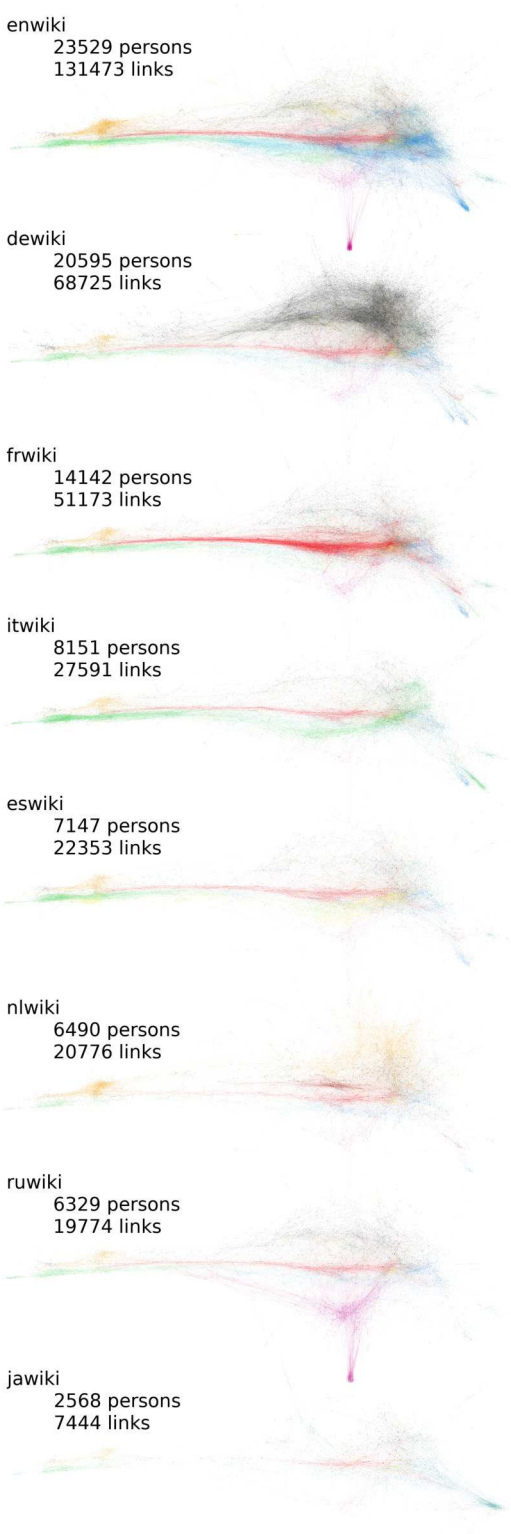


Figure 6.15: Artist network in eight Wikipedia language versions

Summary

As shown by the analysis and the visualization of the resulting network of artists, the identification of persons relevant to art history via co-occurrence-clustered occupations in Wikidata appeared as a feasible alternative to using external authorities such as the ULAN for this purpose. It nevertheless became clear that neither the ULAN nor Wikidata were able to provide a complete coverage of the domain, each having specific foci based on their individual background and the involved parties behind the many contributions they both were composed from. A basic agreement on the structure of fundamental long-term developments in art history as seen from the level of recorded interactions between historical persons from different nations, however, appeared to exist between them. As for the case for combining the contributions of multiple Wikipedia language versions, this suggested that at least at this large-scale level, the combination of person records identified via the two approaches would result in a more balanced dataset, including contributions from multiple world regions and thus taking their specific interests more into account.

Besides the different coverage of nationalities, the complementary nature of the persons discovered via the ULAN mappings and via Wikidata art occupations also became clearly visible in the featured person roles. On the one hand, the set of fine arts related occupations found via co-occurrence in Wikidata included some professions that were almost not present in the ULAN, such as comics and Manga artists. This raised general questions on the authority to decide what counts as fine art, answers to which would expectedly differ between rather institutional sources such as the ULAN and those potentially open to everyone, such as Wikipedia/Wikidata. On the other hand, the ULAN included many non-artists who were nevertheless relevant to the domain of art history by being subjects of artworks, patrons, researchers, philosophers, etc. These persons — as long as they had no "creator" role as well — were not included in the art occupation based subset of Wikidata records, although many of them were present in subsets based on other occupation clusters. Simply including typical non-artist roles in the ULAN, such as popes or kings, would make only little sense since it would add many unrelated persons as well. The exploration of more sophisticated methods of identification, such as extending the occupation-based approach with means of clustering the underlying person-to-person networks for finding related persons from different domains, would thus be another desirable topic for future work.

6.2.4 Network visualizations of other domain specific person groups

The observation that the occupation-based network of Wikidata artists featured a contiguous structure showing large-scale developments in art history, which was moreover quite similar to the one obtained via ULAN mappings, provided a strong incentive to analyze the structures of other domain specific occupation groups of persons as well. As derived from the network of co-occurring occupations shown in Figure 6.12, the occupation clusters listed in Table 6.7 broadly covered the domains fine arts, music,

performing arts, literature, humanities, clergy, media, "science, technology, engineering, mathematics" (STEM), administration and military, businesspersons and sports.

The obtained partition was quite remarkable because it broadly corresponded with the 250 year old categorization made by Joeseph Priestley for his previously mentioned "Chart of Biography" from 1765, described in [Priestley, 1765]. Considering his "Statesmen and Warriors", "Divines and Metaphysicians" ,"Mathematicians and Physicians", "Poets and Artists", "Orators and Critics" and "Historians and Antiquarians", the co-occurrence based partition appeared like a modern update to it. Since a thorough analysis of this variety of possible biography networks was beyond the scope of this work, this subsection discusses preliminary observations for a number of selected subsets which can be considered as an incentive for future work.

Theology and Clergy aka "Divines and Metaphysicians"

Figure 6.16 shows the GCC of the network of Wikipedia biographies having occupations from cluster #19 "Theology and Clergy" after discarding links exceeding a birth date difference of 75 years. Overall, the "clergy" network was found to have the highest link density of the different biography networks and had the largest fraction of "long distance" links. Moreover, it was amongst the more international networks, having almost 50% of its links between persons of different nationality. The visualization of this group of biographies featured a remarkable development which reached back to the beginning of Christianity and represented major developments especially for this world religion.

Starting with the apologist Tatian born in 120AD, a single strand of mainly Christian clergymen spanned from the early days of Christianity across the period of the Roman empire and the following Middle Ages to the beginning of the Modern era, where Martin Luther marked the beginning separation between Catholicism and the emerging Protestantism. From there on, the two well separated strands progressed further and the coloring by nationality also made their geopolitical implications dramatically visible. It was interesting to see increasing ties between the two denominations to appear at around the end of the 19^{th} century. As shown in Subfigure b), a significant proportion of biographies in the protestant strand was assigned with the role theologician, while the catholic part was dominated by priests.

The coloring by role revealed another remarkable section on top of the visualization, representing a strand of Jewish persons almost exclusively assigned with the role Rabbi. Coloring the nodes by birth date rounded to the next half century, Subfigure c) again showed how the temporal pruning of links preserved the chronology embedded in the network. Unfortunately, other world religions besides Christianity and Judaism were very underrepresented in the Figure, raising questions whether this was rooted in the approach to identify relevant persons via co-occurring occupations or on an overall bias present in Wikipedia/Wikidata in this regard.

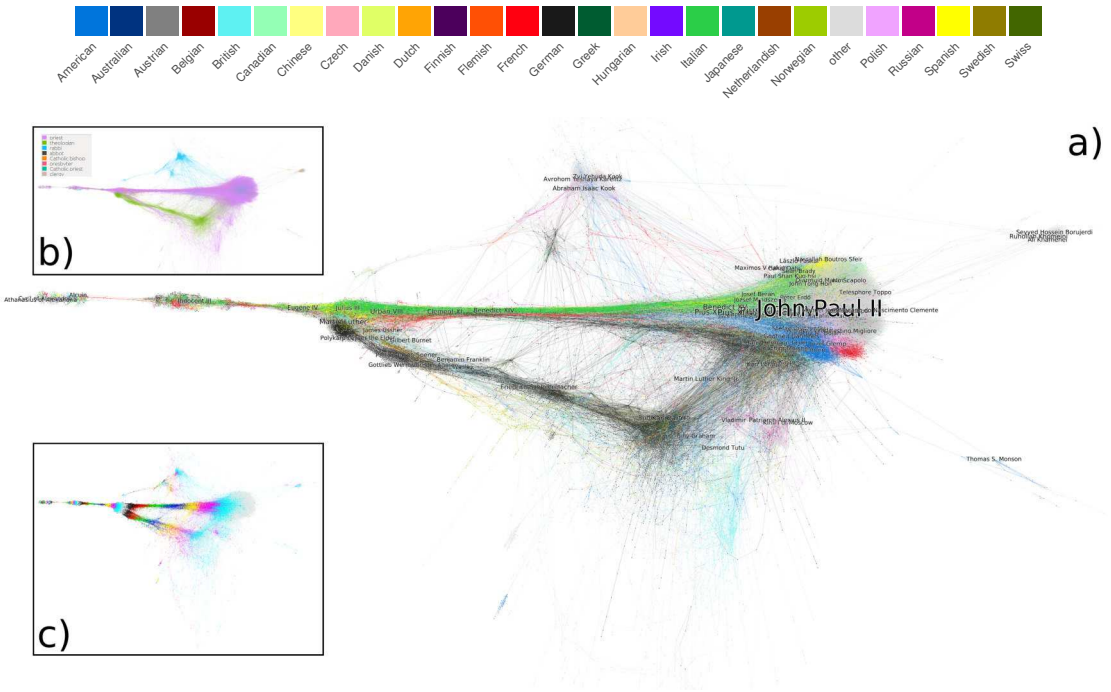


Figure 6.16: Theology and clergy network, 27,853 persons, 145,320 links (**A11 a**)

Government, Administration and Military aka "Statesmen and Warriors"

Figure 6.17 shows the GCC of the 75 year pruned network of biographies having occupations from clusters #5 and #15, "government, administration and military", representing "Statesmen and Warriors" in Priestly's terms. The full network had the highest fraction of "same nationality" links and 97.5% of its links connected persons born not more than 75 years apart. It was remarkable to see a single strand of emperors, kings, administrative and military leaders to stretch from early Byzantine emperors across Charlemange and the developing European monarchies up to the time of the French revolution, where the scenery literally exploded into a multitude of well-separated national strands progressing towards the present day. Although dominated by Europeans, the "pre-revolutionary" single strand also contained important figures from other parts of the world and it was remarkable to find a group of Japanese emperors and members of court to be at its very beginning on the left side of the Figure.

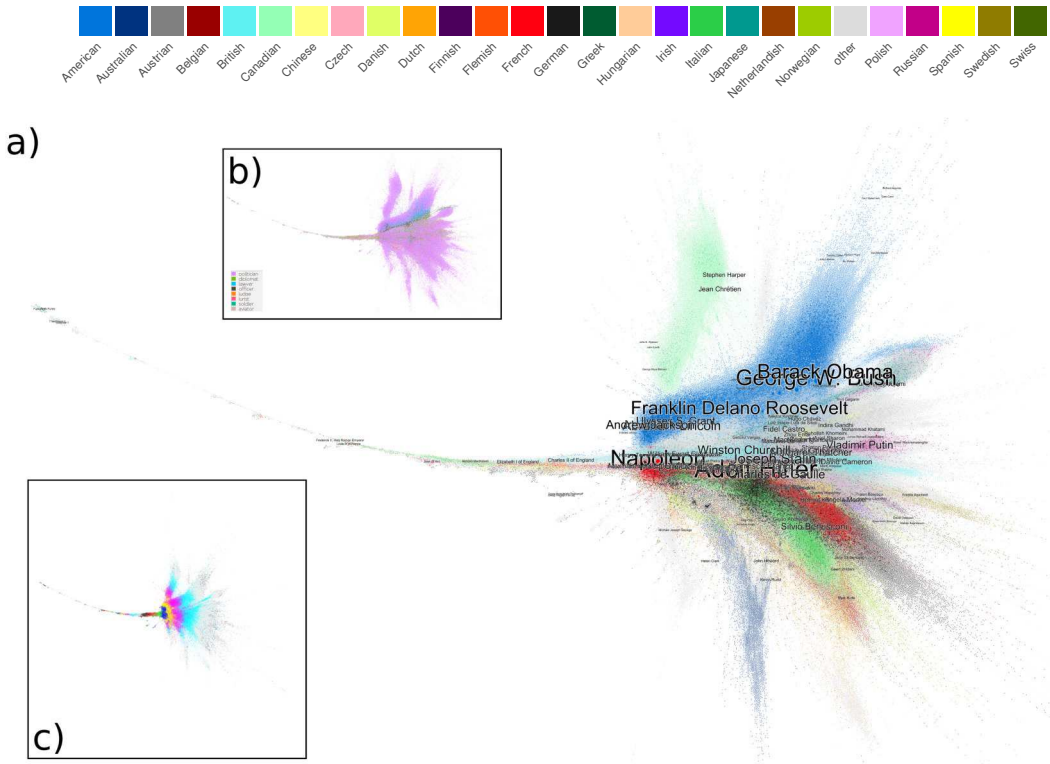


Figure 6.17: Government network, 147,150 persons, 761,864 links (**A11 b**)

Philosophy and STEM aka "Mathematicians and Physicians"

Figure 6.18 shows the GCC of the pruned network of biographies having occupations from cluster #24 "Philosophy and STEM". It started on the left side with two strands of non-Western philopsophers and mathematicians, one group connecting persons born before and around 1000AD from Iran, Arabia and China and another one with Korean and Japanese persons from after 1400AD right below. The upper strand later joined with a group of European Middle-Age scholars around Thomas Aquinas, from where a sparse single strand continued across the age of the Renaissance, featuring artist-engineers such as Leonardo da Vinci who marked the beginning of an increasing separation into distinct disciplines. Joined by the Korean/Japanese strand at around the late 17$^{\text{th}}$ century, the group around the polymath Isaac Netwon appeared as point of bifurcation, from where on two separate strands evolved further towards today, a strand of philosophers, social scientists and economists on top, another one consisting of engineers, astronomers and mathematicians at the bottom which soon also developed a separate strand of physicists. The Philosophy and STEM network was remarkable because each of the separating strands consisted of persons of many different nationalities, which was also reflected in its lowest fraction of same-nationality links amongst the networks (46.3%).

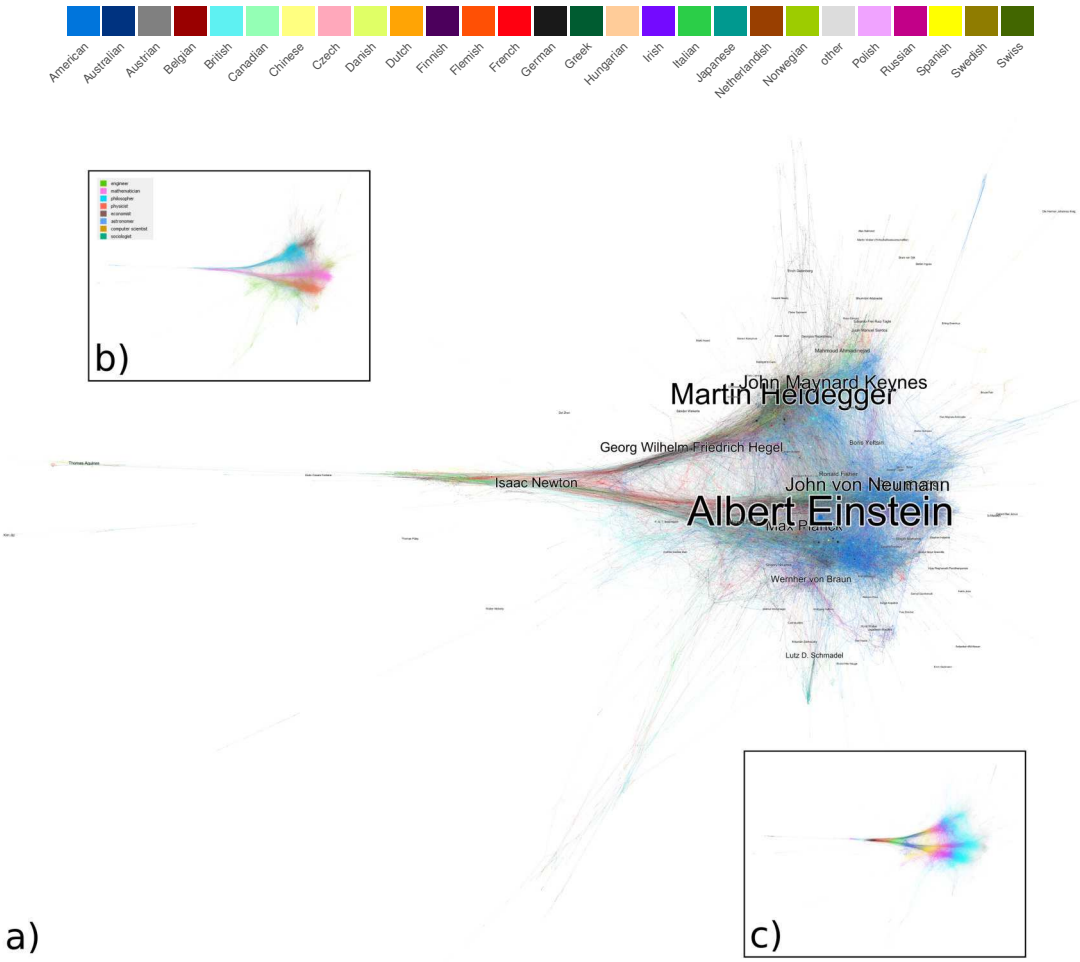


Figure 6.18: STEM network, 30,577 persons, 144,166 links (**A11 c**)

6.2.5 Summary

Looking for an alternative to using the ULAN to identify Wikidata records about persons relevant to art history, the clustering of occupations assigned to Wikidata person records based on the analysis and filtering of their co-occurrence network proved to be a useful approach in this regard. Representing method **M8**, it resulted in remarkably clear-cut clusters of domain-specific occupations, which represented an interesting design artifact (**A9**, Figure 6.12) in its own regard and could be used to derive relevant subsets of persons. It was interesting to find the basic domain categorization which resulted from the clustered occupations to correspond quite well with the person groups defined by Priestley 250 years earlier for his "Chart of Biography". As far as the occupation-based person set of artists was concerned, only about 40% of the persons featured there were also present in the ULAN, which suggested that Wikidata could serve as complementary

data source for relevant person records. It was interesting to see that high overlap existed for persons from "classical" periods in art history, while especially more recent artists had less presence in the ULAN. Looking at persons which were highly connected in Wikipedia and tagged as fine artists in the ULAN but not identified as such via the clustered artist occupations revealed that many of them actually belonged to other domains such as literature and film. As mentioned in the general discussion of the ULAN in Chapter 4, the reasons for such persons to be nevertheless classified as artists most likely were related to specific museum collection items which they had created. While it was in principle desireable to separate such "secondary" artists from "primary" ones, it again pointed out the fact that many other persons who had high relevance for art history, being authors, art historians, critics, etc., could not be identified via artist occupations.

The analysis of the network of Wikipedia hyperlinks which were found for about 73.5% of the persons identified via artist occupations revealed that despite of the only partial overlap, it had a comparable topology to the one obtained for the ULAN mapped persons. Its visualization, representing design artifact **A10** (Figure 6.14), therefore had a quite similar appearance as the one created for the ULAN-mapped Wikipedia biographies, although it featured specific groups of persons as more detached from the main body of the network, such as Soviet painters or Japanese Manga artists. This again underlined the value of using Wikidata as additional data source which appeared to provide a broader view on art history than present in the ULAN. The comparison of the occupation-based networks in different language versions again revealed their individual self-focus-bias, which appeared to be even stronger due to the omission of the many important non-artists that were present in the ULAN-mapped counterparts. This nevertheless again highlighted the value of combining individual language contributions into a more global "whole".

The generalizability of selecting domain-specific biographies with method **M8** could be demonstrated via a selection of visualizations of networks of biographies identified via other occupation clusters. Together representing artifact collection **A11** (Figures 6.16, 6.17 and 6.18), they showed that the clustering of co-occurring occupations yielded meaningful domain-specific subsets of persons whose biographical Wikipedia networks, especially after the application of the temporal link filtering proposed in method **M3**, equally unfolded into chronological representations which, although featuring comparable temporal successions based on the distribution of person birth dates in each subset, revealed very different patterns of interaction between different nations and roles across the ages. It was interesting to find each of the different visualizations to express mayor developments in their respective domains, such as the separation of Catholicism and Protestantism after the reformation and their apparent rapprochement starting around the beginning of the 20th century which became visible in the Theology and Clergy network, the explosion of different national strands of governance and power which appeared to correlate with the onset of nation states throughout the 19th century in the network related to Government, Administration and Military, or the quite international networks of different scientific disciplines which appeared to emerge from a common root in the "Philosophy and STEM" network. The thorough analysis of these different representations was, however, beyond the scope of this work.

CHAPTER 7

Conclusion

The aim of this thesis was to explore the potential of contextual data sources in the problem domain of Digital Cultural Heritage, with a particular focus on art history. This on the one hand included considerations regarding the added value of integrating collections of digitized artworks and related metadata with contextual data about social relationships between art historical actors, on the other hand focused on possible narratives embedded in the latter and how they compared across datasets of different origin.

7.1 Addressing the research questions

This Section relates the work presented in the previous chapters to the three research questions RQ1-3 which were proposed in Section 1.2.1 and are listed again below.

- **Research question 1 (RQ1)** What potential exists in integrating object level metadata with related historical social network data for exploratory search based virtual presentations of art historical content?
- **Research question 2 (RQ2)** What kind of "large scale" historical narratives are embedded in the structure of historical network data related to art history?
- **Research question 3 (RQ3)** How do art historical social networks derived from institutional and/or crowd-sourced data collections differ and what do they have in common?

Integration of social network data with metadata about artworks

RQ1 was addressed via design artifact **A1** (Figure 3.6), the 3D Information Landscape prototype combining artwork data from the Web Gallery of Art (WGA) with data from the Union List of Artist Names (ULAN), presented in Chapter 3. It was accompanied

by method **M1** to semantically integrate and visualize these complementary datasets. Driven by continuous user assessments, the iterative design process resulted in the addition of novel features which drew from this specific combination of data sources, most importantly the explicit inclusion of non-artists and their portaits, which provided additional context for exploratory search. Almost as a side-effect, it prompted for a deeper analysis of the network of portrait relationships, design artifact **A2** (Figure 3.8), and its comparison with existing ULAN person links, which supported the identification of missing facts in the ULAN.

The main contributions of activities in the context of **RQ1** were thus two-fold. On the one hand, the semantic integration of the two used datasets using Semantic Web technologies and their subsequent presentation in a Web-based remote virtual 3D setting served as a technical proof-of-concept that such data-driven environments could be dynamically created using existing technology. Moreover, the performed user studies showed that such environments were of interest and served as incentives for exploration, although they also pointed out a number of issues regarding user interaction. On the other hand, the second main contribution were the insights won from the integration of the two datasets on data level, such as that the ULAN, besides serving as contextual social network resource, enabled the extraction of portrait relationships by allowing the identification of relevant person names in artwork titles. It was interesting to see that the extracted relationships could then be used to point out gaps in the ULAN in turn.

Analysis of the ULAN data

The analysis of the ULAN data, which was discussed in Chapter 4, mainly addressed **RQ2** and revealed a large-scale art historical narrative embedded in the ULAN network. The related activities provided interesting insight into different distributions of persons in linked and unlinked subsets of ULAN records, revealing different levels of "attention" which favoured European male artists from countries belonging to the well-known schools of art history, but also showed that clear effort had been put into representing persons from the art historical "periphery" as well. ULAN person attributes such as nationality, gender and birth and additional attributes about relationship types provided valuable information for the quantitative analysis of the ULAN network, revealing different structural aspects such as assortativity, density and connected components which in part strongly differed for different link types. The visualization of the giant connected component, representing design artifact **A3** (Figure 4.24), revealed a network topology featuring a succession of well-defined national clusters that correlated with major epochs in art history and represented a narrative that corresponded to similar, hand-drawn visualizations made by art historians.

The main contribution of the ULAN analysis was to show that sufficient amounts of social network data about interlinked artists and other important persons in art history formed a network structure whose analysis and visualization provided insight into large-scale developments across multiple centuries. The results motivated to pursue **RQ3** by identifying alternative data sources for comparison.

Aligning the ULAN with Wikipedia

The identification of multilingual Wikipedia content via the ULAN and its integration with ULAN data using framework **M2** represented the first approach to **RQ3**. As discussed in Chapter 5, the data framework enabled a thorough analysis of the Wikipedia hyperlink network of art historical actors using person attributes from the ULAN, revealing that for persons present in both datasets, link information found in the ULAN not only tended to be also represented in Wikipedia to a high degree, but was significantly exceeded by the present hyperlinks as well. A number of structural characteristics and the overall topology of the network with its successing national clusters corresponded as well between the two sources, although the dedicated temporal filtering method **M3** had to be applied in order to be able to create design artifact **A4** (Figure 5.19), the chronological visualization of the Wikipedia network as as juxtaposition to the ULAN version in design artifact **A3**.

The visual representation of the Wikipedia network revealed interesting insight into the structure of biographical relationships in the free encyclopedia and demonstrated that it was able to convey different cultural narratives, such as the parallel development and later fusion of Western and non-Western cultures, which further contributed to **RQ2**. This mode of visualization could also be used in design artifact **A5** (Figure 5.25) juxtaposing individual Wikipedia language contributions with each other, which effectively revealed their individual cultural preferences, thus simultaneously also contributing to **RQ3**. Another contribution to **RQ2** was the analysis and visualization of large-scale interactions on nation level by aggregating the biographical Wikipedia hyperlinks by nationality attributes. Method **M4** was applied to filter the resulting network in order to provide a relative measure to quantify the extent of interactions across nations of different size, which revealed that person connections on Wikipedia in general followed many historical and cultural patterns that had also appeared elsewhere, visualized in design artifact **A6** (Figure 5.27).

The main contributions of aligning the ULAN with corresponding Wikipedia data were the alignment and comparison framework itself and the rigorous overlap analysis, the approach to use ULAN attributes for analyzing the Wikipedia network, which also led to the chronological filtering approach to unravel the chronology embedded in the Wikipedia hyperlink network, the resulting views on its topology and the derived views on the interactions on nation level, achieved after applying the filtering method **M4**.

Linking biographies with art and architecture styles using Wikidata

Further pursuing **RQ2** and **RQ3**, the inclusion of information about art and architecture styles as separate entities into an extended network of art history information on Wikipedia, presented in Section 6.1, resulted in the extended framework **M5**. It allowed to leave behind the ULAN as means to identify and describe art historically relevant persons in Wikipedia and focused on the Wikidata repository for this purpose instead. Since the latter did not offer immediate means to select domain relevant persons, the

identification via links to art and architecture styles appeared as a feasible approach in this regard.

Although the analysis of person roles in the resulting person dataset showed that many fine artists were successfully identified this way, the significant presence of persons belonging to other art disciplines suggested to consider alternative approaches for identification as well. It nevertheless provided interesting insight on the relationships between art styles and person roles, revealing a number of styles with strong interdisciplinary relevance. After chronological filtering using extended filtering method **M6**, the resulting bi-partite person-style network could be visualized as design artifact **A7** (Figure 6.8). This revealed that the person-style network featured a similar chronological structure as it was the case for the person-only networks, which related the different art and architecture styles in a seamingly correct manner and additionally contextualized them via nationality information of persons they were linked to. Despite not containing only fine artists, it appeared as map of art styles which could potentially be used for the contextualization of artworks, which was further supported by the observation that it successfully extended the content of existing, hand-made diagrams of interrelated styles. The projection of the person-style network and its subsequent statistical filtering via method **M7** resulted in design artifact **A8** (Figure 6.10) which provided a more direct view on the interactions between the featured styles. This resulted in a more compact representation which especially emphasized the separation between art and architecture.

The main contributions of the related activities to **RQ2** were to show that the filtered person-style networks were able to convey a historical narrative of stylistic developments by interrelating different styles via interlinked persons in a chronological manner, which allowed to gain insight into key developments such as the changing interrelationship between art and architecture. As far as **RQ3** was concerned, it provided an initial attempt to leave behind the ULAN as means to identify relevant persons, which, however, left space for improvement.

A generalized approach to select domain-specific persons

As alternative approach to selecting art historically relevant persons via art and architecture styles, further focus was therefore put on the person information made available via Wikidata. As discussed in Section 6.2, occupation information was found to be of special interest in this regard and the analysis of its co-occurring assignments to person records using method **M8** returned a remarkable landscape of interrelated occupations which became visible as network featuring well-expressed domain-specific clusterings. While the visualization of this network represented design artifact **A9** (Figure 6.12), its clustering was used to select various domain-specific subsets of Wikidata person records.

The fine art person dataset identified via related occupations featured many persons which were also present in the ULAN, but also contained many without a biographical record there, which especially included more recent born individuals and representatives from the periphery of art history. Another interesting insight was that many ULAN persons which were missing in the occupation-based subset were found to be important

historical actors who only could be considered "secondary" artists, whose identification became enabled this way. The visualization of the chronologically filtered artist network as design artifact **A10** (Figure 6.14) again revealed a historical narrative represented mainly via mutually interacting national clusters, quite similar to the ones already made visible in artifacts **A3** and **A4** but clearly giving more space to person groups less represented there. The extraction and visualization of additional domain-specific person networks in design artifact **A11** (Figures 6.16, 6.17 and 6.18) served as proof-of-concept for the generalizability of method **M8**.

The main contributions of these activities were thus again directed at **RQ2** and **RQ3**, the former addressed via the analysis and visualization of the large-scale narratives embedded in the fine arts via artifact **A10** and also the other domains' person networks via artifact and **A11**, which served as point of reference to distinguish the art-historical from other narratives, the latter via the alternative fine art person dataset which could be identified via method **M8** and artifact **A9**.

7.2 Final discussion & outlook

Initiated by the idea to use social network data about artists and other art historical actors as structural element for the creation of virtual museum architecture for the presentation of artworks in their art-historical context, the focus of this work was soon directed towards much broader questions regarding the nature of the used datasets and what kind of views are embedded therein. The analysis of the ULAN data and especially the distribution of person nationalities in the dataset, particularly in the different subsets which constituted the network of persons, soon revealed that much-contested aspects of "traditional" art history, such as the individual attempts to highlight the importance of emerging nation-states through the greatness of their artistic achievements, discussed for example by Locher in [Locher, 2014], were clearly also embedded in the data. The total view on the giant connected component of the ULAN network underscored this notion even more, where the different national clusters appeared to be quite self-contained and featured only a relatively small proportion of international connections, which nevertheless still resulted in a view that suggested a succession of different nations as centers of art historical importance across the centuries.

Given the intitial motivation to use the ULAN data as substrate for a virtual museum, it was striking to find this view to highly correspond with concepts that had from early on been fundamental to the idea of national art museums which, such as mentioned by Duncan and Wallach in [Duncan and Wallach, 1980] as one important intention for the establishment of the Louvre after the French revolution, had sought to stage the artistic output of the respective nation to represent the culmination of all previous developments which were usually presented in the aggregate form of "national schools". As shown by Barr's Torpedo Diagrams, such notions continued well into the 20th century and it was particularly remarkable to find the visual representations of the Wikipedia biography networks, whose analysis had been motivated by the findings for the ULAN data which sparked interest in additional data sources for comparison, to correspond so well with

Barr's torpedoes, both with respect to their shape as well as to their subdivison into distinct "national" regions. Considering the individual views present in the different language editions of the free encyclopedia revealed their individual self-focus-biases to manifest themselves in the subdivisions of their individual "torpedoes" which tended to grant more space to the cultures/nations that were related to each respective language. While the individual biographical Wikipedia views on art history thus appeared to continue the tradition to focusing on one's own culture, it also became evident that the Wikipedia hyperlink networks of ULAN persons were nevertheless more international than the ULAN associative relationships, although mainly due to still quite self-focused links present in the different language versions which connected persons from language-related cultures/nations with persons from others. The combination of the different networks into a multi-language version, however, resulted in a quite rich and more balanced international linkage which also enabled the analysis and visualization of the aggregated nationality network which revealed interesting patterns that appeared to be of quite global validity. This observation clearly supported the position that open access to a variety of individual points of view could result in a combination which represented more than just the sum of its parts. The quantitate analysis of the Wikipedia hyperlinks of course only represented a high-level showcase which did not take the details behind the links into account, not considering any sentiment or any other semantics in this regard.

The extension of the biography network to a bi-partite version connecting persons and styles, which had in part been motivated by the question to find alternative ways to identify relevant persons in Wikipedia than using the ULAN, provided a broader view on art history content on Wikipedia and also put the focus away from the immediate "national" aspects that were so prominent in the person clusterings of the biographical network. It nevertheless became clear that self-focus bias also existed on this level, with a number of articles about specific art styles only present in language versions which corresponded to related cultures for which the style had played a role. Having a network of links between styles and biographies enabled to further quantify this aspect by analyzing the distribution of person nationalities linked to each style. The individual contributions by the different language versions again resulted in a more global, broad view on art history which included many local styles and also very recent disciplines such as Internet Art. Considering the critique of Barr's "Diagram of Stylistic Evolution from 1890 until 1935" to be selective, which was mentioned by Schmidt-Burkhardt in [Schmidt-Burkhardt, 2005], its contemporary Wikipedia update appeared as response to such criticism, compiling different perspectives into a more diverse whole.

Using Wikidata to retrieve information about art and architecture styles and metadata about persons linked to them in Wikipedia showed the potential of this still emerging, global open data repository to serve as hub for accessing and combining different, but related datasets. As shown by the landscape of co-occurring person roles which resulted from the further analysis of person records there, the platform appeared to offer an extensive data environment for experimentation. It was remarkable to find the clustering of occupations, which was based on a much larger fraction of more recent persons, to resemble historical categorizations to some extent which were conceived a quarter of a

Millenium earlier, suggesting that many basic categories of human activity apparently had not changed over the centuries. Although "only scratching the very surface", the intial analysis of the additional domain-specific person networks, performed after the network of persons selected via art occupations had appeared as a feasible alternative large-scale view on art history, provided plenty of opportunities for future research endeavors.

From this point of view, next steps could include a number of relevant activities. A more refined approach to selecting domain-specific person subsets could for example also take closely related persons from other domains into account and identify their relevance via network characteristics, which would in the art history case add relevant persons with occupations assigned to other domain clusters, such as art historians, -critics, -dealers and -collectors, to the set of persons identified via art occupations. Generally contributing to another important improvement would be to find more high-level descriptions of the encountered networks, taking the encountered person attributes such as nationality, role, gender and birth/death into account. The "variable-granularity" modularity approach presented by Kitromilidis and Evans in [Kitromilidis and Evans, 2018] could serve as useful starting point in this regard. While potentially contributing to the comparability of similar, but not identical, networks such as it was the case for those based on ULAN associative relationships and their Wikipedia hyperlink counterparts, it could moreover be used to visually aggregate individual biographies into clusters based on temporal and other features, potentially resulting in further data-driven updates of historical attempts to represent historical development, such as Friedrich Strass' "Strom der Zeiten" (Stream of time), discussed in Subsection 2.1.2 and shown again in Figure 7.1.

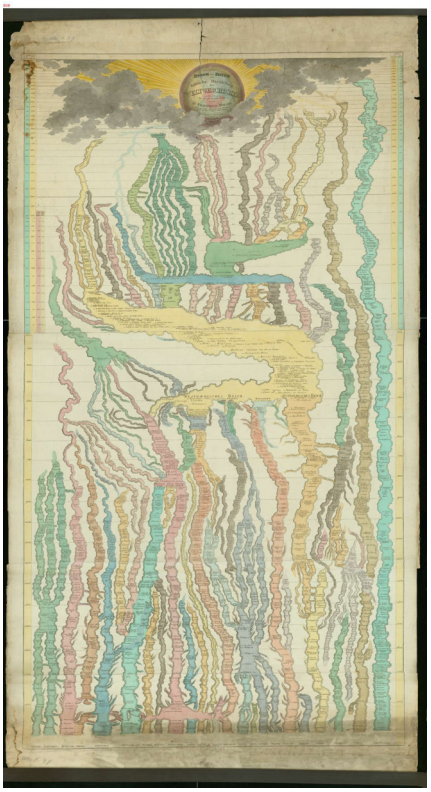


Figure 7.1: Strom der Zeiten, Friedrich Strass, 1828, urn:nbn:de:bvb:12-bsb00003029-3

List of Figures

List of Tables

APPENDIX A

Tables

A.1 Artist rankings

rank	name	nation	count
1	Picasso, Pablo	Spanish	757
2	Dürer, Albrecht	German	616
3	Rubens, Peter Paul	Flemish	600
4	Michelangelo Buonarroti	Italian	537
5	Leonardo da Vinci	Italian	526
6	Raphael	Italian	460
7	Rembrandt van Rijn	Dutch	442
8	Titian	Italian	418
9	Goya, Francisco de	Spanish	391
10	Palladio, Andrea	Italian	377
11	Gogh, Vincent van	Dutch	283
12	Turner, J M. W.	British	270
13	Cézanne, Paul	French	267
14	Brunelleschi, Filippo	Italian	243
15	Klee, Paul	Swiss	240
16	Matisse, Henri	French	224
17	Bernini, Gian Lorenzo	Italian	220
18	Alberti, Leon Battista	Italian	218
19	Courbet, Gustave	French	215
20	Le Corbusier	Swiss	214
21	Schinkel, Karl Friedrich	German	213
22	Caravaggio, Michelangelo Merisi da	Italian	211
23	Delacroix, Eugène	French	209
24	Kandinsky, Vasily.	Russian	205
25	Giotto	Italian	203
26	Poussin, Nicolas	French	202
27	Ingres, Jean-Auguste-Dominique	French	201
28	Manet, Edouard	French	200
29	Blake, William	British	199
30	Giorgione	Italian	191
31	Velázquez, Diego	Spanish	181
32	Piranesi, Giovanni Battista	Italian	180
33	Friedrich, Caspar David	German	177
34	Bosch, Hieronymus	Netherlandish	163
35	Rodin, Auguste	French	162
36	Duchamp, Marcel	French	161
37	Bruegel, Pieter	Flemish	160
38	Ernst, Max	German	159
39	Degas, Edgar	French	156
40	Munch, Edvard	Norwegian	155
41	Cranach, Lucas	German	154
42	Watteau, Antoine	French	152
43	Monet, Claude	French	151
43	Viollet-le-Duc, Eugène-Emmanuel	French	151
45	Gauguin, Paul	French	147
46	Donatello	Italian	145
46	Eyck, Jan van	Netherlandish	145
48	Tiepolo, Giovanni Battista	Italian	144
49	Lotto, Lorenzo	Italian	141
49	Constable, John	British	141

Table A.1: James Elkin's artist ranking

#	name	nation	count	#	name	nation	count
1	Dürer, Albrecht	German	56	70	Toulouse-Lautrec, Henri de	French	8
2	Kandinsky, Vasily.	Russian	49	70	Semper, Gottfried	German	8
3	Goya, Francisco de	Spanish	46	70	Alberti, Leon Battista	Italian	8
4	Raphael	Italian	45	70	Rodriguez, Alonso	Italian	8
4	Picasso, Pablo	Spanish	45	70	Witkiewicz, Stanisław	Polish	8
6	Michelangelo Buonarroti	Italian	41	80	Vasari, Giorgio	Italian	7
7	Rembrandt van Rijn	Dutch	36	80	Piranesi, Giovanni Battista	Italian	7
8	Le Corbusier	Swiss	33	80	Asam, Cosmas Damian	German	7
9	Rubens, Peter Paul	Flemish	31	80	Flandrin, Hippolyte	French	7
9	Matisse, Henri	French	31	80	Lorrain, Claude	French	7
11	Ingres, Jean-Auguste-Dominique	French	29	80	Tintoretto	Italian	7
11	Delacroix, Eugène	French	29	80	Viollet-le-Duc, Eugène-Emmanuel	French	7
11	Degas, Edgar	French	29	80	Brunelleschi, Filippo	Italian	7
14	Rodin, Auguste	French	28	80	Weyden, Rogier van der	Flemish	7
14	Schinkel, Karl Friedrich	German	28	80	Loos, Adolf	Austrian	7
16	Gogh, Vincent van	Dutch	26	80	Schiele, Egon	Austrian	7
16	Leonardo da Vinci	Italian	26	80	Menzel, Adolph	German	7
16	Turner, J M. W.	British	26	80	Constable, John	British	7
19	Manet, Edouard	French	23	80	Marc, Franz	German	7
20	Monet, Claude	French	22	80	Maulbertsch, Franz Anton	Austrian	7
21	Atget, Eugène	French	21	80	Lorenzetti, Pietro	Italian	7
22	Klee, Paul	Swiss	20	96	Sarto, Andrea del	Italian	6
23	Velázquez, Diego	Spanish	19	96	Rossetti, Dante Gabriel	British	6
24	Cézanne, Paul	French	18	96	Carracci, Agostino	Italian	6
25	Kokoschka, Oskar	Austrian	17	96	Giulio Romano	Italian	6
26	Wright, Frank Lloyd	American	16	96	Ernst, Max	German	6
26	Boucher, François	French	16	96	Schwitters, Kurt	German	6
28	Friedrich, Caspar David	German	15	96	Martini, Simone	Italian	6
28	Giotto	Italian	15	96	De Chirico, Giorgio	Italian	6
28	Richter, Gerhard	German	15	96	Sargent, John Singer	American	6
31	Gropius, Walter	American	14	96	Giordano, Luca	Italian	6
31	Bernini, Gian Lorenzo	Italian	14	96	Cano, Alonso	Spanish	6
31	Mantegna, Andrea	Italian	14	96	Fischer von Erlach, Johann Bernhard	Austrian	6
31	Gauguin, Paul	French	14	96	Larionov, Mikhail	Russian	6
31	Klimt, Gustav	Austrian	14	96	Runge, Philipp Otto	German	6
31	Stoss, Veit	German	14	96	Tatlin, Vladimir	Ukrainian	6
37	Brancusi, Constantin	Romanian	13	96	Stethaimer, Hans, the younger	German	6
37	Poussin, Nicolas	French	13	96	Macke, August	German	6
37	Watteau, Antoine	French	13	96	Daumier, Honoré	French	6
37	Blake, William	British	13	96	Boudolf, Jan	Netherlandish	6
37	Bosch, Hieronymus	Netherlandish	13	115	Cortona, Pietro da	Italian	5
42	Duchamp, Marcel	French	12	115	Moreau, Gustave	French	5
42	Giaquinto, Corrado	Italian	12	115	Bellini, Giovanni	Italian	5
44	Mies van der Rohe, Ludwig	German	11	115	Thorvaldsen, Bertel	Danish	5
44	Holbein, Hans	German	11	115	Pissarro, Camille	French	5
44	Giacometti, Alberto	Swiss	11	115	Veronese, Paolo	Italian	5
44	Miró, Joan	Spanish	11	115	Goltzius, Hendrick	Dutch	5
44	Carracci, Annibale	Italian	11	115	Corinth, Lovis	German	5
44	Dalí, Salvador	Spanish	11	115	Schlemmer, Oskar	German	5
44	Beckmann, Max	German	11	115	Maillol, Aristide	French	5
44	Cranach, Lucas	German	11	115	Man Ray	American	5
44	Munch, Edvard	Norwegian	11	115	Burgkmair, Hans	German	5
44	Bartholdi, Frédéric-Auguste	French	11	115	Caravaggio, Michelangelo Merisi da	Italian	5
54	David, Jacques-Louis	French	10	115	Liebermann, Max	German	5
54	Titian	Italian	10	115	Favorsky, Vladimir	Russian	5
54	Donatello	Italian	10	115	Metsys, Quentin	Netherlandish	5
54	Grosz, George	German	10	115	Dix, Otto	German	5
54	Redon, Odilon	French	10	115	Murillo, Bartolomé Esteban	Spanish	5
54	Tonitza, Nicolae	Romanian	10	115	Dufy, Raoul	French	5
54	Chodowiecki, Daniel Nikolaus	German	10	115	Meidner, Ludwig	German	5
54	Eyck, Jan van	Netherlandish	10	115	Vauchelet, Théophile Auguste	French	5
62	Carracci, Lodovico	Italian	9	115	Kirchner, Ernst Ludwig	German	5
62	Malevich, Kazimir	Russian	9	115	Taut, Bruno	German	5
62	Thoma, Hans	German	9	115	Goncharova, Natal'ya	Russian	5
62	Piero	Italian	9	115	Koch, Joseph Anton	Austrian	5
62	Chagall, Marc	Belorussian	9	115	Modersohn-Becker, Paula	German	5
62	Courbet, Gustave	French	9	115	Kollwitz, Käthe	German	5
62	Mondrian, Piet	Dutch	9	115	Terragni, Giuseppe	Italian	5
62	Bonnard, Pierre	French	9	115	Machuca, Pedro	Spanish	5
70	Greco, El	Greek	8	115	Vermeer, Johannes	Dutch	5
70	Hoffmann, Josef	Austrian	8	115	Barlacchi, Tommaso	Italian	5
70	Arp, Hans.	French	8	115	Lorca, Federico García	Spanish	5
70	Palladio, Andrea	Italian	8	115	Poiret, Paul	French	5
70	Magritte, René	Belgian	8				

Table A.2: Heinrich Dilly's artist ranking

Table 1.1 from [Jensen, 2007]

#	name	nation	count
1	Goya, Francisco de	Spanish	157
2	Monet, Claude	French	152
3	Manet, Edouard	French	144
4	Cézanne, Paul	French	142
5	Gogh, Vincent van	Dutch	130
6	Delacroix, Eugène	French	124
7	David, Jacques-Louis	French	122
8	Degas, Edgar	French	114
9	Ingres, Jean-Auguste-Dominique	French	106
10	Gauguin, Paul	French	105
11	Courbet, Gustave	French	102
12	Rodin, Auguste	French	100
13	Daumier, Honoré	French	88
14	Turner, J M. W.	British	83
15	Renoir, Auguste	French	78
16	Gericault, Théodore	French	77
17	Seurat, Georges	French	73
18	Constable, John	British	69
19	Friedrich, Caspar David	German	63
20	Munch, Edvard	Norwegian	59
21	Whistler, James McNeill	American	52
21	Corot, Camille	French	52
23	Toulouse-Lautrec, Henri de	French	51
24	Millet, Jean-François	French	49
25	Cassatt, Mary	American	48
26	Canova, Antonio	Italian	46
27	Ensor, James	Belgian	40
28	Blake, William	British	39
29	Pissarro, Camille	French	38
30	Menzel, Adolph	German	37
31	Gros, Antoine-Jean, Baron	French	28
32	Moreau, Gustave	French	27
33	Girodet, Anne-Louis	French	25
34	Rude, François	French	24
35	Rousseau, Henri	French	23
36	Morisot, Berthe	French	22
37	Runge, Philipp Otto	German	21
38	Puvis de Chavannes, Pierre	French	20
38	Caillebotte, Gustave	French	20
40	Millais, Sir John Everett	British	19
41	Rossetti, Dante Gabriel	British	18
42	Bonheur, Rosa	French	17
42	Sisley, Alfred	French	17
44	Gérôme, Jean-Léon	French	16
45	Böcklin, Arnold	Swiss	15
46	Brown, Ford Madox	British	13
46	Sargent, John Singer	American	13
48	Couture, Thomas	French	11
49	Bouguereau, William	French	10
49	Meissonier, Ernest	French	10

Table 1.2a from [Jensen, 2007]

#	name	nation	count
1	Gogh, Vincent van	Dutch	383
2	Goya, Francisco de	Spanish	261
3	Cézanne, Paul	French	231
4	Gauguin, Paul	French	221
5	Toulouse-Lautrec, Henri de	French	175
6	Degas, Edgar	French	159
7	Turner, J M. W.	British	155
8	Munch, Edvard	Norwegian	154
9	Rodin, Auguste	French	151
10	Delacroix, Eugène	French	147
11	Renoir, Auguste	French	142
12	Monet, Claude	French	131
13	Daumier, Honoré	French	130
14	Manet, Edouard	French	127
15	Ingres, Jean-Auguste-Dominique	French	108
15	Ensor, James	Belgian	108
17	Courbet, Gustave	French	107
18	Blake, William	British	102
19	Whistler, James McNeill	American	99
20	Corot, Camille	French	94
21	Constable, John	British	79
22	Canova, Antonio	Italian	76
23	Seurat, Georges	French	62
24	David, Jacques-Louis	French	59
25	Pissarro, Camille	French	54
26	Gericault, Théodore	French	53
27	Millet, Jean-François	French	51
28	Menzel, Adolph	German	49
29	Rossetti, Dante Gabriel	British	48
30	Friedrich, Caspar David	German	40
31	Moreau, Gustave	French	33
31	Böcklin, Arnold	Swiss	33
33	Morisot, Berthe	French	32
34	Sargent, John Singer	American	31
34	Millais, Sir John Everett	British	31
36	Cassatt, Mary	American	30
37	Sisley, Alfred	French	25
38	Runge, Philipp Otto	German	24
39	Puvis de Chavannes, Pierre	French	22
40	Hunt, Harry Millson	British	18
41	Gérôme, Jean-Léon	French	17
41	Rousseau, Henri	French	17
43	Meissonier, Ernest	French	16
44	Delaroche, Paul	French	13
45	Girodet, Anne-Louis	French	10
46	Bouguereau, William	French	9
46	Bonheur, Rosa	French	9
46	Caillebotte, Gustave	French	9
49	Rude, François	French	7
50	Couture, Thomas	French	5

Table A.3: Robert Jensen's artist rankings

name	nation	birth	deg	comp
Gérôme, Jean-Léon	French	1824	156	gcc
David, Jacques-Louis	French	1748	148	gcc
Bouguereau, William	French	1825	118	gcc
Delaroche, Paul	French	1797	94	gcc
Ingres, Jean-Auguste-Dominique	French	1780	82	gcc
Couture, Thomas	French	1815	82	gcc
Gros, Antoine-Jean, Baron	French	1771	80	gcc
Le Corbusier	Swiss	1887	64	gcc
Wright, Frank Lloyd	American	1867	58	gcc
Rembrandt van Rijn	Dutch	1606	54	gcc
Whistler, James McNeill	American	1834	48	gcc
Rubens, Peter Paul	Flemish	1577	46	gcc
Titian	Italian	1483	46	gcc
Cortona, Pietro da	Italian	1596	44	gcc
Michelangelo Buonarroti	Italian	1475	42	gcc
Boucher, François	French	1703	38	gcc
Rodin, Auguste	French	1840	36	gcc
Matisse, Henri	French	1869	36	gcc
Moreau, Gustave	French	1826	36	gcc
Raphael	Italian	1483	34	gcc
Picasso, Pablo	Spanish	1881	32	gcc
Gropius, Walter	American	1883	32	gcc
Puvis de Chavannes, Pierre	French	1824	30	gcc
Vasari, Giorgio	Italian	1511	30	gcc
Girodet, Anne-Louis	French	1767	30	gcc
Sarto, Andrea del	Italian	1486	28	gcc
Mies van der Rohe, Ludwig	German	1886	26	gcc
Bellini, Giovanni	Italian	1431	26	gcc
Bernini, Gian Lorenzo	Italian	1598	26	gcc
Corot, Camille	French	1796	26	gcc
Delacroix, Eugène	French	1798	26	gcc
Klee, Paul	Swiss	1879	24	gcc
Tiepolo, Giovanni Battista	Italian	1696	24	gcc
Rossetti, Dante Gabriel	British	1828	24	gcc
Piranesi, Giovanni Battista	Italian	1720	24	gcc
Degas, Edgar	French	1834	24	gcc
Duchamp, Marcel	French	1887	24	gcc
Gogh, Vincent van	Dutch	1853	24	gcc
Goya, Francisco de	Spanish	1746	24	gcc
Thorvaldsen, Bertel	Danish	1768	22	gcc
Bruegel, Pieter	Flemish	1515	22	gcc
Asam, Cosmas Damian	German	1686	22	gcc
Pissarro, Camille	French	1831	20	gcc
Mantegna, Andrea	Italian	1426	20	gcc
Kandinsky, Vasily.	Russian	1866	20	gcc
Flandrin, Hippolyte	French	1809	20	gcc
Leonardo da Vinci	Italian	1452	18	gcc
Veronese, Paolo	Italian	1528	18	gcc
Goltzius, Hendrick	Dutch	1558	18	gcc
Carracci, Agostino	Italian	1557	18	gcc
Carracci, Lodovico	Italian	1555	16	gcc
Brancusi, Constantin	Romanian	1876	16	gcc
Donatello	Italian	1380	16	gcc
Brown, Ford Madox	British	1821	16	gcc
Canova, Antonio	Italian	1757	16	gcc
Corinth, Lovis	German	1858	16	gcc
Friedrich, Caspar David	German	1774	16	gcc
Meissonier, Ernest	French	1815	14	gcc
Poussin, Nicolas	French	1594	14	gcc
Grosz, George	German	1893	14	gcc
Monet, Claude	French	1840	14	gcc
Watteau, Antoine	French	1684	14	gcc
Giulio Romano	Italian	1499	14	gcc
Ernst, Max	German	1891	14	gcc
Giacometti, Alberto	Swiss	1901	14	gcc
Richter, Gerhard	German	1932	12	gcc
Schlemmer, Oskar	German	1888	12	gcc
Miró, Joan	Spanish	1893	12	gcc
Velázquez, Diego	Spanish	1599	12	gcc
Rude, François	French	1784	12	gcc
Schwitters, Kurt	German	1887	12	gcc
Carracci, Annibale	Italian	1560	12	gcc
Lorrain, Claude	French	1604	12	gcc
Maillol, Aristide	French	1861	10	gcc
Malevich, Kazimir	Russian	1878	10	gcc
Dalí, Salvador	Spanish	1904	10	gcc
Manet, Edouard	French	1832	10	gcc
Gauguin, Paul	French	1848	10	gcc
Bonheur, Rosa	French	1822	10	gcc
Man Ray	American	1890	10	gcc
Lotto, Lorenzo	Italian	1475	10	gcc
Viollet-le-Duc, Eugène-E.	French	1814	10	gcc
Kokoschka, Oskar	Austrian	1886	10	gcc
De Chirico, Giorgio	Italian	1888	10	gcc
Caravaggio, Michelangelo M. da	Italian	1571	10	gcc
Liebermann, Max	German	1847	10	gcc
Cassatt, Mary	American	1844	8	gcc
Favorsky, Vladimir	Russian	1886	8	gcc
Brunelleschi, Filippo	Italian	1377	8	gcc
Hoffmann, Josef	Austrian	1870	8	gcc
Sargent, John Singer	American	1856	8	gcc
Beckmann, Max	German	1884	8	gcc
Dix, Otto	German	1891	8	gcc
Klimt, Gustav	Austrian	1862	8	gcc
Arp, Hans.	French	1886	8	gcc
Giordano, Luca	Italian	1634	8	gcc
Cano, Alonso	Spanish	1601	8	gcc
Gericault, Théodore	French	1791	8	gcc
Thoma, Hans	German	1839	6	gcc
Piero	Italian	1405	6	gcc
Redon, Odilon	French	1840	6	gcc
Murillo, Bartolomé Esteban	Spanish	1618	6	gcc
Millet, Jean-François	French	1814	6	gcc
Tonitza, Nicolae	Romanian	1885	6	gcc
Böcklin, Arnold	Swiss	1827	6	gcc
Giaquinto, Corrado	Italian	1703	6	gcc
Magritte, René	Belgian	1898	6	gcc
Dufy, Raoul	French	1877	6	gcc
Giorgione	Italian	1477	6	gcc
Toulouse-Lautrec, Henri de	French	1864	6	gcc
Millais, Sir John Everett	British	1829	6	gcc
Semper, Gottfried	German	1803	6	gcc
Chagall, Marc	Belorussian	1887	6	gcc
Renoir, Auguste	French	1841	6	gcc
Schiele, Egon	Austrian	1890	4	gcc
Seurat, Georges	French	1859	4	gcc
Courbet, Gustave	French	1819	4	gcc
Fischer von Erlach, Johann B.	Austrian	1656	4	gcc
Larionov, Mikhail	Russian	1881	4	gcc
Runge, Philipp Otto	German	1777	4	gcc
Meidner, Ludwig	German	1884	4	gcc
Menzel, Adolph	German	1815	4	gcc
Vauchelet, Théophile Auguste	French	1802	4	gcc
Tatlin, Vladimir	Ukrainian	1885	4	gcc
Turner, J M. W.	British	1775	4	gcc
Taut, Bruno	German	1880	4	gcc
Munch, Edvard	Norwegian	1863	4	gcc
Goncharova, Natal'ya	Russian	1881	4	gcc
Koch, Joseph Anton	Austrian	1768	2	gcc
Alberti, Leon Battista	Italian	1404	2	gcc
Caillebotte, Gustave	French	1848	2	gcc
Kollwitz, Käthe	German	1867	2	gcc
Sisley, Alfred	French	1839	2	gcc
Schinkel, Karl Friedrich	German	1781	2	gcc
Vermeer, Johannes	Dutch	1632	2	gcc
Barlacchi, Tommaso	Italian	1500	2	gcc
Dürer, Albrecht	German	1471	22	oth
Holbein, Hans	German	1497	14	oth
Giotto	Italian	1266	14	oth
Burgkmair, Hans	German	1473	10	oth
Loos, Adolf	Austrian	1870	6	oth
Martini, Simone	Italian	1274	10	oth
Blake, William	British	1757	8	oth
Greco, El]	Greek	1541	12	oth
Tintoretto	Italian	1519	10	oth
Cranach, Lucas	German	1472	8	oth
Weyden, Rogier van der	Flemish	1395	6	oth
Metsys, Quentin	Netherlandish	1466	8	oth
Morisot, Berthe	French	1841	2	oth
Stoss, Veit	German	1438	6	oth
Palladio, Andrea	Italian	1508	6	oth
Eyck, Jan van	Netherlandish	1380	6	oth
Modersohn-Becker, Paula	German	1876	2	oth
Chodowiecki, Daniel Nikolaus	German	1726	6	oth
Constable, John	British	1776	4	oth
Kirchner, Ernst Ludwig	German	1880	4	oth
Stethaimer, Hans, the younger	German	1390	2	oth
Terragni, Giuseppe	Italian	1904	2	oth
Machuca, Pedro	Spanish	1480	2	oth
Bonnard, Pierre	French	1867	2	oth
Maulbertsch, Franz Anton	Austrian	1724	2	oth
Mondrian, Piet	Dutch	1872	2	oth
Bosch, Hieronymus	Netherlandish	1440	2	oth
Cézanne, Paul	French	1839	2	oth
Marc, Franz	German	1880	2	oth
Lorenzetti, Pietro	Italian	1276	2	oth
Rodriguez, Alonso	Italian	1568	2	oth
Witkiewicz, Stanisław	Polish	1851	2	oth
Macke, August	German	1887	0	NA
Atget, Eugène	French	1857	0	NA
Bartholdi, Frédéric-Auguste	French	1834	0	NA
Ensor, James	Belgian	1860	0	NA
Lorca, Federico García	Spanish	1998	0	NA
Poiret, Paul	French	1879	0	NA
Hunt, Harry Millson	British	1836	0	NA
Rousseau, Henri	French	1844	0	NA
Daumier, Honoré	French	1810	0	NA
Boudolf, Jan	Netherlandish	1338	0	NA

Table A.4: Persons present in in scholarly rankings as distributed and ranked over ULAN connected components

APPENDIX B

Supplementary Materials

The following supplementary materials are provided online at

`https://github.com/d0rg0ld/thesis`

- Transcripts of the interviews with the domain experts for collecting user feedback for design artifact **A1** (Figure 3.6).
- Network data for exploring design artifacts

  A3 (Figure 4.24)
  A4 (Figure 5.19)
  A7 (Figure 6.8)
  A8 (Figure 6.10)
  A10 (Figure 6.14)
  A11 (Figures 6.16, 6.17 and 6.18)

  using the Gephi[1] network visualization software package.

[1] `https://gephi.org`

Bibliography

[Agresti, 2007] Agresti, A. (2007). *An introduction to categorical data analysis*. Wiley series in probability and mathematical statistics. Wiley-Interscience, Hoboken, NJ, 2nd edition.

[Ahn et al., 2007] Ahn, Y.-Y., Han, S., Kwak, H., Moon, S., and Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*, pages 835–844, New York, NY. ACM.

[Albrecht et al., 2016] Albrecht, K., Dörk, M., and Müller, B. (2016). Culturegraphy. *Leonardo*, 50(2):209–210.

[Alexiev et al., 2020] Alexiev, V., Tarkalanov, P., Georgiev, N., and Pavlova, L. (2020). Bulgarian Icons in Wikidata and EDM. In *Digital Presentation and Preservation of Cultural and Scientific Heritage, Conference Proceedings*, volume 10, Sofia, Bulgaria. Bulgarian Academy of Sciences.

[Amigoni and Schiaffonati, 2009] Amigoni, F. and Schiaffonati, V. (2009). The Minerva system: A step toward automatically created virtual museums. *Applied Artificial Intelligence*, 23(3):204–232.

[Anderson, 2005] Anderson, B. (2005). Imagined communities. In Spencer, P. and Wollman, H., editors, *Nations and nationalism: A reader*. Rutgers University Press, New Brunswick, NJ.

[Andrews et al., 1996] Andrews, K., Pichler, M., and Wolf, P. (1996). Towards rich information landscapes for visualising structured Web spaces. In *Proceedings IEEE Symposium on Information Visualization '96*, pages 62–63, New York, NY. IEEE.

[Aragon et al., 2012] Aragon, P., Laniado, D., Kaltenbrunner, A., and Volkovich, Y. (2012). Biographical social networks on Wikipedia: A cross-cultural study of links that made history. In *Proceedings of the eight Annual International Symposium on Wikis and Open Collaboration (WikiSym '12)*, pages 19:1–19:4, New York, NY. ACM.

[Arends et al., 2010] Arends, M., Froschauer, J., Goldfarb, D., Merkl, D., and Weingartner, M. (2010). Interaction with museum content in Web3d. In *Konferenzband EVA*

2010 Berlin. Elektronische Bildverarbeitung & Kunst, Kultur, Historie, pages 161–167, Heidelberg, Germany. arthistoricum.net.

[Arends and Goldfarb, 2010] Arends, M. and Goldfarb, D. (2010). Social interaction with cultural heritage on the web. In *Proceedings of the International Conference on Web Engineering (ICWE '10)*, volume 6189 of *Lecture Notes in Computer Science*, pages 587–592, Berlin, Germany. Springer.

[Arends et al., 2009] Arends, M., Goldfarb, D., Merkl, D., and Weingartner, M. (2009). Interaction with art museums on the web. In White, B., Isaías, P., and Nunes, M. B., editors, *Proceedings of the IADIS International Conference WWW/Internet 2009*, pages 117–125, Lisbon, Portugal. IADIS Press.

[Assmann, 2008] Assmann, A. (2008). Canon and archive. In Erll, A., Nünning, A., and Young, S. B., editors, *Cultural memory studies: An international and interdisciplinary handbook*, volume 8 of *Media and cultural memory*, pages 97–108. de Gruyter, Berlin, Germany.

[Athenikos and Lin, 2009] Athenikos, S. J. and Lin, X. (2009). The WikiPhil portal: Visualizing meaningful philosophical connections. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(1).

[Auer et al., 2007] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). DBpedia: A nucleus for a Web of open data. In Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., and Cudré-Mauroux, P., editors, *The Semantic Web. ISWC 2007, ASWC 2007*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer, Berlin, Germany.

[Baca and Gill, 2015] Baca, M. and Gill, M. (2015). Encoding multilingual knowledge systems in the digital age: The Getty Vocabularies. *Knowledge Organization*, 42(4):232–243.

[Barabási, 2002] Barabási, A.-L. (2002). *Linked: The new science of networks*. Perseus Publishing, Cambridge, MA.

[Barr, 1936] Barr, A. H. (1986 [1936]). *Cubism and abstract art*. Harvard University Press, Cambridge, MA, 3rd edition.

[Barr, 1933] Barr, Jr., A. H. (1933). Report on the permanent collection. In *The Museum of Modern Art archives, Alfred H. Barr, Jr. papers*, volume 9a.7A. Museum of Modern Art, New York, NY.

[Bearman, 1995] Bearman, D. (1995). Standards for networked cultural heritage. *Archives and Museum Informatics*, 9(3):279–307.

[Bellomi and Bonato, 2005] Bellomi, F. and Bonato, R. (2005). Network analysis for Wikipedia. In *Proceedings of Wikimania 2005, The first International Wikimedia Conference*, Frankfurt am Main, Germany.

[Belting, 1992] Belting, H. (1992). *Die Deutschen und ihre Kunst: ein schwieriges Erbe*. C.H.Beck, Munich, Germany.

[Berners-Lee, 2006] Berners-Lee, T. (2006). Linked Data - Design Issues. `https://www.w3.org/DesignIssues/LinkedData.html`.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5):28–37.

[Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

[Brandes and Köpf, 2001] Brandes, U. and Köpf, B. (2001). Fast and simple horizontal coordinate assignment. In P., M., M., J., and S., L., editors, *Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 31–44, Berlin, Germany. Springer.

[Buffington, 2008] Buffington, M. L. (2008). Creating and consuming Web 2.0 in art education. *Computers in the Schools*, 25(3-4):303–313.

[Callahan and Herring, 2011] Callahan, E. S. and Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology*, 62(10):1899–1915.

[Cameron, 2003] Cameron, F. (2003). Digital Futures I: Museum collections, digital technologies, and the cultural construction of knowledge. *Curator: The Museum Journal*, 46(3):325–340.

[Cameron and Robinson, 2007] Cameron, F. and Robinson, H. (2007). Digital knowledgescapes: Cultural, theoretical, practical, and usage issues facing museums collection databases in a digital epoch. In Cameron, F. and Kenderdine, S., editors, *Theorizing digital cultural heritage: A critical discourse*, Media in Transition, chapter 9, pages 165–191. MIT Press, Cambridge, MA.

[Clauset, 2013] Clauset, A. (2013). Which vertices are important? In *Network analysis and modeling [Lecture notes]*, volume 3. Santa Fe Institute, Santa Fe, NM.

[Clauset et al., 2009] Clauset, A., Shalizi, C., and Newman, M. (2009). Power-Law distributions in empirical data. *SIAM Review*, 51(4):661–703.

[Clough et al., 2009] Clough, P. D., Ireson, N., and Marlow, J. (2009). Extending domain-specific resources to enable semantic access to cultural heritage data. *Journal of Digital Information*, 10(6).

[Cooper, 1994] Cooper, M. (1994). Information Landscapes. *MIT Technical Note*.

[Daquino, 2020] Daquino, M. (2020). A computational analysis of art historical linked data for assessing authoritativeness of attributions. *Journal of the Association for Information Science and Technology*, 71(7):757–769.

[Darnton, 2013] Darnton, R. (2013). The National Digital Public Library is launched! *The New York Review of Books*, 60(7).

[Deleuze and Guattari, 1992] Deleuze, G. and Guattari, F. (1992). *Kapitalismus und Schizophrenie: Tausend Plateaus*. Merve-Verlag, Berlin, Germany.

[Dickerman and Affron, 2012] Dickerman, L. and Affron, M. (2012). *Inventing abstraction, 1910-1925: How a radical idea changed modern art*. Thames & Hudson, London, United Kingdom.

[Dilly, 1992] Dilly, H. (1992). Der kunsthistorische Nachthimmel, Ein Beitrag zur Kritik kunsthistorischer Praxis. In Berndt, A., Kaiser, P., Rosenberg, A., and Trinkner, D., editors, *Frankfurter Schule und Kunstgeschichte*, pages 69–84. Reimer, Berlin, Germany.

[Dolan, 2012] Dolan, J. (2012). Polari - An etymology according to a diagrammatic by Alfred H. Barr (1936). `http://jezdolan.com/photo/polari-an-etymology-according-to-a-diagrammatic-by-alfred-h-barr-1936-2012/220344`.

[Dörk et al., 2011] Dörk, M., Carpendale, S., and Williamson, C. (2011). The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 1215–1224, New York, NY. ACM.

[Dörk et al., 2012] Dörk, M., Carpendale, S., and Williamson, C. (2012). Visualizing explicit and implicit relations of complex information spaces. *Information Visualization*, 11(1):5–21.

[Dörk et al., 2017] Dörk, M., Pietsch, C., and Credico, G. (2017). One view is not enough. *Information Design Journal*, 23(1):39–47.

[Duncan and Wallach, 1980] Duncan, C. and Wallach, A. (1980). The universal survey museum. *Art History*, 3(4):448–469.

[Easley and Kleinberg, 2010] Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, New York, NY.

[Eisler, 1987] Eisler, C. (1987). "Every artist paints himself": Art history as biography and autobiography. *Social Research*, 54(1):73–99.

[Elgammal et al., 2018] Elgammal, A., Mazzone, M., Liu, B., Kim, D., and Elhoseiny, M. (2018). The shape of art history in the eyes of the machine. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pages 2183–2191, Palo Alto, CA. AAAI Press.

[Elkins, 2007] Elkins, J. (2007). Canon and globalization in art history. In Brzyski, A., editor, *Partisan canons*, chapter 2, pages 55–77. Duke University Press, Durham, NC.

[Eom et al., 2015] Eom, Y.-H., Aragón, P., Laniado, D., Kaltenbrunner, A., Vigna, S., and Shepelyansky, D. L. (2015). Interactions of cultures and top people of Wikipedia from ranking of 24 language editions. *PLOS ONE*, 10(3):e0114825.

[Feral, 2011] Feral, D. (2011). Graffiti & street art. `http://www.concretetodata.com/artists/daniel-feral/`.

[Fleischer and Hirsch, 2001] Fleischer, R. and Hirsch, C. (2001). Graph drawing and its applications. In Kaufmann, M. and Wagner, D., editors, *Drawing graphs: Methods and models*, volume 2025 of *Lecture Notes in Computer Science*, chapter 1, pages 1–22. Springer, Berlin, Germany.

[Framestore, 2015] Framestore (2015). Timeline of modern art, Tate. `https://www.framestore.com/timelineofmodernart`.

[Freeman, 1978] Freeman, L. C. (1978). Centrality in social networks. *Social Networks*, 1(3):215–239.

[Fry, 1913] Fry, R. (1913). *Records of journeys to Venice and the Low Countries by Albrecht Dürer*. Merrymount Press, Boston, MA.

[Gansner et al., 1993] Gansner, E. R., Koutsofios, E., North, S. C., and Vo, K. P. (1993). A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*, 19(3):214–230.

[Getty Research Institute, 2015] Getty Research Institute (2015). Getty Vocabularies Editorial Guidelines. `http://getty.edu/research/tools/vocabularies/guidelines/index.html`.

[Geymüller, 1898] Geymüller, H. B. v. (1898). *Die Baukunst der Renaissance in Frankreich*, volume 6 of *Handbuch der Architektur Theil 2*. A. Bergsträsser, Stuttgart, Germany.

[Gibson et al., 2013] Gibson, H., Faith, J., and Vickers, P. (2013). A survey of two-dimensional graph layout techniques for information visualisation. *Information Visualization*, 12(3-4):324–357.

[Giles, 2005] Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901.

[Gillespie, 2014] Gillespie, C. S. (2014). Fitting heavy tailed distributions: the poweRlaw package. *arXiv e-prints*. arXiv:1407.3492.

[Gobbi et al., 2014] Gobbi, A., Iorio, F., Dawson, K. J., Wedge, D. C., Tamborero, D., Alexandrov, L. B., Lopez-Bigas, N., Garnett, M. J., Jurman, G., and Saez-Rodriguez, J. (2014). Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics*, 30(17):i617–i623.

[Goldfarb et al., 2011a] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2011a). Revisiting 3d information landscapes for the display of art historical web content. In *Proceedings of the eight International Conference on Advances in Computer Entertainment Technology (ACE '11)*, pages 46:1–46:8, New York, NY. ACM.

[Goldfarb et al., 2012] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2012). Art history on wikipedia, a macroscopic observation. In *Extended Abstract presented at the ACM Web Science Conference (WebSci '12)*, Evanston, Illinois.

[Goldfarb et al., 2013] Goldfarb, D., Arends, M., Froschauer, J., and Merkl, D. (2013). Comparing art historical networks. *Leonardo*, 46(3):279–279.

[Goldfarb et al., 2011b] Goldfarb, D., Arends, M., Froschauer, J., Merkl, D., and Weingartner, M. (2011b). Combining cultural heritage related web resources in 3d information landscapes. In *Proceedings of the 2011 International Conference on Electronic Visualisation and the Arts (EVA London 2011)*, pages 73–80, Swindon, United Kingdom. BCS Learning & Development.

[Goldfarb et al., 2014] Goldfarb, D., Arends, M., Froschauer, J., Weingartner, M., and Merkl, D. (2014). Collectivizing the Barr Model. *Leonardo*, 47(3):270–270.

[Goldfarb and Merkl, 2016] Goldfarb, D. and Merkl, D. (2016). Bilder der Kunstgeschichte in der Getty Ulan und in Wikipedia. In *Konferenzband EVA 2016 Berlin. Elektronische Bildverarbeitung & Kunst, Kultur, Historie*, pages 33–42, Heidelberg, Germany. arthistoricum.net.

[Goldfarb and Merkl, 2018] Goldfarb, D. and Merkl, D. (2018). Visualizing art historical developments using the Getty ULAN, Wikipedia and Wikidata. In *Proceedings of the 22th International Conference Information Visualisation (iV '18)*, pages 459–466, New York, NY. IEEE.

[Goldfarb and Merkl, 2019] Goldfarb, D. and Merkl, D. (2019). Data-driven maps of art history. *International Journal of Art, Culture and Design Technologies (IJACDT)*, 8(1):1–15.

[Goldfarb et al., 2015] Goldfarb, D., Merkl, D., and Schich, M. (2015). Quantifying cultural histories via person networks in Wikipedia. In *Poster presented at the International School and Conference on Network Science (NetSci '15)*, Zaragoza.

[Gompertz, 2012] Gompertz, W. (2012). *What are you looking at?: 150 years of modern art in the blink of an eye*. Penguin, New York, NY.

[Green-Armytage, 2010] Green-Armytage, P. (2010). A colour alphabet and the limits of colour coding. *JAIC - Journal of the International Colour Association*, 5:1–23.

[Hackenbroch, 1969] Hackenbroch, Y. (1969). Some portraits of Charles V. *The Metropolitan Museum of Art Bulletin*, 27(6):323–332.

[Halavais and Lackaff, 2008] Halavais, A. and Lackaff, D. (2008). An analysis of topical coverage of Wikipedia. *Journal of Computer-Mediated Communication*, 13(2):429–440.

[Hamlin, 2017] Hamlin, A. K. (2017). Approaching intellectual emancipation: Critical reading in art, art history, and Wikipedia. In DiYanni, R. and Borst, A., editors, *Critical reading across the curriculum: Humanities*, chapter 6, pages 104–122. John Wiley & Sons, Hoboken, NJ.

[Hatt and Klonk, 2006] Hatt, M. and Klonk, C. (2006). *Art history: A critical introduction to its methods*. Manchester University Press, Manchester, United Kingdom.

[Havre et al., 2002] Havre, S., Hetzler, E., Whitney, P., and Nowell, L. (2002). ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20.

[Hazan, 2010] Hazan, S. (2010). Musing the metaverse. In *Heritage in the Digital Era, Proceedings of the 14th International Conference on Virtual Systems and Multimedia (VSMM '08)*, Essex, United Kingdom. Multi-Science Publishing.

[Hecht and Gergle, 2009] Hecht, B. and Gergle, D. (2009). Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the fourth International Conference on Communities and Technologies (C&T '09)*, pages 11–20, New York, NY. ACM.

[Heck, 2014] Heck, K. (2014). Formen des Stils. Heinrich von Geymüllers grafische Methoden der Stilgeschichtsschreibung. In Cortjaens, W. and Heck, K., editors, *Stil-Linien diagrammatischer Kunstgeschichte*, volume 2 of *Transformationen des Visuellen*, pages 168–185. Deutscher Kunstverlag, Berlin, Germany.

[Henning et al., 2007] Henning, A., Hagen, L., and Böttcher, A. (2007). Die Gemäldegalerie Alte Meister Dresden in Second Life - Zur Nutzung neuer Medien als Instrument des Beziehungsmanagements. In Bienert, A., editor, *Konferenzband EVA 2007 Berlin. Elektronische Bildverarbeitung & Kunst, Kultur, Historie*, pages 53–58, Heidelberg, Germany. arthistoricum.net.

[Hevner et al., 2004] Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1):75–105.

[Hilles, 2014] Hilles, S. (2014). To use or not to use? The credibility of Wikipedia. *Public Services Quarterly*, 10(3):245–251.

[Hinrichs et al., 2008] Hinrichs, U., Schmidt, H., and Carpendale, S. (2008). EMDialog: Bringing information visualization into the museum. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1181–1188.

[Holloway et al., 2007] Holloway, T., Bozicevic, M., and Börner, K. (2007). Analyzing and visualizing the semantic coverage of Wikipedia and its authors. *Complexity*, 12(3):30–40.

[Huhtamo, 2002] Huhtamo, E. (2010 [2002]). On the origins of the virtual museum. In Parry, R., editor, *Museums in a digital age*, chapter 12. Routledge, Abingdon, United Kingdom.

[Jacomy et al., 2014] Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE*, 9(6):e98679.

[Jensen, 2007] Jensen, R. (2007). Measuring canons: Reflections on innovation and the ninteenth-century canon of European art. In Brzyski, A., editor, *Partisan canons*, chapter 1, pages 27–54. Duke University Press, Durham, NC.

[Joyeux-Prunel, 2014] Joyeux-Prunel, B. (2014). The uses and abuses of peripheries in art history. *Artl@s Bulletin*, 3(1):1.

[Kaiser, 2017] Kaiser, M. (2017). Was uns Biographien über Künstlernetzwerke sagen. Konzepte für eine historische Netzwerkanalyse auf Basis biographischer Texte aus dem österreichischen Biographischen Lexikon (öBL). In Bernad, A. Z., Gruber, C., and Kaiser, M., editors, *Europa baut auf Biographien. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*, pages 383–403, Vienna, Austria. New Academic Press.

[Kamps and Koolen, 2009] Kamps, J. and Koolen, M. (2009). Is Wikipedia link structure different? In *Proceedings of the second ACM International Conference on Web Search and Data Mining (WSDM '09)*, pages 232–241, New York, NY. ACM.

[Kitromilidis and Evans, 2018] Kitromilidis, M. and Evans, T. S. (2018). Community detection with metadata in a network of biographies of western art painters. *arXiv e-prints*. arXiv:1802.07985.

[Klein and Kyrios, 2013] Klein, M. and Kyrios, A. (2013). VIAFbot and the integration of library data on Wikipedia. *Code4Lib Journal*, 2:85–107.

[Konieczny and Klein, 2018] Konieczny, P. and Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*, 20(12).

[Koschützki et al., 2005] Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., and Zlotowski, O. (2005). Centrality indices. In Brandes, U. and Erlebach, T., editors, *Network analysis: Methodological foundations*, volume 3418 of *Theoretical Computer Science and General Issues*, chapter 3, pages 16–61. Springer, Berlin, Germany.

[Kurki and Hyvönen, 2009] Kurki, J. and Hyvönen, E. (2009). Authority control of people and organizations on the semantic web. In *Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009*, Trento, Italy.

[Kwasnik, 1992] Kwasnik, B. H. (1992). A descriptive study of the functional components of browsing. In *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction*, volume 18 of *IFIP Transactions*, pages 191–203, Amsterdam, Netherlands. North-Holland.

[Lincoln, 2016] Lincoln, M. (2016). Social network centralization dynamics in print production in the Low Countries, 1550-1750. *International Journal for Digital Art History*, 2:134–157.

[Lincoln, 2020] Lincoln, M. D. (2020). Network Thinking and Network Analysis in the History of Art. In Brown, K., editor, *The Routledge Companion to Digital Humanities and Art History*. Routledge, New York, Abingdon.

[Lippe, 2011] Lippe, P. v. d. (2011). Wie großmuss meine Stichprobe sein, damit sie repräsentativ ist? Wie viele Einheiten müssen befragt werden? Was heißt "Repräsentativität"? *Diskussionsbeiträge aus dem FB Wirtschaftswissenschaften Univ. Duisburg-Essen (Campus Essen)* , Heft 187.

[Locher, 2014] Locher, H. (2014). The idea of cultural heritage and the canon of art. *Kunstiteaduslikke Uurimusi*, 23(3/4):20–35.

[MacDonald and Alsford, 1991] MacDonald, G. F. and Alsford, S. (1991). The museum as information utility. *Museum Management and Curatorship*, 10(3):305–311.

[Malraux, 1949] Malraux, A. (1949). *Das Imaginäre Museum*, volume 1 of *Psychologie der Kunst*. Klein, Baden-Baden, Germany.

[Manguinhas et al., 2016] Manguinhas, H., Freire, N., Isaac, A., Stiller, J., Charles, V., Soroa, A., Simon, R., and Alexiev, V. (2016). Exploring comparative evaluation of semantic enrichment tools for cultural heritage metadata. In Fuhr, N., Kovács, L., Risse, T., and Nejdl, W., editors, *Research and Advanced Technology for Digital Libraries (TPDL 2016)*, volume 9819 of *Lecture Notes in Computer Science*, pages 266–278, Cham, Switzerland. Springer.

[Manovich, 2002] Manovich, L. (2002). *The language of new media*. MIT Press, Cambridge, MA.

[Mccauley, 2010] Mccauley, D. (2010). Hip-Hop music. http://williamsmccallgallery.com/david-mccauley.html.

[Mihalcea and Csomai, 2007] Mihalcea, R. and Csomai, A. (2007). Wikify!: Linking documents to encyclopedic knowledge. In *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM '07)*, pages 233–242, New York, NY. ACM.

[Moretti, 2013] Moretti, F. (2013). *Distant reading*. Verso Books, London, United Kingdom.

[Nature, 2006] Nature (2006). Britannica attacks. *Nature*, 440(7084):582–582.

[Newman, 2003a] Newman, M. (2003a). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

[Newman, 2010] Newman, M. (2010). *Networks: An introduction*. Oxford University Press, Oxford, United Kingdom.

[Newman, 2003b] Newman, M. E. J. (2003b). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

[Newman, 2006] Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

[Newton, 1941] Newton, E. (1956 [1941]). *European painting and sculpture*. Penguin Books, Harmondsworth, United Kingdom, 4th edition.

[Niggemann et al., 2011] Niggemann, E., De Decker, J., and Lévy, M. (2011). *The new Renaissance - Report of the 'Comité des Sages' reflection group on bringing Europe's cultural heritage online*. European Commission, Brussels, Belgium.

[Oomen and Aroyo, 2011] Oomen, J. and Aroyo, L. (2011). Crowdsourcing in the cultural heritage domain: Opportunities and challenges. In *Proceedings of the fifth International Conference on Communities and Technologies (C&T '11)*, pages 138–149, New York, NY. ACM.

[O'reilly, 2010] O'reilly, T. (2010). What is Web 2.0: Design patterns and business models for the next generation of software. In Donelan, H. M., Kear, K., and Ramage, M., editors, *Online Communication and Collaboration: A Reader*. Routledge, London, United Kingdom.

[Page et al., 1999] Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank citation ranking: Bringing order to the Web. http://ilpubs.stanford.edu:8090/422/. Technical Report, Stanford InfoLab.

[Pampalk et al., 2002] Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM International Conference on Multimedia (MULTIMEDIA '02)*, pages 570–579, New York, NY. ACM.

[Parry et al., 2008] Parry, R., Pratty, J., and Poole, N. (2008). Semantic dissonance: Do we need (and do we understand) the semantic Web? In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Montreal, Canada. Archives & Museum Informatics.

[Pentzold, 2009] Pentzold, C. (2009). Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, 2(2):255–272.

[Perra and Fortunato, 2008] Perra, N. and Fortunato, S. (2008). Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107.

[Priestley, 1765] Priestley, J. (1765). *A description of a Chart of Biography; with a catalogue of all the names inserted in it, and the dates annexed to them.* Author and J. Bowles, London, United Kingdom, 2nd edition.

[Pupyrev et al., 2010] Pupyrev, S., Nachmanson, L., and Kaufmann, M. (2010). Improving layered graph layouts with edge bundling. In Brandes, U. and Cornelsen, S., editors, *Graph Drawing (GD 2010)*, volume 6502 of *Lecture Notes in Computer Science*, pages 329–340, Berlin, Germany. Springer.

[Purchase, 2004] Purchase, H. C. (2004). Evaluating graph drawing aesthetics: Defining and exploring a new empirical research area. In DiMarco, J., editor, *Computer graphics and multimedia: Applications, problems and solutions*, pages 145–179. IGI Global, Hershey, PA.

[Rosenberg and Grafton, 2010] Rosenberg, D. and Grafton, A. (2010). *Cartographies of time: A history of the timeline.* Princeton Architectural Press, New York, NY.

[Rosenzweig, 2006] Rosenzweig, R. (2006). Can history be open source? Wikipedia and the future of the past. *The Journal of American History*, 93(1):117–146.

[Ross, 2013] Ross, N. (2013). Teaching twentieth century art history with gender and data visualizations. *The Journal of Interactive Technology and Pedagogy*, (4).

[Ruffaldi et al., 2008] Ruffaldi, E., Evangelista, C., Neri, V., Carrozzino, M., and Bergamasco, M. (2008). Design of information landscapes for cultural heritage content. In *Proceedings of the third International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA '08)*, pages 113–119, New York, NY. ACM.

[Samoilenko et al., 2016] Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., and Strohmaier, M. (2016). Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity. *EPJ Data Science*, 5(9).

[Samoilenko and Yasseri, 2014] Samoilenko, A. and Yasseri, T. (2014). The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. *EPJ Data Science*, 3(1).

[Schich et al., 2017] Schich, M., Huemer, C., Adamczyk, P., Manovich, L., and Liu, Y.-Y. (2017). Network dimensions in the Getty Provenance Index. *arXiv e-prints*. arXiv:1706.02804.

[Schich et al., 2014] Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., and Helbing, D. (2014). A network framework of cultural history. *Science*, 345(6196):558–562.

[Schmidt-Burkhardt, 2005] Schmidt-Burkhardt, A. (2005). *Stammbäume der Kunst: Zur Genealogie der Avantgarde*. Akademie Verlag, Berlin, Germany.

[Schreiber et al., 2008] Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Osenbruggen, J., Tordai, A., and others (2008). Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):243–249.

[Schweibenz, 2004] Schweibenz, W. (2004). The development of virtual museums. *ICOM News*, (3).

[Sharpe, 2015] Sharpe, D. (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, 20(8).

[Sherratt, 2013] Sherratt, T. (2013). From portals to platforms: Building new frameworks for user engagement. In *Proceedings of the 2013 LIANZA Conference*, pages 20–23, Hamilton, New Zealand.

[Sooke, er 4] Sooke, A. (2016, October 4). Was modern art a weapon of the CIA? *bbc.com*. `http://www.bbc.com/culture/story/20161004-was-modern-art-a-weapon-of-the-cia`.

[Sumi et al., 2011] Sumi, R., Yasseri, T., Rung, A., Kornai, A., and Kertesz, J. (2011). Edit wars in Wikipedia. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 724–727, New York, NY. IEEE.

[Teplitskiy et al., 2017] Teplitskiy, M., Lu, G., and Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9):2116–2127.

[Terras, 2015] Terras, M. (2015). Opening Access to collections: the making and using of open digitised cultural content. *Online Information Review*, 39(5):733–752.

[Tiampo, 2010] Tiampo, M. (2010). Originality, universality, and other modernist myths. In Elkins, J., Valiavicharska, Z., and Kim, A., editors, *Art and globalization*, pages 166–170. Penn State University Press, University Park, PA.

[Trant, 2009a] Trant, J. (2009a). Emerging convergence? Thoughts on museums, archives, libraries, and professional training. *Museum Management and Curatorship*, 24(4):369–387.

[Trant, 2009b] Trant, J. (2009b). Tagging, Folksonomy and art museums: Results of steve.museum' s research. `https://repository.arizona.edu/bitstream/handle/10150/105627/trantSteveResearchReport2008.pdf`.

[Vasari, 1991] Vasari, G. (1991). *The lives of the artists*. Oxford University Press, UK, Oxford, United Kingdom.

[Voß et al., 2014] Voß , J., Bausch, S., Bogner, J., Schmitt, J., Berkelmann, V., Ludemann, F., Löffel, O., Kitroschat, J., Bartoshevska, M., and Seljuzki, K. (2014). *Normdaten in Wikidata*. Lulu Press, Morrisville, NC.

[Vrandečić and Krötzsch, 2014] Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

[Wang et al., 2009] Wang, Y., Stash, N., Aroyo, L., Hollink, L., and Schreiber, G. (2009). Using semantic relations for content-based recommender systems in cultural heritage. In *Proceedings of the 2009 International Conference on Ontology Patterns*, volume 516 of *CEUR Workshop Proceedings*, pages 16–28. CEUR-WS.

[Watts and Strogatz, 1998] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

[White and Roth, 2009] White, R. W. and Roth, R. A. (2009). Exploratory Search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98.

[Wise et al., 1995] Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., and Crow, V. (1995). Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proceedings of Visualization 1995 Conference*, pages 51–58, New York, NY. IEEE.

[Wölfflin, 1915] Wölfflin, H. (1915). *Kunstgeschichtliche Grundbegriffe. Das Problem der Stilentwickelung in der neueren Kunst.* F. Bruckmann A.-G., Munich, Germany.

[Zant et al., 2018] Zant, S. E., Jaffrès-Runser, K., Frahm, K. M., and Shepelyansky, D. L. (2018). Interactions and influence of world painters from the reduced Google matrix of Wikipedia networks. *IEEE Access*, 6:47735–47750.

[Zeng and Mayr, 2018] Zeng, M. L. and Mayr, P. (2018). Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. *International Journal on Digital Libraries*, 20:209—230.

[Zerubavel, 2011] Zerubavel, E. (2011). *Ancestors and relatives: Genealogy, identity, and community*. Oxford University Press, Oxford, United Kingdom.

[Zlatić et al., 2006] Zlatić, V., Božičević, M., Štefančić, H., and Domazet, M. (2006). Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74(1):016115.

[Zöllner, 2010] Zöllner, F. (2010). ‚Eilig Reisende' im Gebiete der Bildvergleichung: Aby Warburgs Bilderatlas ‚Mnemosyne' und die Tradition der Atlanten. *Marburger Jahrbuch für Kunstwissenschaft*, 37:279–304.

[Zweig and Kaufmann, 2011] Zweig, K. A. and Kaufmann, M. (2011). A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining*, 1(3):187–218.