Informatics

Dynamische Netzwerkanalyse mit Zentralitätsmaßen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Software Engineering & Internet Computing

eingereicht von

Luca David Rauchenberger, BSc
Matrikelnummer 01351750

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch
Mitwirkung: Dipl.-Ing. Dr. Velitchko Filipov

Wien, 16. Oktober 2024

Luca David Rauchenberger | Silvia Miksch

Informatics

Dynamic Network Analysis with Centrality Measures

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Software Engineering & Internet Computing

by

Luca David Rauchenberger, BSc
Registration Number 01351750

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch
Assistance: Dipl.-Ing. Dr. Velitchko Filipov

Vienna, October 16, 2024

Luca David Rauchenberger Silvia Miksch

Erklärung zur Verfassung der Arbeit

Luca David Rauchenberger, BSc

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, 16. Oktober 2024

Luca David Rauchenberger

Danksagung

Zuallererst möchte ich mich bei Dipl.-Ing. Dr. Velitchko Filipov bedanken. Seine kontinuierliche Unterstützung, sein wertvoller Rat und sein tiefgehendes Feedback waren von unschätzbarem Wert.

Ich bin außerdem sehr dankbar bei Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch für die Betreuung meiner Arbeit und dafür, dass sie mir die Möglichkeit gegeben hat, meine Forschung am Centre for Visual Analytics Science and Technology (CVAST) durchzuführen.

Mein Dank gilt auch Teresa Kamencek, BA MA, für ihre Unterstützung bei der Vernetzung mit Kunsthistorikerinnen und Kunsthistoriker für den Evaluierungsteil meiner Arbeit und für ihre kunsthistorischen Einblicke.

Zu guter Letzt möchte ich meinen tiefsten Dank meinen Eltern aussprechen. Ohne ihre beständige Unterstützung und Ermutigung hätte ich diesen Punkt nicht erreicht.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Dipl.-Ing. Dr. Velitchko Filipov. His consistent support, invaluable guidance, and insightful feedback have been instrumental throughout my research and writing.

I am also very thankful for Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch supervising my thesis and for giving me the opportunity to conduct my research at the Centre for Visual Analytics Science and Technology (CVAST).

I would also like to thank Teresa Kamencek, BA MA, for her assistance in connecting me with art historian experts for the evaluation part of my thesis and for her insights from an art historical perspective.

Finally, I am deeply grateful to my parents. Without their steadfast support and encouragement, I would not have reached this point.

Kurzfassung

Die Analyse großer, dynamischer Netzwerkdaten, wie zum Beispiel der Database of Modern Exhibitions, stellt aufgrund ihres Umfangs und ihrer Komplexität eine Herausforderung dar. Zentralitätsmaße helfen, diese Komplexität zu bewältigen, indem sie Algorithmen verwenden, um die Bedeutung jedes Knotens des Netzwerks zu quantifizieren. Ein wichtiger nächster Schritt besteht jedoch darin, zu erforschen, wie diese berechneten Zentralitätsmetriken in aussagekräftige visuelle Darstellungen umgewandelt werden können, um Erkenntnisse zu gewinnen und Schlüsselakteurinnen und Schlüsselakteure zu identifizieren.

Zu diesem Zweck haben wir eine umfassende Literaturrecherche durchgeführt, um zu untersuchen, wie Zentralitätsmaße in der Datenvisualisierung angewendet werden. Diese Erkenntnisse führten zur Entwicklung von dome-insights, einem Visual-Analytics-Tool, das Zentralitätsmaße sowohl in seine Visualisierungs- als auch in das Interaktionsdesign integriert. Dome-insights dient als Prototyp, um zu zeigen, wie die Einbindung von Zentralitätsmaßen die Exploration großer, komplexer Netzwerke verbessern und die Entdeckung von Erkenntnissen sowie die Identifizierung von Schlüsselakteurinnen und Schlüsselakteuren erleichtern kann.

Das Tool wurde von Kunsthistorikerinnen und Kunsthistorikern evaluiert und erzielte positive Ergebnisse sowohl in quantitativen als auch in qualitativen Methoden. Es zeigte, dass es in der Lage ist, bedeutsame Erkenntnisse zu gewinnen. Darüber hinaus deuten die Ergebnisse auf das Potenzial von Zentralitätsmaßen für die dynamische Netzwerkanalyse im Allgemeinen hin und unterstreichen ihre Rolle bei der Verbesserung der visuellen Analyse komplexer Netzwerkdaten.

Abstract

Analyzing large, dynamic network data, such as the Database of Modern Exhibitions, presents significant challenges due to the dataset's scale and complexity, as it tracks a decade of European art exhibitions and encompasses thousands of artists with evolving relationships. Centrality measures help address this complexity by using algorithms to quantify the importance of each node. However, an important next step is to explore how these calculated importance metrics can be transformed into meaningful visual representations to extract insights and identify key actors.

To achieve this, we conducted a state-of-the-art literature review to investigate how centrality measures are applied in data visualization, which informed the development of dome-insights, a visual analytics tool that integrates centrality measures into both its visualization and interaction design. Dome-insights serves as a prototype to demonstrate how incorporating centrality measures can enhance the exploration of large, complex networks, facilitating insight discovery and identification of key actors.

The tool was assessed by art historians and delivered positive results in both quantitative and qualitative evaluations, demonstrating its ability to uncover meaningful insights. More broadly, the findings highlight the potential of centrality measures in dynamic network analysis, underscoring their role in enhancing visual analytics for complex network data.

Contents

CHAPTER 1

Introduction

1.1 Motivation & Problem Statement

Various aspects of the world can be conceptualized as networks, consisting of nodes connected by relationships. Viewing it this way facilitates the analysis of structures and patterns, ultimately helping us better understand the complexities of the world [New03]. Dynamic networks go a step further by incorporating time, allowing us to track how these structures and relationships evolve [HS12]. The dataset from the Database of Modern Exhibitions (DoME)[1] exemplifies such a dynamic network, focusing on exhibitions of modern European paintings between 1905 and 1915. With records of 1,367 exhibitions and 13,266 artists spanning over a decade, it has the potential to offer valuable insights into the transformative art world of the early 20th century [BMB$^+$20]. However, extracting from the thousands of nodes and hundreds of thousands of relationships, meaningful insights, which Saraiya et al. define as "an individual observation about the data by the participant, a unit of discovery," poses a significant challenge [SND05, p. 2].

Centrality measures (CMs) [Rod19a] can be helpful with that by identifying important nodes of a network through algorithmically calculating importance metrics for each node based on their position and relationships within the network. For example, degree centrality quantifies a node's connectedness by counting its direct links, while eigenvector centrality, a more sophisticated recursive CM, gauges a node's centrality by considering the importance of its neighbors. And since different CMs capture varying aspects of importance, they together help provide a more holistic and nuanced understanding of a node's role within the network [Rod19b].

While CMs are useful for identifying significant nodes, the complexity of these networks cannot be captured by numbers alone. Keim et al. [KMSZ08] stress the importance of effective visualization in making sense of complex network data, especially when trying

[1] DoME: `exhibitions.univie.ac.at/`. Accessed on June 13, 2024.

to understand how they evolve over time. However, visualizing complex networks can be particularly challenging when they exhibit *small-world* characteristics, as described by Watts & Strogatz [Ing05]. In such networks, most nodes are just a few steps away from each other, thus complicating visualization. The high clustering and dense connections often lead to issues, like visual clutter and scalability, making it difficult to extract meaningful insights. The DoME dataset exemplifies this, as most of its nodes are separated by just one or two degrees, highlighting the *small-world* nature of the network. Moreover, traditional visualization methods, such as node-link diagrams, often struggle with these densely connected datasets, resulting in visual clutter, overlapping edges, and difficulty distinguishing key relationships.

One promising approach to addressing the visualization of *small-world* networks is the so-called *centrality-based* method proposed by Van Ham and Wattenberg [VHW08]. Recognizing that traditional node-link diagrams often struggle with *small-world* networks, they leverage CMs to selectively remove edges with lower importance. This approach reduces clutter while at the same time preserving key structural relationships and clustering within the network. Their *centrality-based* method applied to the node-link diagram demonstrates how CMs can guide visualization and interaction design to provide clearer and more meaningful insights into a network's structure.

1.2 Aim of the Work

This research builds upon the *centrality-based* approach proposed by Van Ham and Wattenberg [VHW08], applying it to the field of visual analytics for dynamic network analysis. Similar to their enhancement of node-link diagrams, this research effort explores how CMs can now guide visualization and interaction design in the context of visual analytics. Using the DoME dataset as a representative example of large, dense, and complex networks, this research focuses on identifying effective ways to leverage CMs, both static and dynamic, to enhance network analysis. As part of this research effort, a key deliverable will be the development of a web-based analytics tool that employs various CMs. Therefore, this thesis aims to answer the following research question:

1.2.1 Main Research Question

Can centrality measures enhance visual analytics approaches for dynamic network visualization to improve insight-based analysis and help identify key actors during the analysis? This question investigates the intersection of the use of CMs and their visualization. It seeks to understand the impact of CMs on network analysis to obtain relevant insights.

In order to address the *Main Research Question*, the following sub-research questions (SRQs) are presented:

1.2.2 Sub-Research Questions

SRQ1: Which visualization methods are used to communicate insights into network analysis using centrality metrics? This sub-question aims to examine how centrality metrics derived from network data are visualized in state-of-the-art literature. Its objective is to understand which visualization techniques are employed and how they are used to convey the information encapsulated within these metrics.

SRQ2: Do novel dynamic centrality measures more effectively capture temporal changes in networks compared to their static counterparts? This question aims to explore two distinct approaches for depicting changes within dynamic networks. It compares the efficacy of conveying change in a dynamic network through visualizing static centrality metrics versus using the dynamic centrality measure, *change centrality* [FPA+12].

SRQ3: How well can a visual analytics approach for dynamic network visualization facilitate deeper insights and assist in identifying key actors using different centrality measures? This research question delves into the effectiveness of employing and consequently visualizing centrality measures for the analysis of network data. It examines the environment's impact to facilitate a deeper understanding and easier discovery of insights.

Each of the SRQs contributes to the overarching goal of enhancing the utility of CMs through visualization methods, with the ultimate objective of gaining valuable insights from social dynamic network data.

1.3 Domain

The broad nature of social networks makes them widely applicable across various fields. In order to effectively explore potential enhancements in the visualization of centrality metrics, this research narrows its focus to a specific domain within social networks. This research effort concentrates on the digital humanities, with a particular focus on art history. Using the Database of Modern Exhibitions [BMB+20], this focus allows for a concrete demonstration of how CMs and their visualization can reveal key insights, such as identifying influential artists and analyzing shifts in artistic movements over time.

1.4 Structure

In Chapter 2, we introduce core concepts of social networks and CMs, with a focus on both static and dynamic CMs.

Chapter 3 reviews the state-of-the-art literature on the visualization of CMs for the analysis of social and dynamic networks, covering both static and dynamic approaches.

In Chapter 4, we design the *dome-insights* prototype, an interactive tool for visualizing and analyzing dynamic network data, drawing insights from the literature review to inform its design.

Chapter 5 focuses on the technical implementation of the prototype, detailing the key aspects of its functionality.

Chapter 6 evaluates the prototype using a value-driven and qualitative assessment to assess the tool's effectiveness.

In Chapter 7, we interpret and discuss the findings from the evaluation phase, relating them to the three SRQs guiding the study.

Chapter 8 summarizes the contributions of this thesis and outlines potential for future research.

Finally, the Appendix includes supplementary materials.

CHAPTER 2

Basic Concepts

2.1 Social Network

A graph is a mathematical structure used to model relationships between objects. In a graph, nodes represent the objects, and lines represent the connections between these objects. A social network, which is a type of graph, models a social system consisting of actors and the ties between them. The concept of a social network emphasizes the relational structure among actors. It highlights how these connections influence behaviors and outcomes within the network, making it an important perspective for understanding social dynamics [Was94].

2.2 Centrality in Network Analysis

Rodrigues [Rod19b] describes centrality as a means to evaluate the importance or centrality of individual nodes within a network. Centrality can capture different aspects of a node's role, such as how connected it is, its position in controlling the flow of information, or other context-specific criteria. Consequently, the importance of a node is relative. This relativity underscores that the significance of a node is not only based on its immediate connections but also its position within the broader network structure. As a result, CMs are versatile tools, particularly valuable for understanding complex systems, like social networks where node importance is highly dependent on a node's relationships with other nodes.

CMs are used to quantify the importance of a node in a network based on different definitions of what it means to be important or "central". They are calculated using various mathematical approaches, each highlighting different aspects of centrality [Rod19b]. This subjectivity is further illustrated by work done by Choudhury and Uddin, which uses a composite CM consisting of degree, closeness, and betweenness centrality to express a specific kind of importance in order to align with the authors' research interests [CU23].

Traditionally, these CMs have been applied in static networks, where the connections between nodes are fixed. However, networks can also be dynamic, where connections change over time, leading to the distinction between static and dynamic CMs [Rod19b].

Static CMs

In this section, four static CMs will be presented, which evaluate the network at a single point in time. They are among the most popular and widely used CMs for providing insights into the roles and significance of nodes based on the network's current structure.

- **Eigenvector Centrality:** [Rod19b] defines eigenvector centrality as a measure of a node's importance based on the idea that connections to highly important nodes contribute more to the node's importance than equal connections to less important nodes. This measure is particularly useful in situations where the importance of a node is not just a function of how many connections it has (degree centrality), but also of the importance of the nodes it is connected to. This makes it useful for identifying influential individuals in social networks.
- **Degree Centrality:** Degree centrality is one of the simplest ways to measure a node's importance within a network. It is determined by counting the number of direct connections (or edges) that a node has with other nodes in the network [Rod19b].
- **Closeness Centrality:** Closeness centrality is a measure of how near a node is to all other nodes in a network. It captures the idea of a node's accessibility within the network by considering how quickly information can spread from that node to others [Rod19b].
- **Betweenness Centrality:** Betweenness centrality is a measure of how often a node appears on the shortest paths between other nodes in the network. It highlights nodes that serve as bridges or connectors within the network, which are crucial in controlling the flow of information or resources [Rod19b].

Dynamic CMs

The study of dynamic CMs, which takes into account the time-varying nature of networks, is still relatively unexplored compared to their static counterpart. Dynamic CMs aim to capture the importance of nodes within the context of these changing connections by providing insights that static CMs cannot.

Various static CMs have been adapted to work with temporal networks, mostly by generalizing static CMs to handle the temporal dimension [Rod19b]. For example, [TMC+17a] describes an eigenvector-based CM for dynamic networks that modifies the traditional eigenvector CM by introducing additional layers to incorporate the temporal dimension. This method enhances the analysis of networks over time by distinguishing between connections within the same time layer and those across different time layers. Doing so enables a more nuanced understanding of node importance as it evolves, capturing both the temporal aspect of the network and the relative importance of nodes as the network structure changes over time.

Beyond modified static CMs for dynamic network data, there are also various novel dynamic CMs that were conceived with dynamic networks in mind from the outset:

- **Change Centrality:** Change centrality is a metric designed for dynamic networks, which evaluates how much a node's connections change between two points in time. It measures the extent to which a node's neighborhood is altered over time, capturing both the appearance and disappearance of links. Unlike static CMs, change centrality focuses on temporal variations, offering insights into network dynamics, such as detecting shifts in the network's structure [FPA+12].

- **ICC & OCC:** *outgoing contact chain* (OCC) and *ingoing contact chain* (ICC) are dynamic CMs that capture how a node's connections change over time. While the *OCC* measures how a node reaches other nodes through direct and indirect outgoing links, the *ICC* captures how a node is connected by incoming links over time. These metrics provide a more accurate assessment of a node's centrality in dynamic systems where the timing and order of interactions are of interest, such as in information diffusion, like epidemic spread [SFS+19].

CHAPTER 3

Literature Review

Focusing on the visualization of centrality metrics in the social sciences and digital humanities, this section aims to review relevant literature on both static and dynamic networks. To the best of our knowledge, no comprehensive research has been conducted on how CMs can be used to enhance visual analytics or on the most effective methods for visualizing these metrics in social network analysis. Consequently, this literature review will address this gap by conducting foundational research. To offer different perspectives on the state-of-the-art usage of centrality metrics, the review is divided into the following two parts:

1. **Introduction to CMs in Social Networks** Social networks span a wide array of fields, with CMs applied in various ways to visualize and analyze these networks, from traffic flow in urban planning to flood control in environmental studies. This review begins with a paper-centric approach, involving an examination of research efforts across different contexts to lay a foundation for exploring how CMs are utilized in various domains. By focusing on individual studies, we aim to identify both the nuances and commonalities in the application of these metrics. Accordingly, this section will summarize key papers and discuss how CMs are employed to convey insights, illustrating their broad and deep application potential.

2. **Review of CMs in Dynamic Social Networks** The second part of this review aims to conduct a comprehensive, state-of-the-art analysis of how centrality metrics are visualized in the study of dynamic social networks. Unlike the previous section, this part of the literature review will adopt a methodological approach to examine contemporary research efforts to identify the most frequently used visualization methods and analyze how these approaches effectively convey insights into dynamic networks. By applying this methodology, we will not only map out the current landscape but also assess the effectiveness of different visualization techniques.

3.1 Introduction to CMs in Social Networks

This section adopts a paper-centric approach to explore how CMs are applied and visualized across diverse contexts. The review is based on a search conducted on ProQuest[1] using the following keywords:

"CMs" & "Network" & "Visualization"

We summarize the top three papers listed in the search results, sorted by relevance, to provide an initial and broad perspective on the methods used to visualize centrality metrics and convey insights into social network data.

3.1.1 Summarized Papers

Peechapat et al. investigate the power dynamics within Thailand's music industry in an effort to "improving efficiency and equity in the digital content industries" [PP24, p. 1]. To obtain interesting insights from the network data, the study uses CMs and employs various, different visualization methods to convey the underlying power dynamics. Initially, a series of large node-link diagrams (see Figure 3.1) plot each actor, where the centralities are encoded by node size, and color identifies the type of actor.

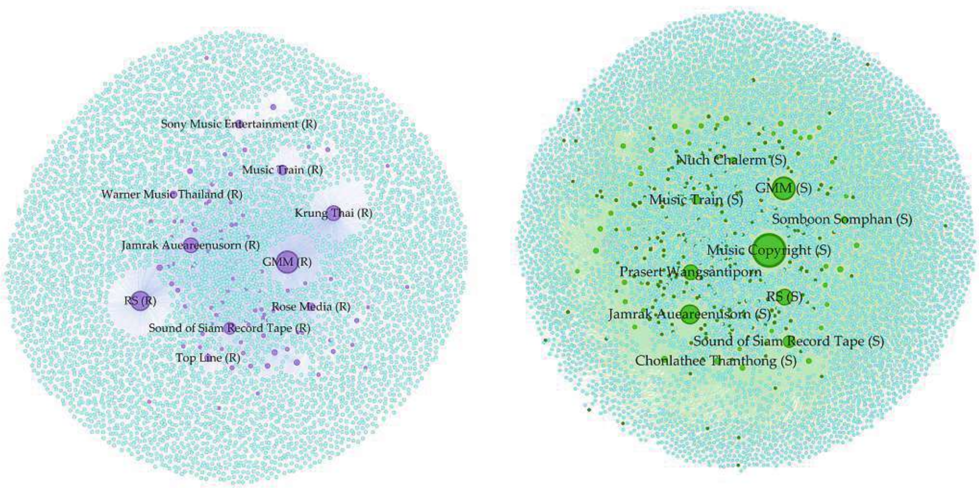

Figure 3.1: Large node-link diagrams plotting different actor groups of Thai music industry [PP24].

These node-link diagrams illustrate changing dynamics by plotting varying music owner groups against artists. The study notes that, due to the large number of actors, this can obscure the visibility of central figures. Thus, to focus more on the individual connections, the authors filter out less connected nodes for a more focused analysis of the relationships

[1]ProQuest: proquest.com. Accessed on May 21, 2024.

between the actors with the highest centrality and therefore limit themselves to actors with the greatest number of connections (see Figure 3.2). These graphics highlight the most important connections between actors and various music owner groups for deeper analysis.

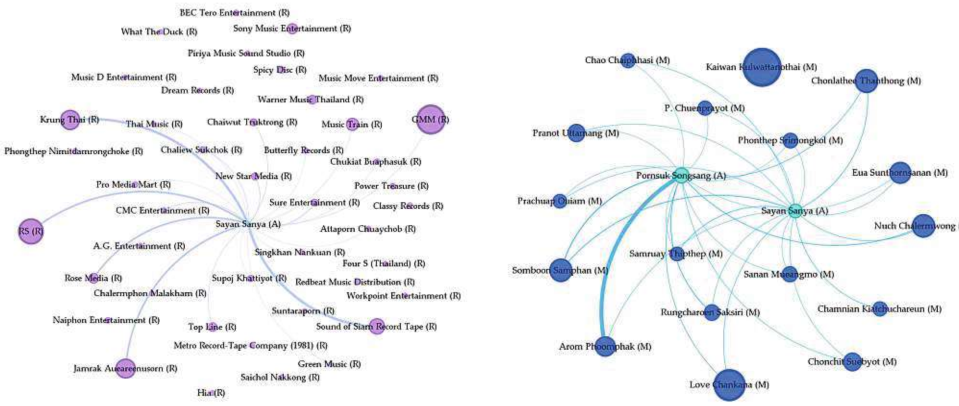

Figure 3.2: Focused node-link diagrams plotting different actor groups of Thai music industry [PP24].

Lastly, the authors use line charts (see Figure 3.3) to plot the degree of connections for each music owner group on a logarithmic scale. This method is noted to convey the varying distribution of centrality of the owners and artists.

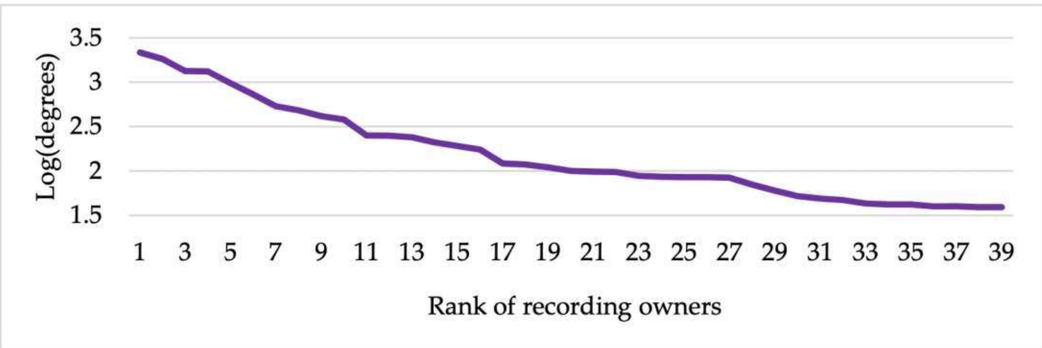

Figure 3.3: Line chart plotting degree centrality distribution of different actor groups of Thai music industry [PP24].

This visualizes the distribution of influence across the network, highlighting how a few actors hold significant centrality compared to many much less connected nodes, ultimately confirming that a few key actors concentrate most of the collaboration between powerful players.

In an effort to combat floods, a crucial part of flood risk governance pertains to diverse actors working together effectively. An integral aspect of this involves understanding the governing dynamics among different actors involved in flood risk governance.

To better understand the underlying power dynamics, the study “Social network analysis of EU flood risk management plans: Case Finland” [BEPK24] utilizes a directed multi-mode node-link diagram (see Figure 3.4) to represent various actors of different groups.

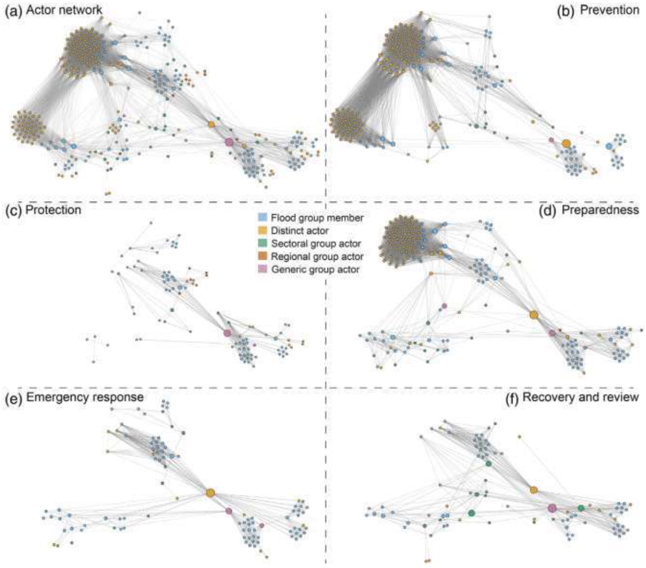

Figure 3.4: Comparison of different categories responsible for flood risk governance [BEPK24].

In this visualization, node size corresponds to degree centrality, and edges are colored based on the source node. This encoding highlights the connections and influence of each actor within the network. Importantly, the paper categorizes flood prevention measures into five main groups: prevention, protection, preparedness, emergency response, and recovery and review. Different actors are assigned to these categories based on their role in flood risk governance. Banafa et al. [BEPK24] visualize the varying centralities within these subcategories by creating side-by-side graphs for each category to observe changes in importance given the group. To ensure comparability, the positions of the nodes for the actor are kept consistent across each graph, preserving the viewer’s mental map [ELMS91].

Lastly, the study displays a scatterplot (see Figure 3.5) exhibiting multiple encodings of information to offer a comprehensive view of how actors are positioned relative to each other based on various factors: The x-axis represents the count of risk management measures, the y-axis shows weighted degree centrality, and the marker size indicates normalized betweenness centrality.

Lee et al. [LS23] investigate whether areas with high centrality based on commuting flows are adequately covered with point of interest (POI) facilities. The study conducts a correlation analysis between centrality metrics (degree and eigenvector centrality) and

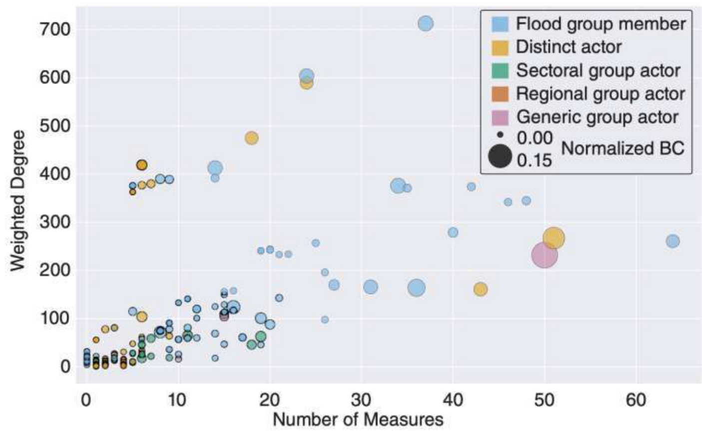

Figure 3.5: Scatterplot plotting degree centrality and betweenness centrality simultaneously [BEPK24].

POI coverage to determine if Seoul's areas have sufficient urban facilities to meet the needs of their populations. Degree centrality and eigenvector centrality are used to quantify the significance and centrality of different regions based on commuting data.

By correlating these CMs with the distribution and density of POI facilities, the study offers valuable insights for facility management and urban planning, particularly in rapidly growing and dynamically changing urban environments. The centralities are conveyed on a spatial map (see Figure 3.6) of the Seoul metropolitan area, with high levels of centrality encoded in a darker color tone.

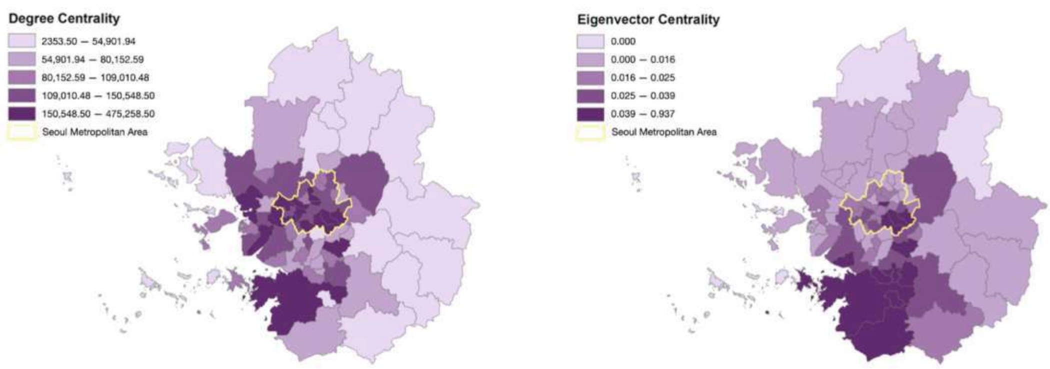

Figure 3.6: Side-by-side comparison of degree and eigenvector centrality indicating regional significance for commuting flows [LS23].

3.2 Review of CMs in Dynamic Social Network

This section methodologically reviews how centrality metrics are visualized and applied in dynamic social networks. By examining current literature, it highlights the methods and techniques used to visually convey evolving insights from network data.

3.2.1 Methodology

In selecting an appropriate methodology for conducting this literature review, a systematic literature review (SLR) was initially considered. The SLR method is an extensively used research methodology and is commonly used in scholarly research. However, as Okoli [Oko15] notes, the rigorous requirements of an SLR entail a significant time investment due to their comprehensive nature. A properly conducted SLR averages around 1,100 hours [PR08] and thus clearly exceeds the practical constraints of a master's thesis. Okoli also emphasizes that most thesis projects do not require the high standards of rigor typical of an SLR [Oko15]. Consequently, an SLR was deemed impractical for this research effort.

Selected Methodology

Petticrew et al. [PR08] explain that a state-of-the-art review is an effective way to bring readers up to date on the most recent research, without the exhaustive demands of an SLR. Consequently, a state-of-the-art literature review was chosen as a suitable alternative, providing an overview of current practices while fitting within the constraints of a master's thesis.

In an effort to balance thoroughness and feasibility, this thesis employs key elements of Okoli's proposed methodology for an SLR [Oko15] as part of its state-of-the-art literature review to ensure a structured approach and reproducibility of its outcome. The modified approach entails the following steps:

1. **Identify the purpose:** Clearly stating research objectives to guide the review.
2. **Search for literature:** Employing a detailed, reproducible search strategy for literature.
3. **Apply practical screen:** Categorize literature by relevance.
4. **Extract data:** Extract specific visualization method used to convey CMs.
5. **Study Synthesis:** Analyzing data quantitatively and qualitatively and synthesizing the studies.

This approach will serve as the framework for the state-of-the-art literature review on the utilization of visualization methods for CMs in social networks.

Applied Methodology

The previously outlined modified literature review methodology is implemented in the following way:

1. **Identify the purpose:** The objective of this review is to systematically analyze state-of-the-art literature focusing on how centrality metrics are visualized and applied in dynamic social networks.

2. **Search for literature:** A comprehensive literature search was conducted using the ProQuest database. The following keywords were chosen to align with the research objective:

   "CMs" & "Dynamic Network" & "Visualization"

   The search was further refined by applying several filters to ensure relevance and quality:

   - **Peer-reviewed articles:** To ensure the inclusion of academically rigorous studies.
   - **Time frame:** Limited to the last 10 years to capture the most recent advancements and trends in the field.
   - **Subject area:** Only include research with the subject area "social networks" to align with the objective of the review.
   - **Language:** Restricted to English-language publications to ensure comprehension and consistency in analysis.

3. **Apply practical screen:** After the initial search, a manual screening process was employed to categorize the literature by relevance. The screening involved evaluating each paper based on the following criteria:

   - **Relevance to Social Dynamic Network Data:** Only papers that analyzed dynamic social networks were considered.
   - **Use of CMs:** Papers that did not employ CMs were excluded.
   - **Visualization of Metrics:** Only studies that visualized the centrality metrics were retained. Papers that discussed CMs without accompanying visualizations were deemed irrelevant.

4. **Extract Data:** For each paper that passed the relevance screen, a detailed data extraction process was implemented. This involved manually reviewing the selected studies to identify and extract the following information:

   - **Visualization Methods:** The specific techniques used to visualize the CMs were noted.

- **Insights Conveyed:** The key insights that the visualization aimed to communicate were extracted, focusing on how the visualization method helped in understanding the significance of the nodes within the network.

5. **Study Synthesis:** The final step involved synthesizing the extracted data through both quantitative and qualitative analysis in a discussion concluding the review:

   - **Quantitative Analysis:** Assess the frequency and distribution of various visualization methods that were used.
   - **Qualitative Analysis:** Discuss the methods and strategies that were employed to convey the insights and temporal aspects of social network data.

Following the "Search for Literature" step, a total of 46 papers were identified. After applying the "Practical Screen" step, this number was narrowed down to 21 papers. From these selected papers, methods of visualizing centrality metrics were categorized into a static or dynamic visualization technique category. Although all the reviewed network data is dynamic, in this context, "static visualizations" refer to methods that represent the data without depicting temporal changes within the diagram itself. In contrast, "dynamic visualizations" explicitly convey temporal changes within the figure, allowing viewers to observe how relationships and centrality metrics evolve over time.

The following sections provide an overview of the different visualization methods identified as part of the "Extract Data" step:

3.2.2 Static Visualizations

Node-Link

Node-Link diagrams are used to convey the relationships and differences in centrality between actors in a network. Centrality is communicated through node size [SZC18, SFTN18, LV20] and color [SZC18, SFTN18, SFS+19]. While [SZC18, SFTN18] express differences in centrality with different colors along a spectrum, [SFS+19] takes a different approach and highlights only the most important nodes, leaving all other nodes grey. Additionally, [SFS+19] marks all links of important nodes as well.

[SFTN18] simultaneously conveys multiple centralities by using double encodings of nodes: the size of a node might signify degree centrality while its color indicates betweenness centrality. In contrast, [SFS+19] takes a different approach and encodes centrality simultaneously in both size and color. In [LV20], color is used to encode additional information for comparison, so while node size represents centrality, color provides important details about the actors. Furthermore, the thickness of the links is used to indicate the intensity of the relationships between actors in [SFTN18, LV20].

While [SFTN18, LV20] apply a force-directed algorithm to position the nodes, [SZC18] employs an ego network that positions the nodes based on their actual geographic location.

Scatterplot

To convey the distribution of degree centrality, [SFS+19, MRBC19, SBPS23] use scatterplots to plot the frequencies for different degree centralities, with the x-axis representing degree centrality and the y-axis representing frequency. Additionally, [SFS+19, MRBC19] employ a log-log while [SBPS23] uses both a linear scale and a log-log scale. To compare different distributions, [SFS+19] plots various centralities simultaneously, allowing for a comparison of their distributions, whereas [SBPS23] uses multiple small side-by-side scatterplot diagrams. Moreover, [MRBC19] examines the degree distribution from a time-integrated perspective by accumulating all the connections that occur over a certain period and comparing the static vs. time-integrated distribution within the same plot.

Beyond degree centrality distribution, [SFS+19, AIP+18] employ scatterplots to measure the relationship dynamics between different CMs for nodes in a network. [AIP+18] uses traditional scatterplots to illustrate the relationship between different CMs for each node. They plot various side-by-side scatterplot diagrams to explore how variations in network structure and different settings relevant to the nodes affect their centrality and alter the dynamics of the relationship. [SFS+19] also displays the dynamics between two centralities but uses a variation of the scatterplot that groups overlapping points into bubbles, labeling each bubble with the frequency of nodes.

Boxplot

[LV20, JC24] use boxplots to show the distribution of individual actors by plotting every actor along the axis of the boxplot. Both studies employ a side-by-side comparison between different CMs. Furthermore, [JC24] compares different groups of metrics for a given centrality measure. Both visualizations from [LV20, JC24] show significant outliers, and [LV20] address this by providing two different approaches to displaying data with significant outliers: cutting off extreme outliers and using a log-transformed distribution to fit all actors into the boxplot. In contrast, [JC24] plots the values unmodified, resulting in extremely narrow boxes.

Table

Tables are used to compare centralities in various ways. While [SZC18, TS19] compare actors in a network based on various centralities such as betweenness centrality, [AIP+18, YHC+22] further subcategorize CMs into more granular variations of the measure, such as sociocentric and egocentric centrality. Regarding the ordering of actors being compared, [TS19] orders actors alphabetically, whereas [SZC18] ranks them based on centrality values in descending order. A different approach is taken by [SFTN18], who group actors of the network into categories and track changes over the course of the study by assessing the number of actors whose individual centralities improved or worsened within a group. Lastly, various studies use aggregated centralities to describe the nodes' importance. [SFTN18] use averages from various CMs to assess the significance of each group. In contrast, [YZGW22] calculates the mean of actors for a given time period and displays the standard deviation alongside the minimum and maximum values for actors within the network. Meanwhile, [SFS+19] characterizes the network's actors over a time period by depicting their centralities using statistical measures such as min, max, median, and mean.

Other Methods

[SFTN18] uses a radial diagram to convey centrality, where the closer a node is to the center, the more central it is, and nodes that are on the same orbit have equal centrality.

3.2.3 Dynamic Visualizations

Node-Link

Node-link visualizations of CMs incorporate time in two distinct ways. The first method involves aggregating time into static values [SZC18, SSM+23], such as summarizing data over a two-month period [SSM+23]. The second method conveys time through comparative snapshots at different points. This can be achieved by presenting small node-link diagrams side-by-side [AIP+18, YZGW22] or by providing more detailed snapshots that span across multiple pages, offering extensive context to guide the reader through varying scales of analysis [TEWG23]. In [YZGW22], the researchers enhance these time snapshots by adding new nodes in each successive diagram, illustrating how the network evolves. They assign numbers to nodes to indicate the order of appearance, helping identify new nodes appearing over time.

Encoding in node-link visualizations is primarily achieved through node size and label size. The magnitude of centralities is often represented by node size [TEWG23, YZGW22, SZC18] or by the color of the nodes [SZC18]. Alternatively, the label size is also used to encode these metrics [YZGW22]. In some cases, multiple CMs are encoded within a single graph; for example, [YZGW22] visualizations display degree centrality with node size and betweenness centrality with label size. In [SZC18], the magnitude of centrality metrics is conveyed through node size as well as color simultaneously to emphasize varying magnitudes. In [TEWG23], nodes are colored red or blue to indicate whether the centrality is above or below average, respectively. To indicate the weight of interactions, [TEWG23] adjusts the thickness of the edges between nodes, using varying line weights to represent the strength of connections.

Node positioning in node-link diagrams employs several strategies. [YZGW22] uses an unordered layout employing a packing algorithm to prevent overlapping nodes. [AIP+18, SZC18] position nodes on top of a map, aligning them according to the geographical locations of the represented entities. [TEWG23] uses a radial layout that enhances the clarity of relationships within the node-link sociogram. The nodes in the node-link diagrams used by [SSM+23] are fixed in reserved spots on the radial layout, making it easier to track changes across temporal snapshots.

The number of nodes displayed in a node-link diagram varies depending on the purpose. For example, [YZGW22] begins with fewer than ten nodes and incrementally adds more nodes in each subsequent temporal snapshot, eventually reaching hundreds, to illustrate trends. Conversely, [YMC21] limits displayed nodes based on their degree values, filtering out less important nodes to maintain focus on the most significant ones.

Line-Plot

Line plots track the evolution of centrality within a network by displaying temporal changes along the x-axis. To enhance the analysis, a second line is added to the plot, facilitating direct comparisons between different centrality metrics [TEWG23].

Table

Dynamic data is conveyed in tables by highlighting temporal development through comparisons of different snapshots of time, as demonstrated in works like [SBPS23, MHZ23, TS19, TEWG23, YMC21]. To structure dynamic data for comparison, various methods are utilized. For instance, comparing different CMs using summary statistics (min, max, median, mean, 25%, and 75%) as seen in [SFS+19, YZGW22, TS19]. Additionally, listing significant actors in descending order, either by their name, as in [MHZ23, LV20, TS19], or by centrality, as in [MHZ23, YMC21], or presenting unordered lists as in [SBPS23], are common practices within dynamic tables. Furthermore, dynamic tables are used by [TEWG23] to provide temporal context to other figures and display centrality metrics over time for other figures that don't express time.

3.3 Discussion

This section reviews the various visualizations found in the literature and their applications. It discusses why and how these methods were used, highlighting their effectiveness in analyzing and interpreting various aspects of presenting dynamic social network data.

3.3.1 Temporal Visualization

The following table illustrates the distribution of visualization methods identified across 21 reviewed papers. The heatmap provides an overview of the frequency with which each method was employed and its classification as either static or dynamic. From the analysis, seven distinct visualization methods were identified. The static methods dominate, with tables being the most frequently used (15 instances), node-link diagrams (9 instances across both static and dynamic categories), and followed by scatterplots (6 instances) (see Table 3.1).

Notably, despite the focus on dynamic social network data, a significant majority of the used methods are "static" because they do not incorporate a temporal component. This finding reveals an interesting trend: even when dealing with dynamic data, many studies choose methods that present the information in a fixed, static state.

Table 3.1: Frequency of visualization methods used in 21 reviewed papers, categorized by static and dynamic techniques. Darker cells indicate more frequently used methods.

Type	Node Link	Scatter Plot	Line Plot	Box Plot	Bar Chart	Table	Spatial Map	Other	Total
Static	4	6	4	2	3	11	2	1	33
Dynamic	5	0	1	0	1	4	0	0	11
Total	9	6	5	2	4	15	2	1	44

Static Visualization

Even though these visualization methods do not visually convey the passage of time, they are still utilized in ways that effectively communicate dynamic network data. By leveraging static methods in innovative ways, researchers can present time-related insights without the need for fully dynamic visualizations. The following list contains observed approaches that illustrate how static methods can be adapted to capture and convey temporal dynamics:

- **Temporal Static Values:** This approach involves either aggregating temporal data into static values that can be communicated through static methods or calculating temporal metrics that yield a single, static value representing time. For instance, as demonstrated in [SSM⁺23], temporal data from a period of several months is aggregated into static values, which can then be represented in a node-link diagram. This effectively condenses temporal changes into a static visualization, providing a overview of time-related trends without the complexity of dynamic visualization. In [SFS⁺19], the metrics are calculated to aggregate temporal aspects into a single static value, making it easier to represent and interpret the temporal evolution of network centrality.
- **Augmenting Static Data with Temporal Context:** Another strategy involves using static data from a specific time snapshot and enhancing it with additional figures, such as tables, to provide temporal context. For instance, [YMC21] employs a static snapshot of network data while incorporating numerous tables that add temporal context.

Dynamic Visualization

Dynamic visualization methods explicitly show the passage of time, illustrating how data evolves. The literature review identified three main approaches for integrating temporal changes into visualizations:

- **Single Diagram with Temporal Data:** This method integrates data from different time points within a single diagram, allowing for a direct comparison of how metrics change over time. It is particularly effective in helping viewers build a mental map of the data, facilitating comparisons of values across different periods. A frequent and traditional variant of this method includes tables that encode time for various snapshots, offering a straightforward way to track changes. However, the challenge with this approach lies in the limited space available within a single diagram; including multiple time points can lead to clutter, which may compromise clarity, especially with complex datasets [SBPS23, TEWG23, YMC21, MHZ23].

- **Side-by-Side Mini Diagrams:** This approach involves displaying multiple mini diagrams within a single figure, with each diagram representing a different point in time. Arranged side-by-side, these diagrams allow readers to observe temporal changes while keeping each mini diagram uncluttered, as each time point has its own dedicated space. However, the physical separation of diagrams can make it challenging to compare exact changes between time points, requiring the reader to mentally track the progression across the figures. An example of this approach is found in [AIP+18, YZGW22], where a geographic view of a node-link network is paired with small side-by-side representations of centrality changes across three different time points. This method effectively captures network dynamics over time, offering clear snapshots for each moment but requiring effort to track the overall progression.

- **Numerous Full-Size Graphs:** This method employs multiple full-size graphs, with each graph representing a different time point. It is often used when showing significant details and giving context is crucial, as full-size graphs allow for a more detailed exploration of the data at each time point. This approach provides significantly more context within each figure, as demonstrated in [TEWG23, SSM+23], where this approach was used to explore complex network dynamics in depth. However, a key challenge is that readers may find it difficult to stay on top of changes across different time points.

Each of these dynamic visualization approaches has its strengths and trade-offs, and the choice of method should be tailored to the specific needs of the analysis. For data requiring precise comparisons and strong mental mapping, a single diagram that incorporates temporal data may be the most effective, as it minimizes change blindness [ELMS91] and allows for easy comparisons. In contrast, side-by-side mini diagrams offer a compromise with a well-organized and uncluttered presentation but with an increased risk of change blindness when trying to track changes across diagrams. Lastly, conveying time with full-size graphs is able to provide significant context but also poses the largest risk of change blindness as they are often spread across different pages, potentially overwhelming the reader with disjointed temporal information. Ultimately, the selection of a dynamic visualization method should balance these trade-offs, aiming to best convey the dynamic nature of the data while meeting the specific objectives of the analysis.

3.3.2 Methods Use Cases

- **Node Link:** Node-link diagrams are frequently employed as a versatile method for visualizing social networks. They provide a clear and intuitive representation of nodes and edges, making them a popular choice for analyzing the structure and dynamics of networks [CM11]. This translates to their ability to intuitively represent relationships and centrality metrics. Their effectiveness arguably stems from the diverse range of encoding techniques available, which can be strategically selected and applied based on the specific analytical needs of a study. When the goal is to highlight the most central actors within a network and their key relationships, researchers often choose this method and encode centrality using node size, as it provides an immediate visual cue about the importance of each actor. This approach, prevalent in studies like [SZC18, SFTN18], uses larger node sizes to indicate higher centrality, making it easy to identify key players at a glance.

  Color serves as another powerful encoding tool, often used in conjunction with node size to provide a dual layer of information. This method is particularly effective when multiple centrality metrics need to be conveyed simultaneously. For instance, in [SFTN18], node size might represent degree centrality, while color indicates betweenness centrality. This dual encoding allows for a more nuanced comparison of actors, helping researchers distinguish between nodes that are influential due to their connectivity versus those that play a critical intermediary role within the network.

  However, in scenarios where the primary objective is to emphasize only the most critical nodes and their connections, selective color usage becomes advantageous. The approach seen in [SFS+19], where important nodes are highlighted in color while less significant ones are grayed out, exemplifies this technique. Selective highlighting directs the viewer's attention to the most relevant parts of the network, making it particularly useful in large and complex networks where excessive information could otherwise overwhelm the user and color helps to clearly identify key actors.

- **Scatter Plot:** Scatterplots are predominantly utilized in the reviewed literature for visualizing and deriving insights related to distribution. For instance, the degree distribution of degree centrality illustrates how many nodes possess a given number of direct connections. Scatterplots excel in highlighting the network's structure by revealing whether most nodes have few connections or if there are hubs with many connections. This visualization helps in identifying whether the network follows a power-law distribution, a common feature in social networks that indicates the presence of a few influential actors. For example, in [SBPS23], the use of scatterplots demonstrated that the network followed a power-law distribution, underscoring the presence of highly connected nodes that serve as critical hubs. To enhance the clarity of these patterns, scatterplots are often employed with log-log scales[YZGW22]. This logarithmic scaling facilitates visualizing whether the network adheres to a power-law distribution, as it compresses the scale to make patterns more discernible. For instance, in [SFS+19], the log-log scatterplot provides a more accurate representation of the degree centrality distribution.

Beyond the analysis of degree distribution, scatterplots are also instrumental in comparing actors based on two different CMs plotted on the axes. For example, in [SFS+19], scatterplots were used to compare two CMs side by side, offering an understanding of how these metrics interact and vary among actors.

- **Table:** Tables are a very versatile tool for comparing centralities in network analysis. They allow for the comparison of actors based on various CMs, as seen in studies like [SZC18, TS19], where actors are compared using measures such as betweenness centrality. Some studies, like [AIP+18, YHC+22], further break down centralities into more granular subcategories of individual centralities, such as sociocentric and egocentric centrality. Tables are also an effective means to convey dynamic data by highlighting temporal developments. Studies like [SBPS23, MHZ23, TS19, TEWG23, YMC21] use tables to compare different time snapshots, often utilizing summary statistics, like those in [SFS+19, YZGW22, TS19]. These tables can also list significant actors by name or centrality, as seen in [MHZ23, LV20, TS19], or present unordered lists, as in [SBPS23].

3.3.3 Visualization Approaches

Conveying the information contained within centrality metrics is a complex and multifaceted task, with no one-size-fits-all solution for visualizing these metrics. This complexity arises because different visualization methods each have their own strengths and limitations, and no single method can adequately capture all the nuances of network metrics. For instance, while listing metrics in a table may provide an organized view of raw data, it often fails to reveal deeper patterns and relationships. Similarly, a single graphic, such as a node-link diagram, may effectively display the structure of a network but may not succeed in conveying more complex insights or patterns inherent in the data.

This point is supported by the reviewed literature, which shows that, on average, $\tilde{2}.1$ different visualization methods are used per research effort to convey centrality metrics (see Table 3.1). This suggests that researchers often employ a variety of methods to effectively communicate the information within centrality metrics.

From the 21 reviewed papers, three approaches have been identified for how figures are commonly used to convey complex insights (see Figure 3.7):

- Holistic Approach, which integrates multiple methods to provide a comprehensive view
- Complementary Approach, where different methods are used in conjunction to highlight various aspects of the data
- Different Focus Approach, where each method emphasizes distinct facets of the network metrics

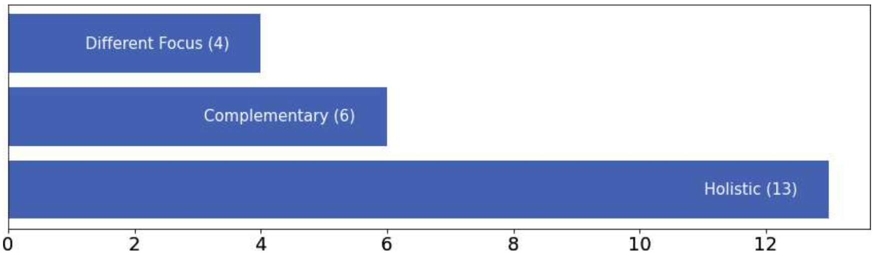

Figure 3.7: Bar Chart Illustrating the Frequency of Visualization Approaches Used in the Reviewed Literature for Conveying Centrality Metrics

Importantly, these approaches are not mutually exclusive, with various papers employing multiple combinations of these approaches to effectively convey insights. The remainder of this section elaborates on how the reviewed studies utilized these approaches in different scenarios as outlined in the following paragraphs:

- **The holistic approach:** The holistic approach involves using multiple visualization methods in a complementary manner to effectively communicate insights from social network data. This approach leverages the strengths of different visualizations to provide a comprehensive understanding of the network's dynamics from various perspectives.

  For instance, [SFS+19] employed this approach by first presenting table data to provide an overview of the data by displaying aggregated values such as the median and mean of the network's centralities. Using a table provides a clear and organized way to present raw numerical data, allowing for easy comparison and identification of general trends within the network. Next, a large node-link diagram was used to visually highlight hubs, or nodes with significant importance within the network. This type of visualization effectively demonstrates the network's relationship structure and helps to identify key actors who play crucial roles within the network. Additionally, a scatterplot was employed to show the frequency distribution of degree centrality within the network. This visualization conveyed an understanding of how centrality is distributed across the network, indicating whether most actors have similar centrality values or if there are significant disparities. Finally, the relationship between two different CMs was plotted on another scatterplot. This allows for the exploration of correlations between these metrics, helping to reveal how different aspects of centrality relate to each other within the network.

  By combining these various methods, the holistic approach accomplishes what a single method could not: it provides a multi-dimensional view of the data, uncovering insights from various angles that might otherwise remain hidden. This comprehensive analysis is conducive to communicating complex social dynamics and power structures within networks. This approach is also the most frequently used, with nearly 90% of all papers employing a combination of different visualization methods.

- **Complementary Approach:** The complementary approach involves using two visualizations side by side, where each method either complements the other or offers an additional perspective, resulting in a more nuanced and thorough understanding of the data. This approach is particularly effective in cases where a single visualization might miss important aspects or fail to capture the full complexity of the network being studied.

  For instance, in one study [JC24], researchers used two box plots placed next to each other, each representing a different CM. This method allowed for the comparison of the same dataset visualized through different lenses, offering a deeper understanding of the distribution and variance within the CMs. Moreover, [SZC18] illustrates the usefulness of the complementary approach by pairing a node-link diagram, which visualizes the power dynamics within a social network, with a geographic map that shows where the nodes from the node-link diagram are spatially located. The node-link diagram effectively demonstrates the structure of power relationships, identifying key influencers within the network. Meanwhile, the accompanying map provides a geographic context, showing how these power dynamics are distributed across physical space. This combination enables a more comprehensive understanding of how social influence is geographically distributed, offering insights that neither visualization could provide alone.

- **Different Focus Approach:** The different focus approach utilizes a single visualization method applied with varying focal points, such as different filters or scenarios, to highlight how specific dynamics within the network change under different conditions. This approach is particularly effective for examining how certain variables impact the network, offering insights into the shifting power dynamics.

  For example, in [SIA23], the researchers analyzed the changing betweenness centrality of actors within a network under different scenarios. By applying different filters to the same visualization, they were able to observe how the importance of certain nodes changed depending on the scenario. This approach effectively demonstrated how the centrality of various actors fluctuated, providing a clearer understanding of the factors that drive these changes. The different focus approach is particularly useful for understanding how changes in a variable affect network dynamics.

CHAPTER 4

dome-insights - Design

The primary goal of this thesis is to explore how a visual analytics approach can be improved by incorporating CMs to provide deeper insights and help identify key actors in dynamic networks. A critical part of addressing this objective involves the development of a visual analytics prototype. In this chapter, we introduce dome-insights[1], an interactive tool designed for visualizing and analyzing dynamic network data. This prototype builds upon the findings from our literature review, where we examined various techniques for applying CMs in dynamic contexts (see Section 3.3).

4.1 Application Structure

The *dome-insights* application is structured using a two-tier architecture of complementary methods. This design is inspired by Shneiderman's *Visual Information-Seeking Mantra*, "Overview first, zoom and filter, then details on demand," which emphasizes the importance of providing users with a high-level overview before allowing them to explore finer details [Shn03]. This mantra aligns with the strengths of CMs, which quantify each node's relationships into a single metric, arguably an abstraction or overview of the node's relationships. Thereby simplifying the network, CMs are very suitable for building the foundation of an overview for even complex networks. Furthermore, CMs are also useful for identifying significant relationships as part of a more detailed view, allowing users to focus on central co-exhibitors from an artist. Consequently, the two-tier architecture (see Figure 4.1) exhibits the following views:

Overview and Ego-View

1. ***Overview*:** This view enables users to explore the full dataset, identifying key figures and broader trends. Artists are displayed as nodes, with sizes proportional

[1]dome-insights: `github.com/tridelt/dome-insights`. Accessed on October 13, 2024.

to their centrality scores, providing a visual representation of centrality. Color coding adds additional information in the form of the artists' nationalities. While this view gives a high-level "overview first" of the network, to reveal details, the user can simply left-click on any node, and it automatically opens its *Ego-View*.

2. ***Ego-View*:** In contrast, the *Ego-View* provides "details on demand" on an artist's co-exhibition relationships. Given the network's complexity, we highlight only a node's most significant relationships as determined by an algorithm powered by CMs. Thanks to a sophisticated temporal visualization method employed as part of the *Ego-View*, users can explore these relationships and see how they evolve over time.

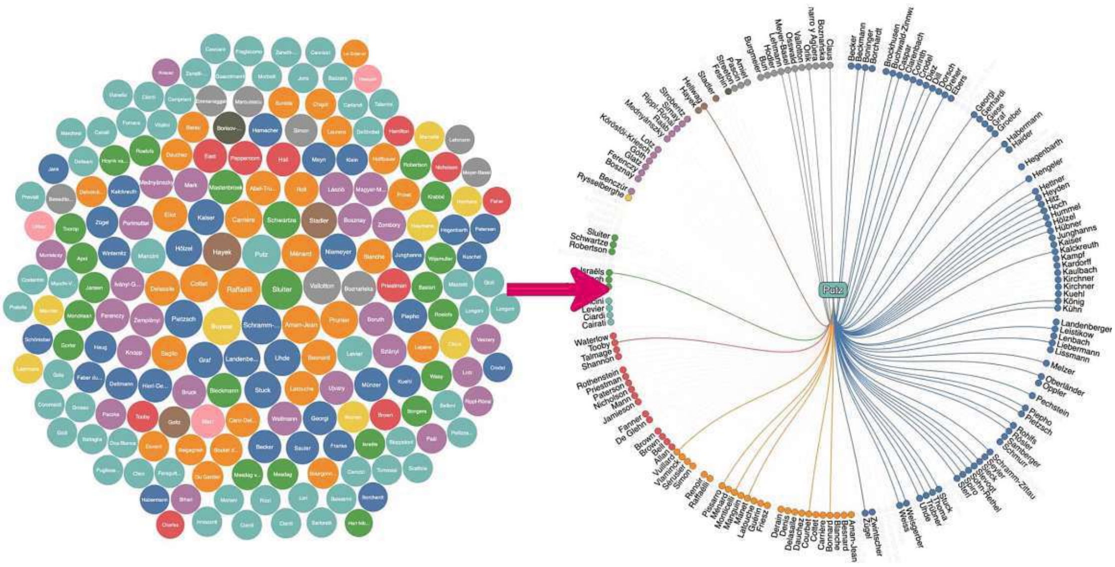

Figure 4.1: dome-insight's two-tier architecture, showing both the *Overview* (left) and *Ego-View* (right).

The following sections will elaborate on insights from the previous literature review chapter that influenced the design and visualization methods used as part of dome-insights:

4.2 Overview

A common method for visualizing centrality in network analysis is the node-link diagram, where node size represents centrality scores and links illustrate connections. This approach is widely used and effective for visually comparing centrality metrics [SZC18, SFTN18, LV20, SFS+19, SSM+23, AIP+18, YZGW22, TEWG23]. In our prototype's *Overview* view, we build on this method by omitting the link connections, relying solely on node size to represent an artist's centrality. By abstracting connections in this way, we simplify the visualization, particularly for large datasets, like ours, which include over 13,000 artists and nearly 5,000,000 co-exhibition relationships. This practice enables users to identify key figures without the distraction of visual clutter from plotting millions of relationships.

To further enhance the visualization, color encoding of the nodes is used, in the case of this dataset it encodes the country of birth of artists. The position of the nodes is determined by a layout algorithm, which positions nodes according to their centrality. Artists closest to the center are the most central, while those in outer layers have progressively lower centrality. This layered design offers an intuitive way to compare centrality by associating in addition to node-size also proximity to the center with importance (see Figure 4.2).

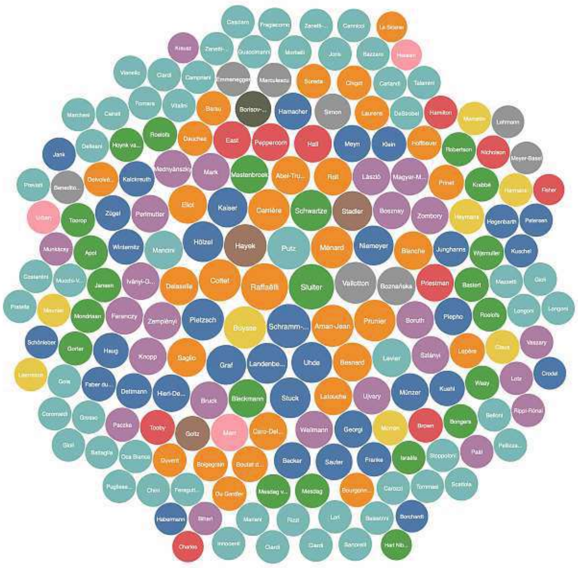

Figure 4.2: Bubble plot from the *Overview* view, where node size indicates the artist's centrality, and color represents the country of birth.

Combined with node size and color encoding, this layout offers a clean and concise view of the network, facilitating effective exploration and identification of significant figures. This layout not only emphasizes the most central figures but also naturally has the effect of grouping artists with similar centrality in concentric layers, creating a clear visual hierarchy and thus simplifying the detection of patterns relating to importance. The result is that nodes with similar centrality scores are positioned at a similar distance from the center.

For example, the patterns on the left side (see Figure 4.3) reveal the dominance of artists from Great Britain, thanks to the color encoding that highlights their prominence with distinct red-coded bubbles. This allows users to quickly identify the presence of different nationalities and compare the dominance of certain countries.

The radial layout on the right side (see Figure 4.3) exhibits clear layering. This layered structure effectively communicates varying levels of centrality, making it easy to interpret

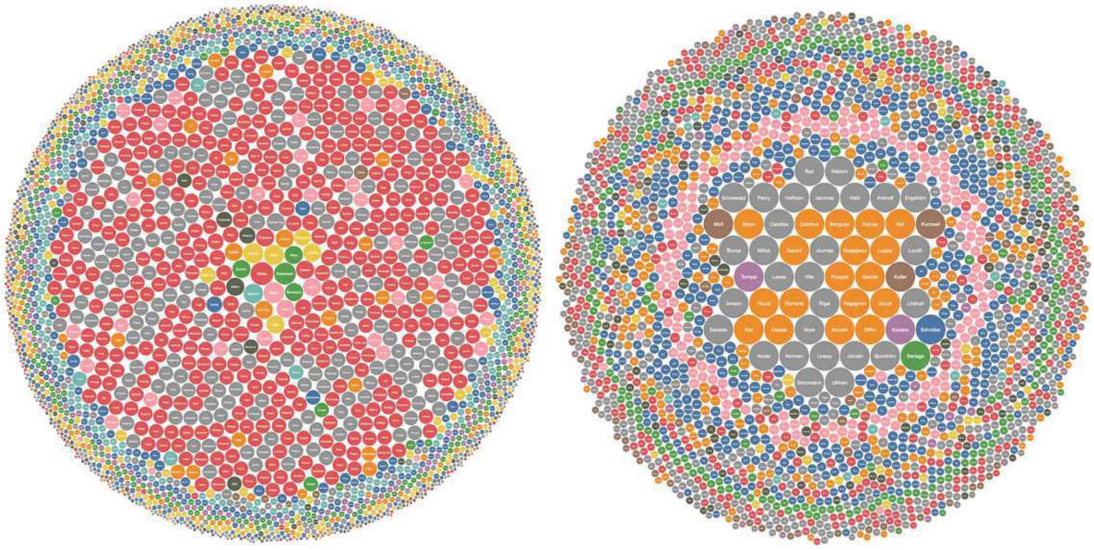

Figure 4.3: Two examples illustrating the different layering and patterns that emerge from this visualization method.

and compare the dominance of different nationalities within the network, with decreasing importance of layers as we move farther away from the center.

By combining node-size scaling with country-based color coding, the *Overview* simplifies the complexity of large datasets. This approach allows users to quickly identify influential figures and recognize broader patterns, making the *Overview* an effective first step in understanding network centrality without overwhelming the viewer with too much information.

4.2.1 Selection

In the reviewed literature, numerous research efforts present data from multiple perspectives to explore the network. This often involves comparing different CMs and segmenting data by factors like country or highlighting the top 10% most central nodes. Furthermore, for temporal network data, snapshots (i.e., timesliced networks) are also commonly used to demonstrate how networks evolve over time [BBDW17]. These diverse perspectives are crucial for providing a more nuanced understanding of network dynamics by uncovering different facets of the data.

Inspired by these aforementioned approaches from the literature, our prototype employs various selection tools that similarly also allow users to explore the network from multiple angles but empowered with interactive features. In the *Overview* users can switch between CMs, filter by country, or focus on specific time periods, enabling custom exploration. This dynamic system recreates and extends insights typically seen in static graphs from the reviewed papers but adds sophistication by offering real-time control. By interacting with the data directly, users can inspect countless perspectives and uncover patterns that would not be apparent from a single viewpoint (see Figure 4.4).

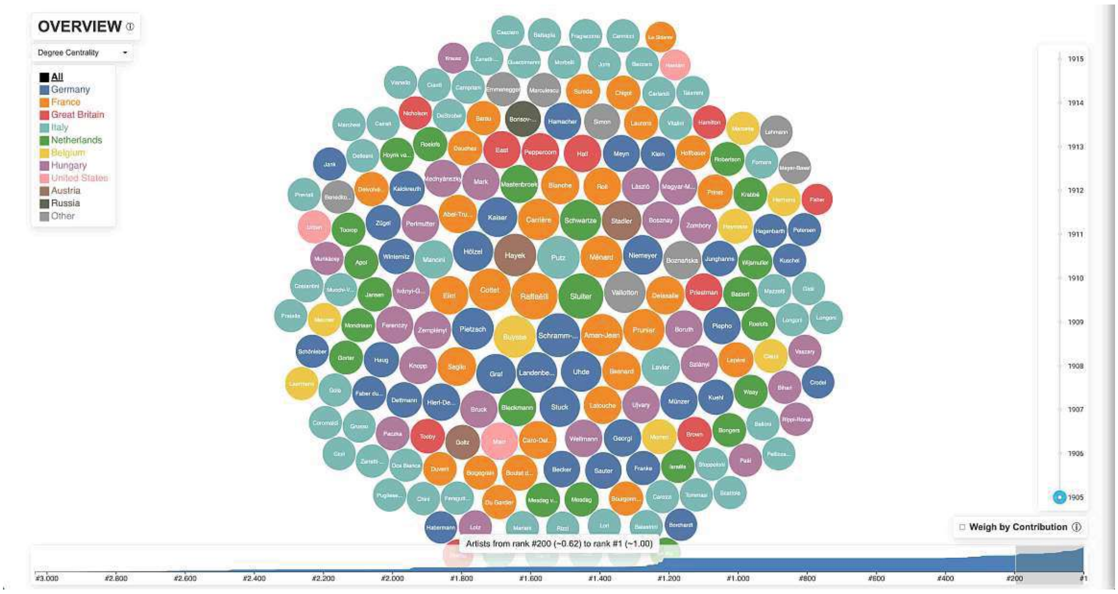

Figure 4.4: Various selection tools are positioned around the edges of the canvas, allowing users to explore the network.

CM Selection

One of the most frequently explored aspects in the reviewed network analysis literature is the use of various CMs to examine the relative importance of nodes within a network. For example, in [SSM$^+$23], centrality metrics were visualized using multiple node-link diagrams, with different graphs showing the same network but using distinct CMs. This allowed for a side-by-side comparison of centralities to identify which nodes were central across different CMs. Similarly, [FSIK19, LS23] employed node-link diagrams to visualize centrality metrics in separate diagrams, offering users a clear comparison between various forms of centrality. In contrast, [SBPS23, YZGW22] used tables to compare the three most central actors, presenting how their centrality rankings shifted when calculated using different CMs. In summary, using a combination of CMs, such as degree, betweenness, eigenvector, and closeness centrality, is common to reveal different facets of centrality within a network.

Building on this insight, our prototype employs a CM selection tool that allows users to dynamically switch between degree, betweenness, eigenvector, and closeness centrality. This flexibility enables users to explore various dimensions of the nodes' centralities and their differing scores depending on the chosen CM. Similarly to reviewed research efforts, this also illuminates the data from different perspectives but with real-time, interactive control (see Figure 4.5). A dropdown menu at the top left allows users to select from these four different CMs.

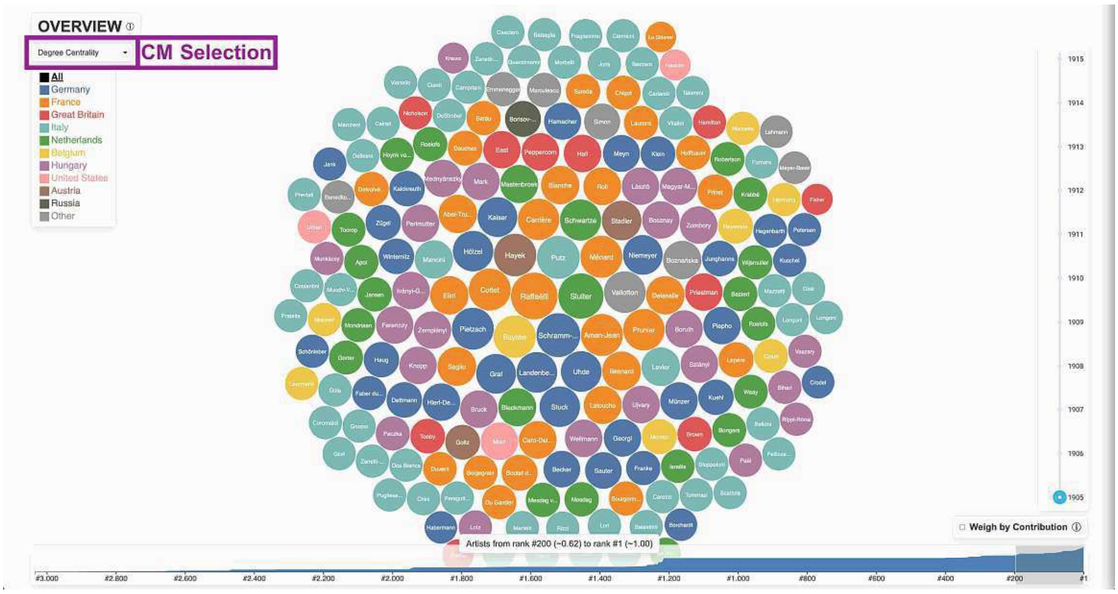

Figure 4.5: Dropdown menu for selecting CMs to explore different dimensions of artists' centralities. Currently, degree centrality is selected, and the node sizes on the canvas are calculated based on this currently selected CM.

Time Selection

Conveying temporal changes is crucial for understanding how dynamic network data evolves over time. In the literature, various approaches have been employed to capture these changes. For example, [SSM+23] visualizes the same node-link diagram at different points in time, offering snapshots that reveal how centralities have shifted over the years. [AIP+18] presents side-by-side snapshots of the same data at different moments, showing how the network evolves over time. In addition to graphical approaches, tables are often used to add temporal context to visualizations. Furthermore, dynamic data is frequently structured in tables that highlight temporal development through comparisons of different time snapshots, as demonstrated in works like [SBPS23, MHZ23, TS19, TEWG23].

Building on these insights, our prototype features a time selection tool that allows users to dynamically explore the network's evolution. By adding a vertical slider on the right side of the canvas, users can variably select a year to examine the artists' centralities during that point in time (see Figure 4.6).

By providing real-time exploration of temporal changes, our visual analytics tool enables a more interactive analysis across all available time points. Notably, when the year is changed, the displayed artists and their corresponding information are updated to reflect the selected year, while the selected country of birth and selected CM remain constant. For example, a user might be interested in exploring German artists through the lens of

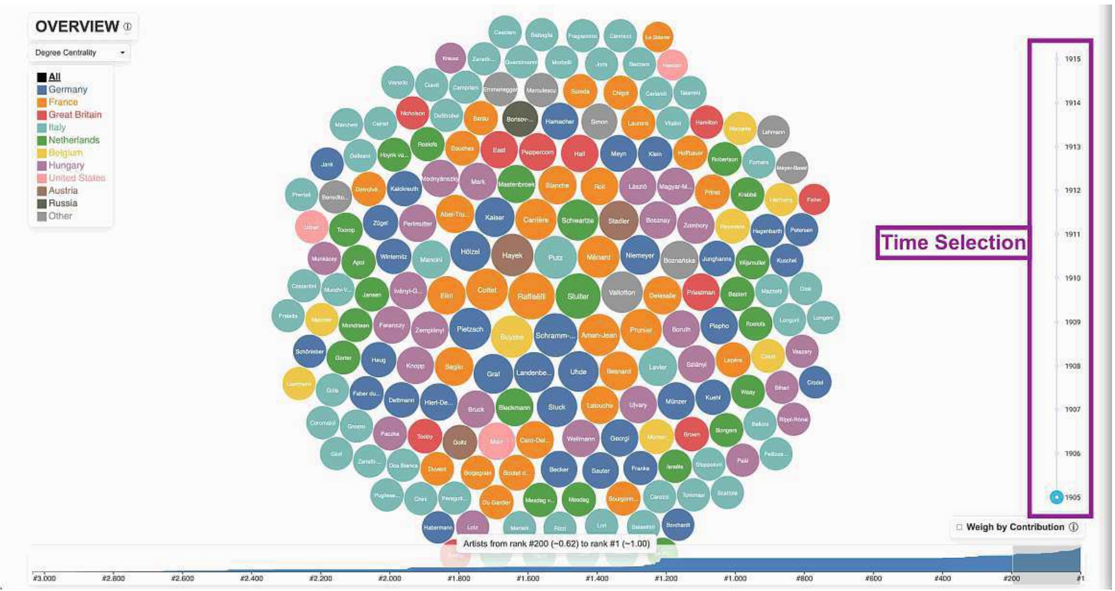

Figure 4.6: Vertical slider for selecting specific years to filter artists, with the year 1905 currently selected, showing only artists exhibiting in that year while maintaining other filters.

eigenvector centrality. As they switch through different years, these other selections are maintained to allow users to focus on this specific area of interest and observe how the centralities of these artists evolve.

Country Selection

Another interesting observation from the literature review is the ability to filter specific groups within the network to conduct a more focused and in-depth analysis. Researchers often choose to narrow their focus to a subset of the dataset to better understand the dynamics within a particular group. For example, in [SZC18], the network is filtered geographically to allow for a more focused analysis of a specific area, while in turn [BEPK24] applies group-categorization filters to focus on a particular subgroup of the network to facilitate gaining insights that might otherwise be obscured when considering a larger network.

Inspired by this, our prototype implements a country selection filter to allow users to focus on artists based on their country of birth. This feature achieves the same level of the targeted analysis found in the literature, but with the added flexibility of visual analytics: by filtering based on country of birth, users can narrow down the dataset in real-time, reducing the noise from unrelated data and focusing on the user's interest. The power of this filter, much like with the temporal selection, arises from coupling it with other active selections. For instance, users can select a specific country, paired with a CM and time period, enabling exploration of various combinations (see Figure 4.7).

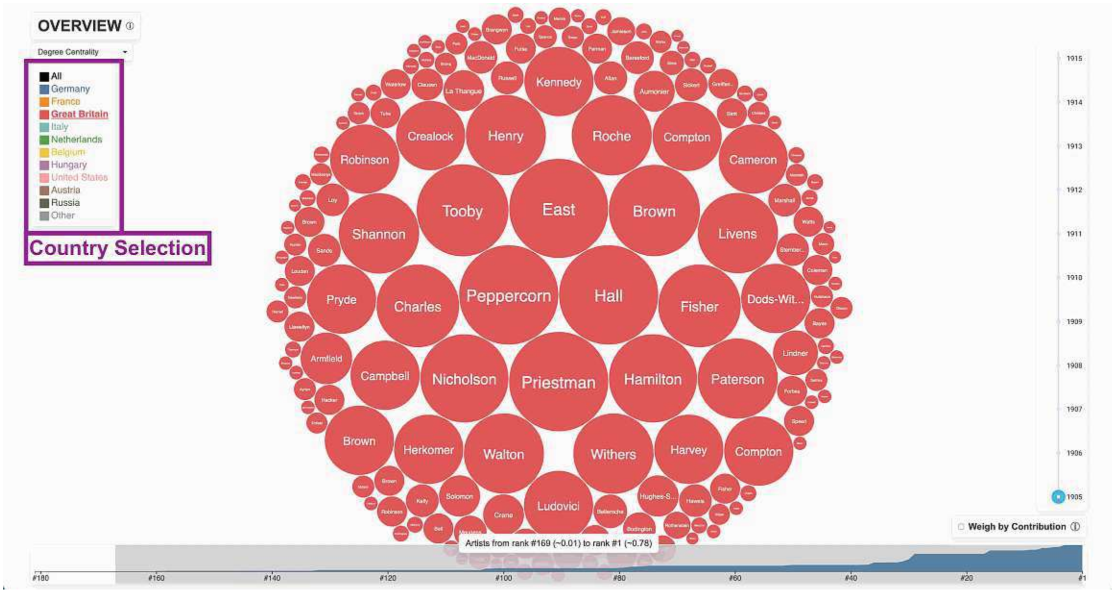

Figure 4.7: Country selection filter showing only artists that were born in the currently selected country. Currently, only artists born in Great Britain are displayed.

The country selection filter is located on the left side of the interface. When a user selects a country, the application recalculates and updates the display to show only the artists from that region. This filter also affects the artists available for selection by the brush tool, enabling users to further refine their analysis within the filtered group. Choosing a specific country might reduce the number of visible artists from 3,000 to only a few hundred.

Distribution Selection

A considerable part of the reviewed research efforts visualize the distribution of centrality scores. This focus can be attributed to the importance of understanding power dynamics within the network, particularly in identifying whether centrality is concentrated among a few key actors or more evenly distributed. For instance, [SFS+19, MRBC19, SBPS23] utilized scatter plots to illustrate how centrality is distributed across networks, often revealing patterns such as power law distributions where a small number of nodes exhibit significantly higher centrality. [LV20, JC24] used a different approach choosing box plots to highlight the concentration of centrality within a select group of individuals.

Inspired by this, our prototype includes a distribution visualization tool that ranks artists based on their centrality metrics. This distribution is displayed at the bottom of the canvas, with artists arranged from the lowest centrality on the left to the highest on the right (see Figure 4.8). By examining this visualization, users can quickly grasp how centrality is spread across the network, providing valuable insights into the hierarchy of the network.

In addition, inspired by research efforts like [FSIK19, LS23], which focused only on specific nodes, namely the most central, for more targeted analysis, dome-insights builds on this approach by introducing interactivity through the brush tool. This allows users to not only view the whole distribution but also explore a specific range along the centrality spectrum. With the brush tool, users can isolate whether they wish to focus on the top 10% most central nodes, moderately central nodes, those with lower centrality, or select a specific selection around an interesting distribution pattern.

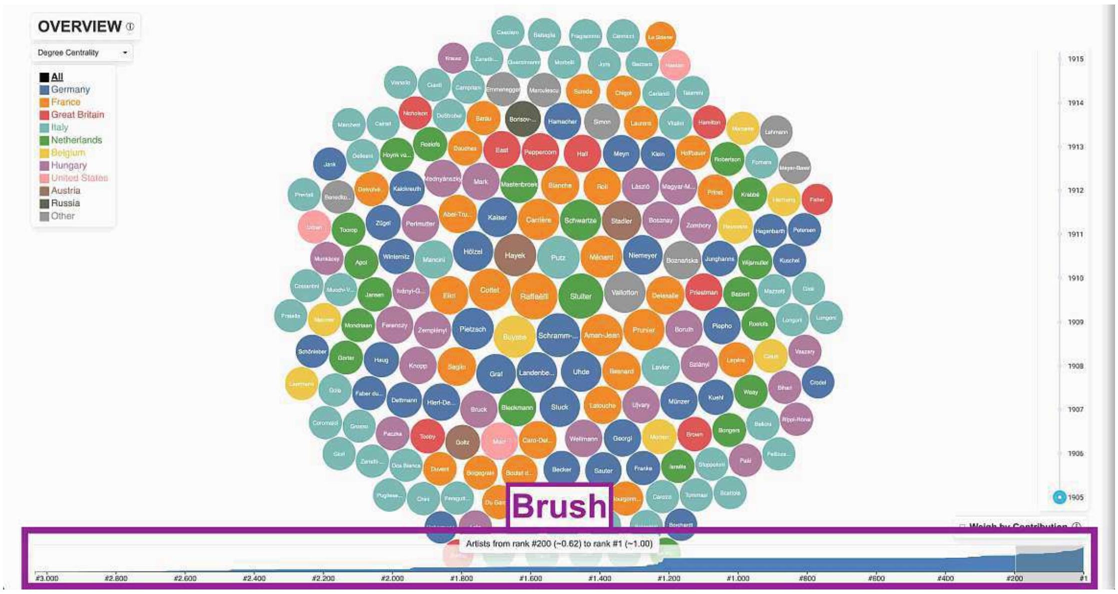

Figure 4.8: Brush selection tool displays the distribution of artists based on their rankings according to the currently selected CM and country.

As users adjust the brush, a label above the tool clearly communicates the selected range, showing the centrality scores and rankings of the chosen selection (see Figure 4.9). This feedback enhances the user's consciousness of the centralities of the currently selected individuals.

Figure 4.9: The label indicates the selected range of artists, from the 200th highest-ranked with an approximate centrality metric of 0.62 to the highest-ranked artist with a centrality metric of around 1.00.

Suite of Selection Tools

In conclusion, our visual analytics prototype takes inspiration from the techniques frequently employed in the reviewed literature (see Section 3.3), where static visualizations, tables, and comparisons are used to analyze complex network data. By implementing features such as CM selection, temporal filtering, country filtering, and distribution selection, our system allows users to explore the network interactively, facilitating the discovery of patterns and relationships from multiple perspectives in real-time.

4.2.2 Metadata

Across the papers reviewed in the literature, a common aspect of presenting results is the frequent display of metadata. In numerous research efforts, tables are employed to present this supplementary data, offering biographical details, temporal data, or context-specific information that help researchers interpret network dynamics and CMs [SBPS23, TS19, SFS+19, MRBC19, MHZ23].

Inspired by the widespread use of metadata in these studies, our prototype incorporates similar metadata directly into the visualization through interactive tooltips. Dome-insights provides tooltip metadata for all nodes of the network. As users hover over a node in the bubble chart, a tooltip appears (see Figure 4.10). In our context, the tooltip displays relevant static details in the tooltips, such as first name, last name, country of birth, and year of birth helping to provide contextual information.

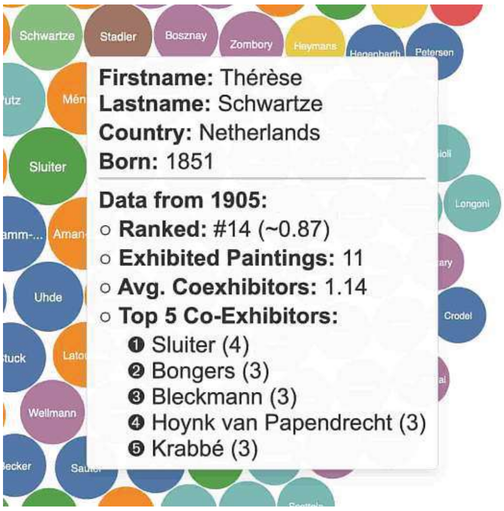

Figure 4.10: On hovering over a node, a tooltip appears, showing detailed information about an artist, including static and year-dependent metadata.

Furthermore, many reviewed research efforts present temporal data through charts or tables, often providing aggregated data that details how a network's characteristics change over time, information that is often crucial for understanding the evolving dynamics of the network. For example, [SBPS23, TS19] provide aggregated information of the network over certain periods of time, while [MHZ23] offers metadata on an individual basis for specific points in time, thus allowing for more granular insights into temporal changes. Moreover, studies like [SFS+19, MRBC19] aggregate data using statistics such as *min*, *max*, *median*, and *mean*.

Inspired by these methods, our tooltips extend the concept of aggregating temporal data by offering year-specific insights about each artist (see Figure 4.11). For each year, the tooltips display concrete statistics specific to the artist, such as the artist's centrality rank in that particular year and other domain-specific data, including the total number of exhibitions they participated in, the total number of paintings exhibited, and the average number of co-exhibitions with each of their co-exhibitors in a given year. This information adds depth to the analysis by not only showing the artist's centrality but also providing important details about their level of activity and contribution to exhibitions during that time.

Figure 4.11: Tooltip displaying year-specific data for an artist for the year 1905.

The tooltip also provides information on an artist's top co-exhibitors for a given year (see Figure 4.12). This is particularly useful when compared with an artist's "Avg. Co-Exhibitions" metric for a given year, as it allows users to determine whether the artist's most frequent co-exhibitors are above the average or not. If they are, this might be indicative of them belonging to an "inner circle" of frequent co-exhibitors that stand out from all the other co-exhibitors that year.

Figure 4.12: Tooltip displaying an artist's most frequent co-exhibitors for the currently selected year.

The different types of data from the tooltip provide valuable domain-specific information that plays an important role in contextualizing the abstract centrality scores.

4.2.3 Weigh by Contribution

Traditional CMs, such as degree, closeness, betweenness, and eigenvector centrality, provide different insights into network structure, addressing questions like who is the most connected, influential, or controls the flow of information. The context-agnostic nature of these CMs makes them widely applicable, offering a broad, generalized view of a network's dynamics.

However, their context-agnostic nature also can be seen as a limitation. They do not account for specific nuances in the dataset. Notwithstanding, attributable to the fact that CMs are essentially just computations, they can be adjusted to address this. For instance, this customization potential was demonstrated by [CU23], who modified traditional CMs by weighting a combination of existing CMs to better align with their research goals.

Inspired by their approach, we also adjust CMs, but based on characteristics of the underlying dataset, and thus add a domain-specific influence. In this example dataset, we introduce a domain-specific angle by reflecting the notion of artistic success, capturing how frequently and extensively an artist's work is exhibited.

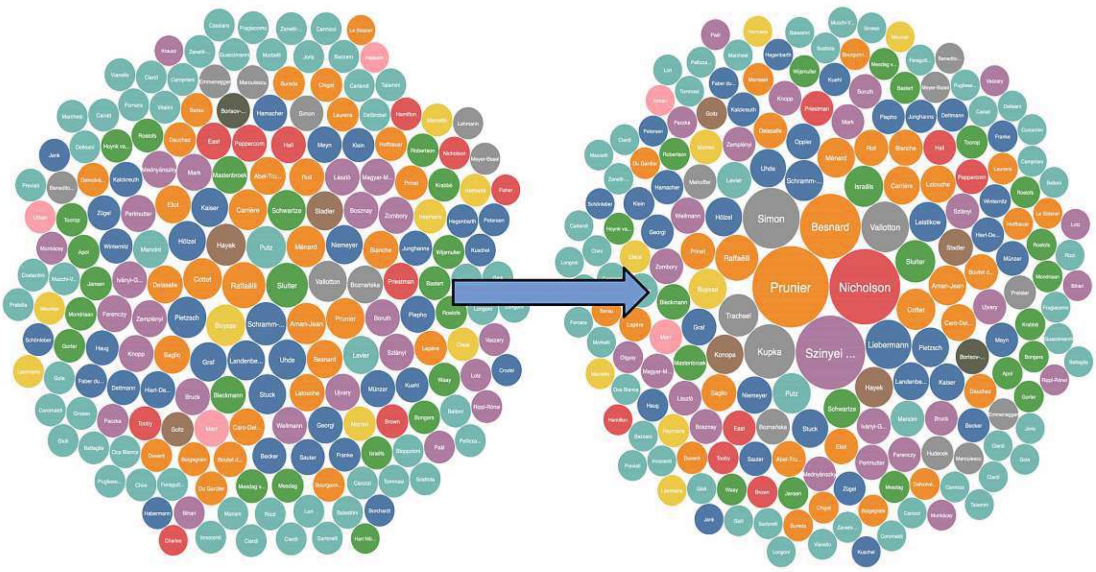

Figure 4.13: The *weigh by contribution* feature adjusts artists' centrality scores and their corresponding node sizes based on their yearly average contribution to exhibitions. The left side shows the network before the feature is activated, and the right side shows the modified centrality values after the checkbox is checked.

By applying a boost to calculated centrality metrics based on an artist's yearly average contribution to exhibitions, this approach aims to add a new perspective to the meaning of a node's centrality. This boost modifies the currently selected CM without replacing or overriding it. Artists who exhibited more extensively in a given year receive an increase in their centrality score, while those with fewer exhibited works receive less of a boost. This adjustment suggests that artists who frequently exhibit multiple works may be viewed as more successful and arguably more central than those who exhibit fewer paintings (see Figure 4.13).

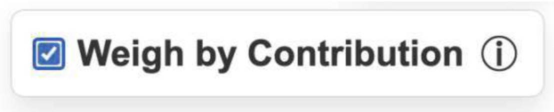

Figure 4.14: Checking this *weigh by contribution* checkbox modifies an artist's centrality based on average yearly contribution to exhibitions.

Importantly, the *weigh by contribution* feature can be turned on or off, allowing users to switch between standard centrality calculations and the contribution-adjusted view (see Figure 4.14). This offers a more nuanced view by continuing to capture the significance of an artist's network position and relationships but also incorporating domain-specific data that reflects concrete information from the dataset.

4.3 Ego-View

While the *Overview* view in our prototype provides a broad understanding of centrality across the entire network, helping users identify key figures and trends, it abstracts away the detailed relationships between individual artists. To address this limitation, we developed the *Ego-View*. This view focuses specifically on visualizing the artist's co-exhibition relationships: the `exhibited_with` connections. By doing so, it provides a much more detailed perspective on how these connections evolve. In this section, we will explain what we aimed to accomplish with the *Ego-View* and the design decisions that helped us achieve these goals. The three overarching objectives to be accomplished by this visualization are as follows:

- **Identifying the Most Relevant Co-Exhibitors:** It was essential to focus on the most important co-exhibitor relationships, as displaying all connections would overwhelm the user and obscure valuable insights. We needed a way to filter and display only the most relevant co-exhibitors.
- **Effectively Displaying Relationships:** Another objective is displaying relationships in a clear and structured manner, making it easy for users to understand who is connected with whom. This requires a layout that organizes relationships visually in an intuitive way, allowing users to follow the connections between the ego and their co-exhibitors without unnecessary complexity or clutter.
- **Tracking Temporal Evolution and Identifying Patterns:** Artists' networks change over time as they co-exhibit with different co-exhibitors. Hence, one objective was to give users a clear understanding of how these relationships evolve.

Ultimately, in pursuit of creating a visualization method that accomplishes these three goals, we drew from insights in the literature to craft the *Ego-View* (see Figure 4.15). By leveraging findings from previous network analysis research, we were able to design a tool that balances the complexity of dynamic network data with ease of interpretation. Our approach enables users to interactively explore the network while maintaining clarity and focus on the most important relationships over time.

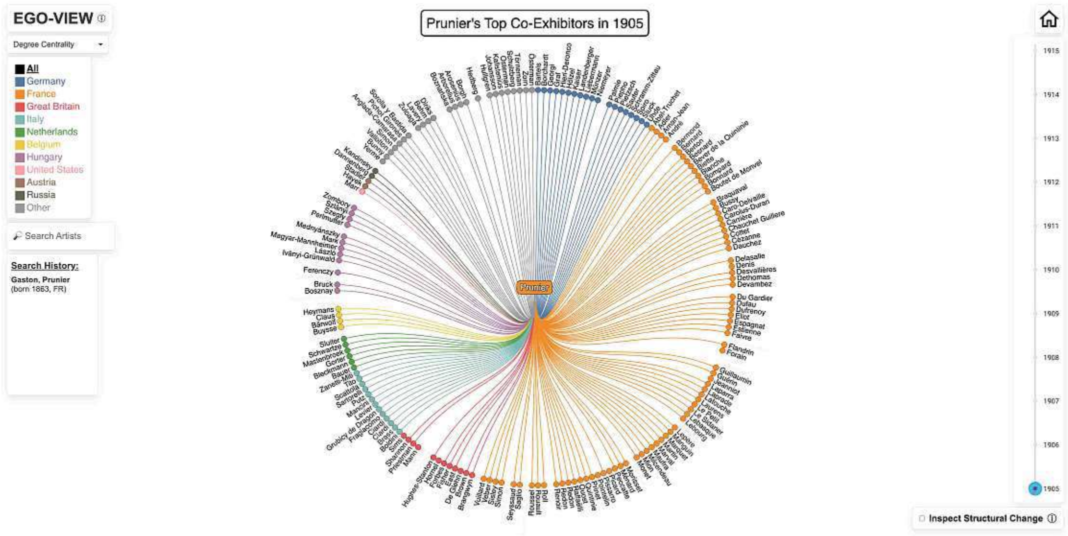

Figure 4.15: The *Ego-View* focuses on an individual artist's connections, This view provides "details on demand" of co-exhibitors with high centrality based on the currently selected CM.

In the following sections, we will now explore how each of these objectives was addressed through specific design choices and elaborate on the features from the *Ego-View*:

4.3.1 Effectively Displaying Relationships

Focusing on individual artists and their most significant connections is crucial for understanding their centrality and position within the network. While the *Overview* provides a broad understanding of network dynamics by highlighting key figures and trends, it abstracts away the specific relationships that shape an artist's centrality. The *Ego-View* aims to address this gap by providing details on the personal network of an artist, specifically focusing on the co-exhibition relationships that connect the artist to others within the network.

Ego networks are particularly effective for achieving this focused examination, as they center on a single node (the ego) and include the nodes directly connected to it (the alters) [BEJ18]. In our dataset, the ego represents an individual artist, and the alters are other artists with whom the ego has co-exhibited. This localized view simplifies the complexity of the overall network, allowing users to examine the immediate and most influential relationships in depth, offering a complementary perspective to the broader network insights provided by the *Overview*.

Additionally, to enable deeper analysis and comparison, our design allows for multiple ego networks to be explored simultaneously. Users can interactively explore the network by *right-clicking* on alters to reveal their own respective ego networks, effectively turning any *right-clicked* alter into a new ego (see Figure 4.16). This multi-layer exploration facilitates comparative analysis by enabling users to examine and contrast various co-exhibitor networks of different artists at the same time.

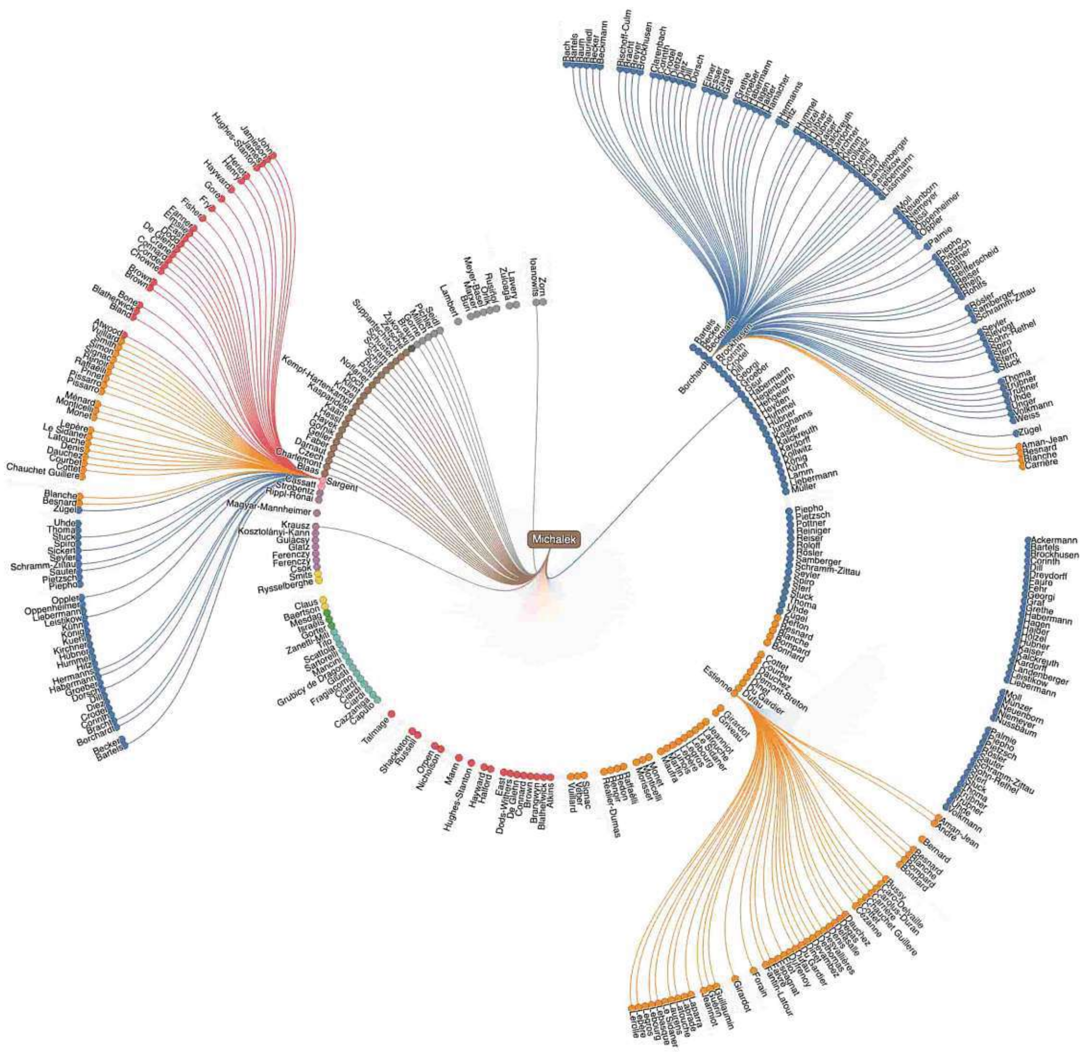

Figure 4.16: Comparison view showing multiple ego networks, including the selected artist's "alter-egos" and their respective ego networks, to facilitate comparative analysis.

To efficiently display the relationships in our ego network visualization, we adopted a radial layout [YFDH01], which organizes nodes in concentric circles around the central ego. This structure is well-suited for visualizing hierarchical data emanating from a single point, making it useful for exploring multi-level co-exhibition relationships without overwhelming the user. As users *right-click* on alters, their networks expand outward, integrating seamlessly with the existing layout while avoiding visual clutter.

Concretely, for a structured and compact radial layout, we chose the *Reingold–Tilford* "tidy" algorithm [RT81] for constructing hierarchical node-link diagrams. This algorithm ensures that all alters and the alters' alters are placed on two separate orbits around the ego. It offers several advantages for our use case. It maximizes space efficiency by expanding outward, allowing multiple levels of connections to be displayed clearly without overlap. It also enhances clarity by minimizing edge crossings and node overlaps, making it easier for users to follow connections.

4.3.2 Identifying the Most Relevant Alters

Despite the conscious choice of the "tidy" radial layout, which optimizes space usage, the size of our dataset, containing over 13,000 artists and hundreds of thousands of relationships, poses a significant challenge. Displaying every relationship would lead to overwhelming visual clutter, making it difficult for users to extract meaningful insights. As a result, we needed to filter the network to focus only on the most relevant alters.

To address this, we leverage CMs, taking advantage of their established effectiveness in identifying and outlining central nodes within networks. We devised an algorithm (see Section 1) that utilizes CMs to highlight only the most significant co-exhibitor relationships. By prioritizing alters based on their centrality score, the system ensures that users can focus on an artist's most central co-exhibitors. This approach enables users to focus on key relationships without being distracted by less impactful connections, an approach mirroring van Ham and Wattenberg's work [VHW08].

Country Selection & CMs

The previously introduced country and CM selectors from the *Overview* are carried over to the *Ego-View*, also enabling the user to dynamically filter and explore artists by applying different selections.

The country selection tool filters out all artists except those from the selected country from the "Top 200 Co-exhibitors" (see Figure 4.17). Changing the selected country or CM recalculates the "Top 200 Co-exhibitors" based on the new criteria. For example, selecting degree centrality and France would display an ego network from 1905-1915 featuring the 200 French artists with the highest centrality scores according to degree centrality.

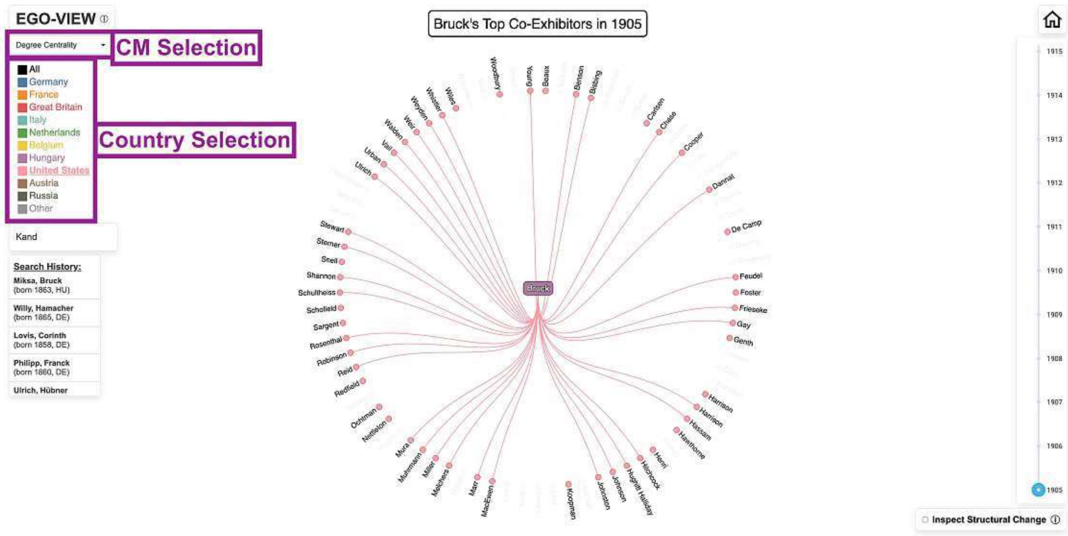

Figure 4.17: Country and CM selectors in the *Ego-View* with the United States selected, filtering out all other countries from the co-exhibitors list and calculating the top co-exhibitors based on the currently selected degree centrality.

Time Selection

The time selector in the *Ego-View* functions differently than in the *Overview*. Changing the year does not alter the displayed artists but rather it dynamically updates the visualization encodings to reflect who exhibited and co-exhibited in the currently selected year (see Figure 4.18).

4.3.3 Tracking Temporal Evolution and Identifying Patterns

Understanding how an artist's relationships evolve over time is crucial for analyzing their centrality and position within the network. As artists co-exhibit with different co-exhibitors throughout their careers, their network connections change. Capturing this temporal dimension presents unique challenges, particularly in maintaining clarity and usability in the visualization.

One of the primary challenges in visualizing dynamic networks is enabling users to track changes over time without causing confusion or cognitive overload. In our approach, we implemented a time slider, allowing users to explore the network at various snapshots, switching between years. Similar to the approaches seen in [SSM$^+$23, AIP$^+$18], this approach also offers time slices but in addition the control to dynamically change the currently selected slice.

Notwithstanding, it revealed a significant challenge, displaying the most central co-exhibitors for any given year virtually always also means different nodes for any given

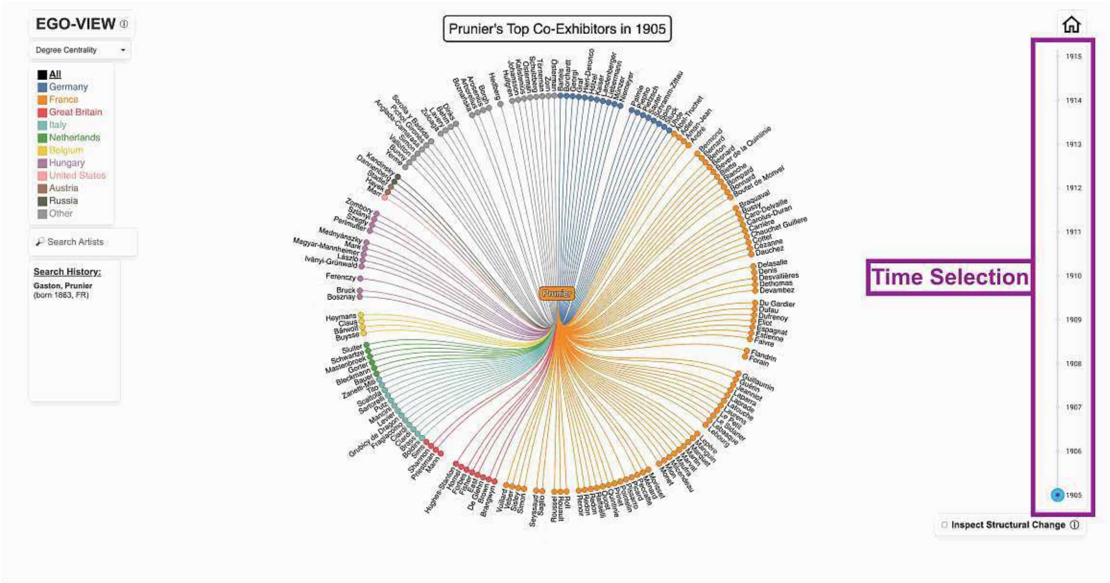

Figure 4.18: With the time selector set to 1905, only artists' nodes that exhibited in 1905 are opaque. Furthermore, only links from artists that co-exhibited in 1905 are opaque.

year, making it thus very difficult for users to follow evolving relationships. Consequently, users had to constantly reorient themselves by finding where nodes moved or they disappeared altogether. This made tracking relationships difficult or impossible.

To address this, instead of calculating the most central nodes for each individual year, we take a different approach. Our method aggregates the most central nodes over the entire 10-year period. This allows us to adopt the approach from [SSM+23], which used node-link diagrams with fixed node positions across different time intervals. By keeping the nodes' positions static over different points in time, users can now observe how relationships develop and change without the confusion caused by moving nodes between time slices. This makes it much simpler to identify patterns, such as strong co-exhibition relationships.

However, using fixed positions presents another challenge: not all alters are consistently active across every time period. Furthermore, while some artists maintain long-standing co-exhibition relationships, others may only exhibit together once. To capture these temporal differences, we implemented a sophisticated encoding system that adjusts the nodes' and links' opacity to reflect the exhibition status. There are three possible cases that arise from this encoding system (see Figure 4.19).

In the first case, the alter exhibited and in addition also co-exhibited with the ego, hence both the alter's and the ego's node as well as the link connecting the alter and ego are opaque (see Figure 4.20). In the second case, the alter exhibited but did not co-exhibit with the ego, as a consequence only the alter's node is opaque, while the link remains transparent, signifying that even though the alter exhibited that year, the alter did

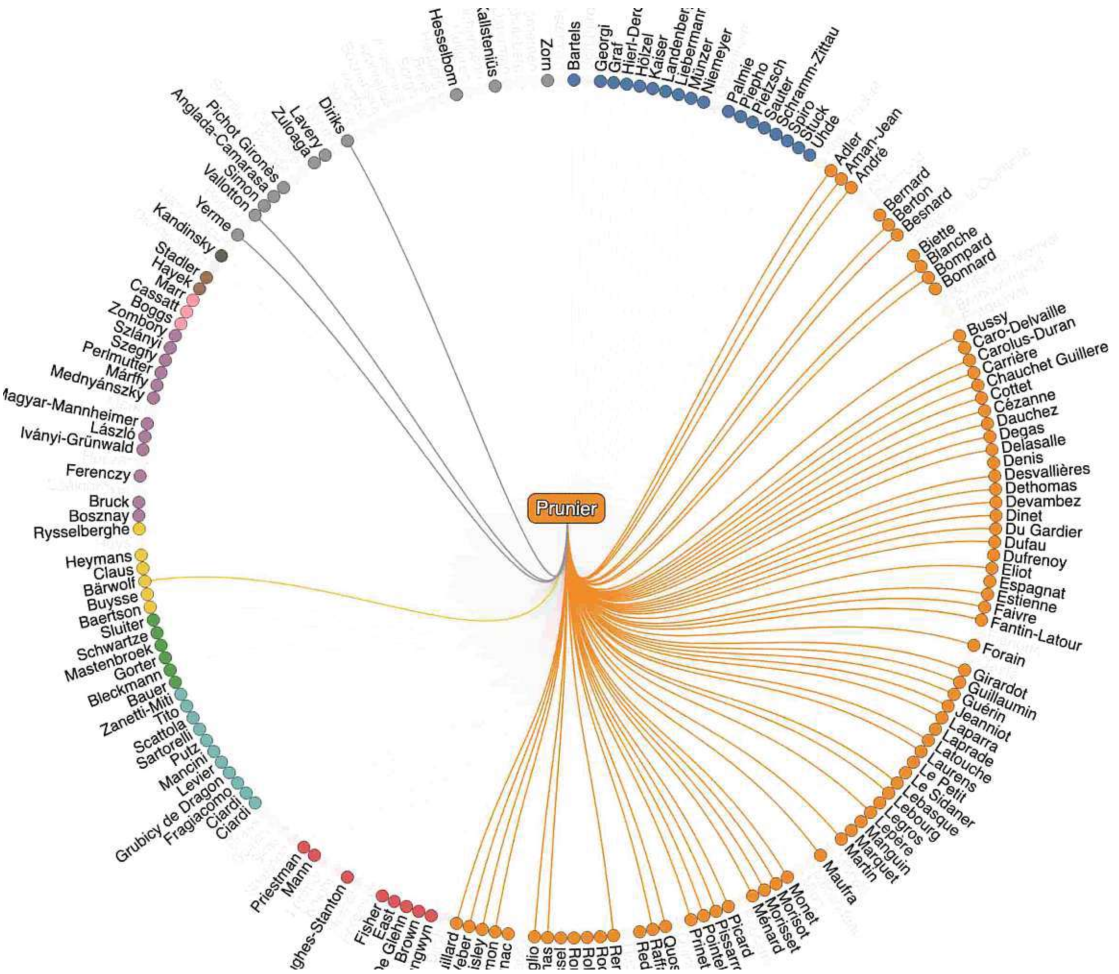

Figure 4.19: The transparency of nodes and links indicate whether artists exhibited and co-exhibited in the currently selected year, respectively.

however not co-exhibit with the ego (see Figure 4.21). Finally, in the third case, if the alter did not exhibit in that year, neither the node nor the link is visible, with both remaining transparent to represent the absence of any exhibition or co-exhibition activity (see Figure 4.22).

This combination of fixed-node positioning and temporal encodings provides the means for users to track and explore how an artist's relationships evolve over time. Users can easily detect key patterns, such as periods of consistent exhibitions or the long-term stability of co-exhibition relationships.

4.3.4 Convenience Features

The prototype offers two convenient tools to support the exploration of the dataset: the search bar and a history. The search bar allows users to quickly find any of the

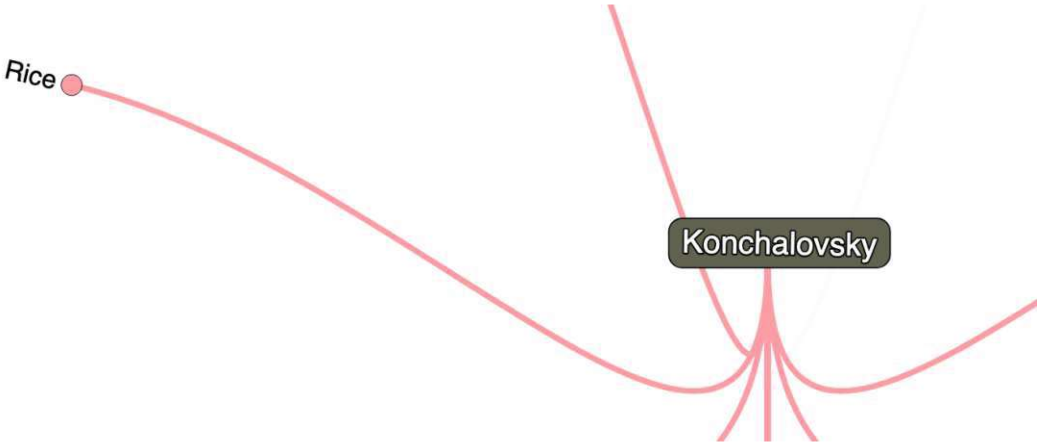

Figure 4.20: The alter's node as well as the link connecting the alter with the ego is opaque, indicating the alter exhibited and further shows that the alter also co-exhibited with the ego during the selected year.

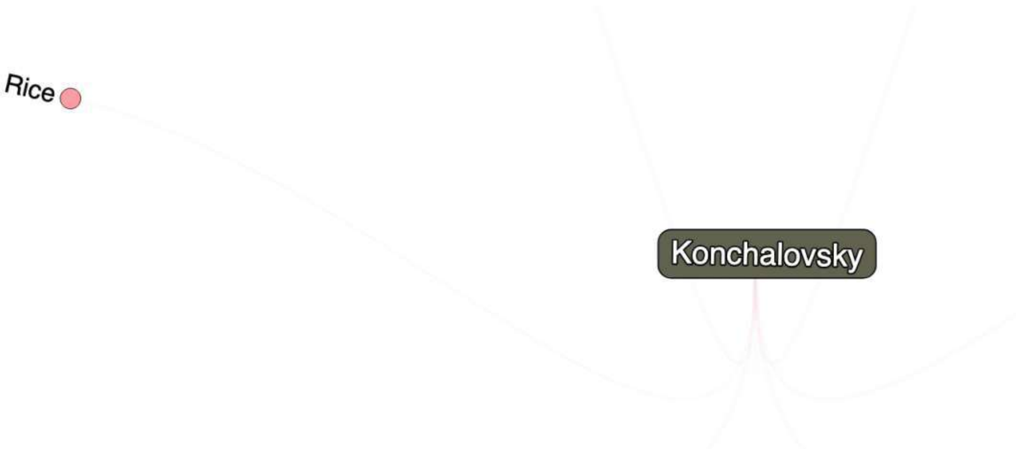

Figure 4.21: The alter's node is opaque indicating that the alter exhibited that year. However, since the link connecting the alter with the ego is transparent, they did not co-exhibit that year.

13,000+ artists by typing in their name, with real-time suggestions helping to narrow down potential matches (see Figure 4.23). The history records previously explored artists, allowing users to revisit already viewed artists.

4.3.5 Inspect Structural Change

Understanding how relationships within a network evolve over time is essential for gaining insights into shifting dynamics and centralities. Tracking these temporal changes helps identify variations in relationships and the evolving importance of participants. Visualization techniques, such as the *Ego-View*, can provide details on how an artist's

Figure 4.22: The alter's node as well as the link connecting the alter with the ego is transparent, indicating the alter did not exhibit that year and consequently also no co-exhibition could have taken place.

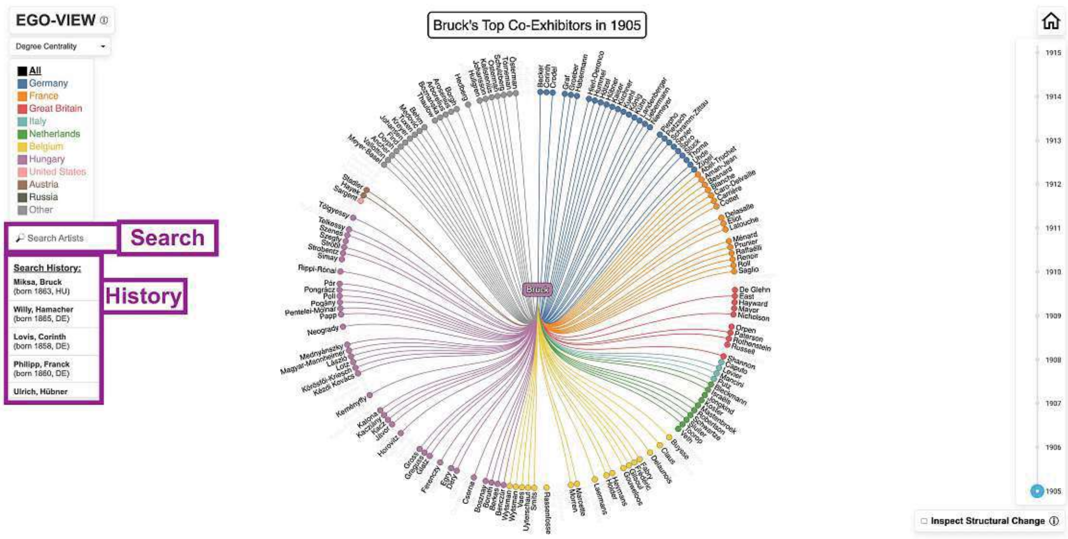

Figure 4.23: Two convenience features: search bar for quickly locating artists and a history for revisiting previously explored ego networks.

co-exhibition network evolves. However, detecting where these changes occur can be challenging when switching between time points due to change blindness [ELMS91], which occurs when users struggle to notice and remember differences between two periods.

In an effort to address this issue, we use a dynamic CM called *change centrality* [FPA+12]. This CM quantifies the degree of change in an artist's co-exhibition network between two selected time periods, condensing these changes into a single value. For instance, *change centrality* is useful to easily detect nodes that have experienced a significant change in their respective network of co-exhibitors.

While this approach offers valuable insights into network dynamics, it does not specify where within the network the change has occurred, nor does it reveal the underlying causes of these shifts. Rather, the "inspect structural change" feature quantitatively expresses the extent of change in an artist's network between two points in time. This information can guide users toward areas of potential interest, such as artists who experienced significant change over a short period, or conversely, those whose networks remained stable over many years. However, it is important to note that *change centrality* only provides a number representing this change, without offering deeper insights into the causes or specific areas within the network that contributed to these changes. Despite these limitations, the measure serves as a useful addition to the exploratory analysis of evolving dynamics in an artist's co-exhibition networks.

By enabling the "inspect structural change" feature (see Figure 4.24), users can adjust the time slider with an additional knob (see Figure 4.25) to select a range of years and analyze the change that occurred during that particular period. The calculated *change centrality* values are visualized through color-coded nodes on a gradient scale, ranging from blue (indicating minimal change) to red (indicating significant change) (see Figure 4.26). Nodes experiencing the maximum change due to the artist not exhibiting in either of the selected years are colored black, thus clearly distinguishing them from nodes who exhibited in both of the selected years.

Figure 4.24: When the *inspect structural change* checkbox is selected, a 5-color scale appears as a legend to interpret color changes.

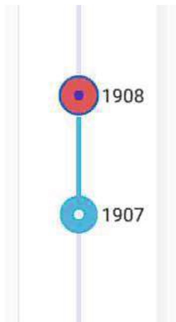

Figure 4.25: When the *inspect structural change* checkbox is selected, a second knob appears on the time slider.

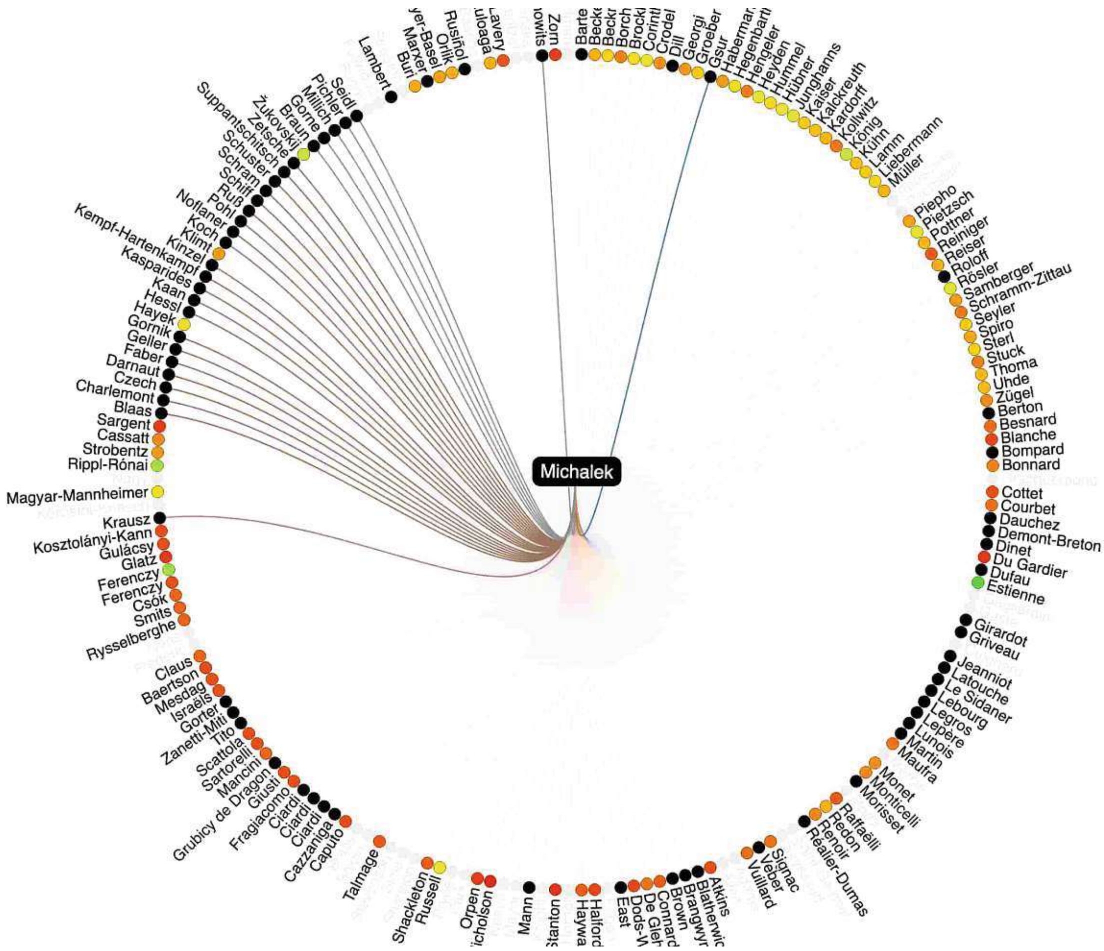

Figure 4.26: Nodes are color-coded to represent structural change, with colors indicating the amount of change that occurred for each artist individually: blue for minimal change, red for extensive change, and black for when the artist didn't exhibit in either of the two selected years.

CHAPTER 5

dome-insights - Implementation

This chapter focuses on the implementation of dome-insights[1], building upon the previous design chapter. The following sections cover the core aspects of the implementation, emphasizing technical details vital for the prototype's performance and functionality.

5.1 Architecture

To implement the design features outlined in the previous section, a two-tier architecture was utilized. The frontend web app was developed using `Angular CLI` version `17.1.1`[2] and `Node.js` version `21.6.1`[3], utilizing the *D3.js* library[4] for the creation of visual components. The backend is powered by *FastAPI* (version `0.111.0`)[5], a lightweight web framework designed for building APIs, providing an ideal platform for fast prototyping.

The frontend is transpiled into static files and deployed via *GitHub Pages*[6], a straightforward option for quickly getting a web app deployed and facilitating continuous deployment. The backend is hosted on *Heroku*[7], running on a single `m-dyno` worker. The backend application is served using *Uvicorn* (version `0.30.1`)[8], an `ASGI` server chosen for its excellent compatibility with *FastAPI*[9] (see Figure 5.1).

[1] dome-insights: `github.com/tridelt/dome-insights`. Accessed on October 13, 2024.
[2] Angular CLI v17.1.1: `github.com/angular/angular-cli`. Accessed on September 4, 2024.
[3] Node.js: `nodejs.org/en/`. Accessed on September 4, 2024.
[4] D3.js: `d3js.org`. Accessed on September 4, 2024.
[5] FastAPI: `fastapi.tiangolo.com/`. Accessed on September 4, 2024.
[6] GitHub Pages: `pages.github.com`. Accessed on September 4, 2024.
[7] Heroku: `heroku.com`. Accessed on September 4, 2024.
[8] Uvicorn: `uvicorn.org/`. Accessed on September 4, 2024.
[9] FastAPI: `fastapi.tiangolo.com/`. Accessed on September 4, 2024.

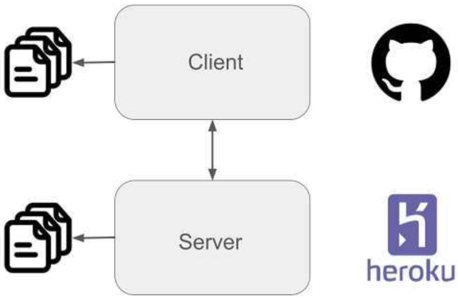

Figure 5.1: Two-tier architecture of the web application, featuring an Angular frontend and a FastAPI backend.

5.2 Data

The data for this project is sourced from the DoME dataset[10], originally stored in CSV files and later imported into a Neo4j database to improve organization and querying capabilities.

As outlined earlier, the main focus is on exploring the network of co-exhibitions, i.e., the relationships between artists based on when they exhibited together. These co-exhibition relationships form the foundation of the analysis. In the dataset, co-exhibitions are modeled using the following structure:

```
(a:Artist)-[:EXHIBITED]->(ex:Exhibition)<-[:EXHIBITED]-(b:Artist)
```

While this structure effectively captures the participation of artists in shared exhibitions, it presents a challenge for calculating centrality metrics. The issue stems from the fact that the network is heterogeneous, consisting of artists, exhibitions, and other types of entities. However, centrality metrics from nodes of exhibitions can arguably not be meaningfully compared with centrality metrics from nodes of artists.

To address this, a new relationship, titled `exhibited_with`, was created to directly connect artists based on their co-exhibition (see Figure 5.2). This allows for the construction of a new, homogeneous graph containing only artist nodes and allows for a more focused representation of the co-exhibition relationships.

These `exhibited_with` relationships ultimately serve as the foundation for all features of the prototype. They are crucial for identifying key figures within the art world and analyzing evolving relationships. Because CMs are an abstraction, they allow us to condense the vast dataset, containing millions of `exhibited_with` relationships, into just a few thousand centrality scores. This reduction not only simplifies the complexity of the data but also opens up new opportunities for efficiently managing and delivering the data to the frontend.

[10]DoME: `exhibitions.univie.ac.at/`. Accessed on June 13, 2024.

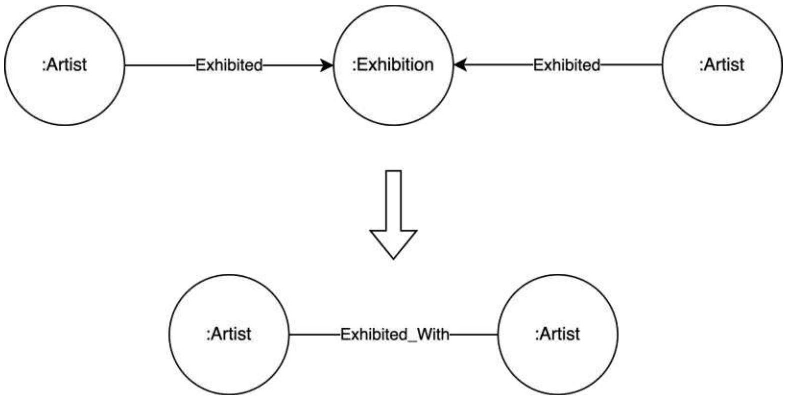

Figure 5.2: Diagram illustrating the transformation of a co-exhibition relationship from a heterogeneous to a homogeneous one.

As a result, the way data is handled for the prototype is influenced by its small size and the need for low latency to ensure a smooth user experience. Although *Neo4j* is powerful for complex graph queries, the prototype does not require such advanced queries. Instead, much of the necessary data can be predicted and pre-calculated in advance. Since the dataset is relatively small, most of it can be served alongside the static files of the website, avoiding the need for real-time queries. Even though some more advanced features require on-demand basic data processing, this still can be done by a backend without a live instance of *Neo4j*. By preparing and processing the data outside of the database, the prototype loads faster and performs more smoothly, without relying on complex live database queries.

Consequently, the prototype employs two primary approaches for data delivery:

1. **Pre-loaded Data**: Since the centrality metrics can be pre-calculated and summarized, the resulting dataset is compact enough to be served directly alongside the static files of the web app. This pre-loaded data, including each artist's centrality score, is retrieved when the page loads, ensuring a fast and responsive user experience. Storing this data in JSON files allows the application to perform efficiently without requiring real-time calculations.

2. **On-Demand Data**: For more complex features that go beyond pre-calculated centrality metrics, data is processed dynamically on the backend side. Real-time calculations are performed as needed, allowing for more requests without overloading the frontend. This hybrid approach maintains responsiveness while accommodating more advanced data needs.

5.3 Overview

The main component of the *Overview* view is a bubble plot that displays differently sized bubbles, created using D3.js's `d3.pack()` layout. It computes the hierarchical structure of the data and positions the nodes accordingly. The data is structured into a hierarchy using D3.js's `d3.hierarchy()` function, which allows each node's value to determine its size. An SVG container holds the visualization, while the nodes are positioned based on the layout's computed x and y coordinates. Finally, circles are drawn for each node, with their radii defined by the packing algorithm to visually represent the data.

Another vital component of the *Overview* is the brush tool, built using D3.js's `d3.brushX()` function. This tool allows users to filter the displayed artists based on their centrality. To create a brushable area in D3.js, the `d3.brushX()` function is set up with `extent()` to define the selection bounds, allowing users to select a range of artists within these limits. The x-axis is drawn using `d3.axisBottom()` to display the time range, and an area plot is created using `d3.area()` to visually represent the data. The brush group is appended to the SVG container with `svg.append("g").call(brush)`, and a default selection is applied with `d3.brush.move()` to initialize the interaction.

The following two sections explore the main logic powering the core features of the *Overview*, followed by the *weigh by contribution* feature, which dynamically adjusts the visualization based on the number of paintings by each artist.

5.3.1 Main Logic Enabling the Frontend Features

The data logic that powers the *Overview* component is fully handled on the frontend side, leveraging pre-loaded data for efficient processing. Rather than relying on a traditional backend, all necessary data is pre-calculated and served at runtime. This ensures that no on-demand data fetching is required, allowing for access to all potential user requests with very low latency.

The visualization tools in the *Overview* component are driven by distinct modules within the Angular application. These modules control the bubble plot, brush tools, and other selection elements such as the year slider, CM dropdown, and the *weigh by contribution* checkbox. These tools enable users to refine and filter the artist data based on various criteria, thus facilitating dynamic interaction with the dataset.

All of these tools are coordinated through the `overview-service.ts` file, which serves as the central hub for managing user interactions and synchronizing updates across the interface. This service tracks user inputs and efficiently filters data from the pre-loaded dataset to update the visualizations in real-time. Whenever a user interacts with the interface of the *Overview*, the following steps are performed by `overview-service.ts` to process the data and update the visualization to the new selection:

1. Load data for all artists from a pre-loaded JSON file.

2. Filter the data based on the current country selection.
3. Modify CM metrics based on whether the *weigh by contribution* is selected.
4. Sort the data based on the selected CM, ranking artists from highest to lowest.
5. Filter the data to include only artists within the range selected using the brush tool.

Once the relevant data is processed, it is published to a `BehaviorSubject` within `overview-service.ts`. By pre-loading and filtering the data on the frontend, the system minimizes latency and enhances performance, providing users with instant feedback as they explore the visualization.

5.3.2 Weigh by Contribution Feature

The goal of the *weigh by contribution* feature, as outlined in the design chapter, is to capture an artist's contribution to exhibitions over the course of a year. This feature allows for a more domain-specific representation of an artist's success.

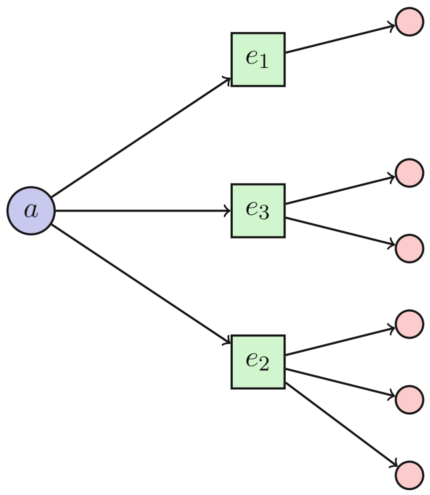

Figure 5.3: Visualization of the different contributions an artist a makes to three exhibitions (e_1, e_2, e_3).

To gauge an artist's contribution for the *weigh by contribution* feature, we calculate their participation in each exhibition for a given year by determining their percentual contribution to each event. The artist's yearly contribution is then derived by averaging these individual percentages across all exhibitions they participated in that year. This approach provides an approximation of the artist's success, as it accounts for the varying degrees of involvement in multiple exhibitions (see Figure 5.3).

The centrality adjustment formula accounts for the number of paintings an artist contributed to exhibitions within a given year (see Equation 5.1).

$$C'(a) = C(a) \times \left(1 + \beta \times \frac{1}{|E(a,y)|} \sum_{e \in E(a,y)} \left(\frac{N_a^e}{N^e}\right)\right) \tag{5.1}$$

Where:

- $C(a)$ is the artist's base centrality score.
- β is a scaling factor that adjusts how strongly the contribution is weighted.
- $|E(a,y)|$ is the number of exhibitions in which artist a participated during year y.
- N_a^e is the number of paintings contributed by artist a to exhibition e in year y.
- N^e is the total number of paintings in exhibition e in year y.

The expression inside the parentheses is the already weighted factor, based on an artist's relative contributions to exhibitions in a given year, with which an artist's base centrality is multiplied. The addition of 1 ensures that artists with minimal contributions retain their base centrality score, while the scaling factor β controls how much these contributions impact the final score.

The system calculates the artist's average relative contribution across all exhibitions they participated in during a specific year. This contribution value is then multiplied by the base centrality score, $C(a)$, to yield the adjusted centrality, $C'(a)$. By adjusting the value of β, one can fine-tune the weight the yearly average contribution shall have, allowing for more or less emphasis on an artist's exhibitions within a given year when calculating centrality.

5.4 Ego-View

The *Ego-View* component provides a sophisticated temporal visualization of an artist's (ego's) network of co-exhibitors (alters) over a 10-year period, dynamically capturing shifts in centrality and relationships through a radial hierarchy. This visualization is implemented using D3.js to construct a *Radial Tidy Tree* layout, which arranges the network in a circular format. The layout is based on the Reingold-Tilford "tidy" algorithm, which ensures a visually pleasing representation of hierarchical data by arranging nodes in a tree structure that maximizes readability [RT81]. The algorithm organizes the data efficiently, projecting it radially to display co-exhibitors' nodes. To render the relationships between the ego and co-exhibitors, D3.js's `linkRadial` function is used to draw the connections (links) between nodes.

The following two sections explore the main logic powering *Ego-View*, followed by the *change centrality* feature, which supports exploring change in an artist's co-exhibitor network between two years.

5.4.1 Main Logic Enabling the Frontend Features

Unlike the relatively straightforward logic behind the *Overview* view, which handles basic metadata for a group of artists, the *Ego-View* visualization requires a more advanced approach. This is because the *Ego-View* not only focuses on the ego's co-exhibitors but also integrates a temporal component, displaying yearly data to show which co-exhibitors were active and which co-exhibited with the ego in any given year. As a result, the system requires on-demand data processing to support the dynamic and interactive nature of the visualization.

Adding to the complexity is the need to optimize the user experience for interactions with the timeline. To enable a smooth experience when exploring the timeline of an artist, going back and forth through the 10 years of the dataset must function without repeatedly contacting the backend. Consequently, this requires all necessary data for temporal exploration to be delivered in a single request and the backend to construct and deliver a complex Data Transfer Object (DTO) that contains all the data required for different years for a selected ego network.

Before describing the backend processes in detail, it is first necessary to outline the specific data elements required to support the *Ego-View*. The data structure consists of three key components: static metadata, dynamic yearly metadata, and additional encoding information required to support the visualization.

The static metadata includes the following basic, static details about each artist:

- `id`: Unique identifier for the artist.
- `firstname`: The artist's first name.
- `lastname`: The artist's last name.
- `country`: The artist's country of origin.
- `birthdate`: The artist's year of birth.

In addition to the static data, the dynamic metadata is pre-aggregated and organized by year. This dynamic metadata allows the system to display information for any selected year. The following fields capture this information:

- `total_paintings`: Total number of paintings created by the artist for each year.
- `avg_coexibs`: Average number of co-exhibitions the artist had with each co-exhibitor for each year.
- `coExhibitors`: A nested array containing the artist's top co-exhibitors for each year.

- `n_exhibitions`: The number of exhibitions the artist participated in for each year.
- `score`: A list of yearly centrality scores, representing the artist's centrality or contribution for each year.
- `children`: A nested list containing the top 200 co-exhibitors of the ego. Each co-exhibitor entry includes the same yearly metadata (e.g., total paintings, co-exhibitions, years active, centrality scores) as described for the ego. In addition, this data includes metadata for each co-exhibitors and also includes their activity status and co-exhibition status with the ego for each year.

Lastly, the DTO also includes information specifically designed to support the radial tree visualization itself. This data is contained in two key lists:

- `years_active`: A list of years during which the artist was active. This data determines whether the node representing the artist is opaque or transparent.
- `years_coexhib`: A list of years in which the artist co-exhibited with the ego, controlling the transparency of the links connecting the ego and the alters.

The following is an abridged example of a DTO, in JSON format, for the Russian painter Vassily Kandinsky:


```
{
    "id": "5",
    "name": "Kandinsky",
    "firstname": "Vassily",
    "lastname": "Kandinsky",
    "country": "RU",
    "birthdate": "1866",
    "total_paintings": {
        "1905": 17,
        "1906": 47,
        "1907": 98,
        ...
    },
    "avg_coexibs": {
        "1905": 1,
        "1906": 1.2,
        "1907": 1,
        ...
    },
    "coExhibitors": {
```

```
        "1905": [],
        "1906": [],
        ...
    },
    "years_active": [
        "1905", "1906", "1907", ...
    ],
    "score": {
        "1905": 0.2642,
        "1906": 0.8537,
        "1907": 0.2733,
        ...
    },
    "children": [
        ...
    ],
    "n_exhibitions": {
        "1905": 3,
        "1906": 8,
        "1907": 4,
        ...
    }
}
```


As mentioned before, to keep the visualization responsive and avoid frequent backend queries, all necessary data to facilitate smooth temporal exploration is sent to the frontend in a single response. This includes pre-aggregated data for the entire 10-year period, ensuring smooth operation as users explore different years for the same ego. However, in the case when a user makes a different selection. For instance, when the user chooses a different ego, or the user changes the country or CM filter, then, the backend runs an algorithm to calculate new relevant data reflecting the selection changes made by the user and sends a new DTO to the frontend. This algorithm works as follows, using the example of Vassily Kandinsky:

1. **Identify the Top Co-Exhibitors:** The algorithm first iterates through each year and aggregates Kandinsky's top 20 co-exhibitors based on their centrality metrics for each year. If the number of unique co-exhibitors does not reach 200, the algorithm iterates again, this time selecting 20 + `iteration_i` co-exhibitors in subsequent rounds, until 200 unique co-exhibitors are aggregated or the pool of available co-exhibitors is exhausted.
2. **Rank and Organize Co-Exhibitors:** Once the top co-exhibitors are identified, they are ranked according to their centrality scores. This ranking determines their positions in the radial layout.

3. **Populate Metadata for Each Co-Exhibitor:** For each co-exhibitor, all necessary metadata is populated.

4. **Determine Yearly Activity and Relationships:** The system also determines the activity status of each co-exhibitor for each year. This data is required for the frontend to adjust the visualization dynamically, showing which of Kandinsky's co-exhibitors were active and who co-exhibited with Kandinsky each year.

5.4.2 Change Centrality Feature

Change centrality is a dynamic CM that measures the extent to which a node's neighborhood experiences change between two points in time [FPA+12]. The *change centrality* for each artist is calculated by comparing the changes between two different years according to the following algorithm (see Figure 5.4). This algorithm calculates the symmetric difference between the set of their connections at two times and divides it by the union of those connections. This produces a change ratio that reflects how much a node's immediate connections have changed. The change centrality further combines these ratios for different degrees of separation to account for the changing connections of neighbors and more distant nodes, providing a more comprehensive measure of the network's structural evolution [FPA+12].

ChangeCentrality$(G, t1, t2)$

for $i \in V(G_{t2})$ **do**
  if $i \in V(G_{t1})$ **then**
    $CC_i = 0$
  else
    $CC_i = 1$
  end if
  for all $j \in [1..diameter]$ **do**
    $\bigcup_{ij} = \{v \in V : d_{t1}(i,v) = j\} \cup \{v \in V : d_{t2}(i,v) = j\}$
    $\bigcap_{ij} = \{v \in V : d_{t1}(i,v) = j\} \cap \{v \in V : d_{t2}(i,v) = j\}$
    if $\bigcup_{ij} \neq \emptyset$ **then**
      $CC_i += \left(\left|\bigcup_{ij}\right| - \left|\bigcap_{ij}\right|\right) / \left|\bigcup_{ij}\right|$
    end if
  end for
end for
return $[CC]$

Figure 5.4: Algorithm for calculating *change centrality* as proposed in "Visual Analysis of Dynamic Networks using change centrality" by Federico et al. [FPA+12].

To optimize performance, the *change centrality* value for each artist was pre-calculated for every year combination. Similar to the way the centrality scores are handled, these pre-calculated change metrics are served together with the other static files, eliminating the need for dynamic processing.

CHAPTER 6

Evaluation

In this chapter, we evaluate the effectiveness of dome-insights.

6.1 Introduction to Value-Driven Methodology

6.1.1 Methodology

Evaluating visualization tools presents significant challenges, as the insights they generate are often subjective and highly dependent on the context. What one user finds valuable may differ from another's experience, influenced by variations in tasks, goals, or backgrounds [Nor06]. Additionally, the effectiveness of a visualization tool cannot be easily captured by a single metric, as it entails assessing how well a tool enables users to uncover patterns, understand complex data, and derive meaningful insights. These qualitative aspects, such as user experience and comprehension, are difficult to quantify, making it challenging to evaluate a visualization tool's overall utility and quality through standardized measures [YKSJ08].

To address these challenges, we adopt the ICE-T value-driven evaluation framework, which aims to capture the broader benefits of visualization tools. This method emphasizes a holistic assessment by focusing on the four key aspects forming the foundation for value-driven evaluation, ICE-T [Sta14]:

- **Insight**: "spurring the generation of insights and insightful questions"
- **Confidence**: "generating confidence and knowledge about the data's domain and context"
- **Essence**: "conveying the essence of the data"
- **Time**: "minimizing the time to answer diverse questions"

This comprehensive approach offers a structured and repeatable framework for generating more consistent evaluations across varied contexts [Sta14].

6.1.2 Heuristic

While Stasko et al.'s description of a value-driven approach [Sta14] effectively highlights key components for assessing a visualization's value, it lacks a concrete method to put the framework into practice and actually conduct the evaluation. Addressing this, Wall et al. [WAM+18] developed heuristics to translate the theoretical concepts of ICE-T into actionable steps.

Wall et al. [WAM+18] accomplished this by developing a structured three-level hierarchy: high-level components (ICE-T), mid-level guidelines, and low-level heuristics. They developed heuristics by breaking down each component of ICE-T into mid-level guidelines that capture the essence of its intended goal, such as how well a visualization supports insight generation or builds user confidence. These guidelines are further refined into low-level heuristics: specific, rateable statements that provide concrete criteria for assessing a visualization's effectiveness. These heuristics make it possible to translate the abstract principles of value-driven evaluation into measurable terms, allowing evaluators to systematically assess key aspects, such as uncovering insights, effectively communicating complex data, and fostering trust in data representation. Through iterative refinement, by focusing on rateability and applicability to visualizations, Wall et al. ultimately derived 21 heuristics (see Table 1 in Appendix).

Wall et al. [WAM+18] demonstrated the viability of these heuristics in an evaluation with a small set of participants, who were tasked to rate each visualization on a Likert scale (1 = strongly disagree, 7 = strongly agree). The individual heuristic scores were then averaged to determine their guideline scores, which ultimately were averaged to the overall ratings for each of the four ICE-T components. This aggregation process provides a comprehensive and nuanced evaluation of the visualization's value as defined by ICE-T. By translating theoretical concepts into practical evaluation steps, this method fulfills the value-driven approach envisioned in Stasko et al.'s original framework [Sta14].

6.2 Evaluation of dome-insights

The evaluation of the visual analytics prototype, *dome-insights*, is divided into two complementary parts: a value-driven evaluation and a qualitative assessment.

For the value-driven evaluation, we closely follow the heuristics-based ICE-T framework developed by Wall et al. [WAM+18], with minor adaptations to suit our study. Unlike Wall et al.'s evaluation of three visualizations, we focus on a single visualization, dome-insights. Wall et al. confirm in their research paper that this is still a valid application of their method. Based on their recommendation, we selected five participants for the evaluation. According to a study by Wall et al., five raters are sufficient to produce consistent and reliable results in heuristic evaluations [WAM+18]. In line with the methods proposed by Wall et al., the participants in our study all have some familiarity with the domain of the visualized data and had no prior exposure to the prototype, ensuring unbiased, first-time user experiences. The five raters evaluated the tool using the 21 heuristic-based questions from Wall et al.'s heuristic approach (see Table 1 in Appendix). Similarly, the participants also evaluated the questions using Likert-scale-based and single-choice answers [WAM+18].

In addition to the value-driven evaluation, we included a qualitative evaluation through four open-ended questions designed to investigate the effectiveness of the *inspect structural change* feature compared to the traditional year slider, addressing SRQ2 (see Section 8.2 in Appendix).

6.2.1 Evaluation Procedure

The evaluation will be conducted remotely over the course of one hour and divided into three phases:

1. **Introduction (15 minutes):** Participants will receive an introduction to the prototype, explaining how the functionality of the prototype works and how CMs were used to enhance the exploration of the network and gain insights.
2. **Exploration (30 minutes):** Participants will independently explore the tool, testing key functionalities, like the *inspect structural change* feature to understand how it supports structural and temporal data analysis.
3. **Questionnaire (15 minutes):** Participants will complete a Google Forms[1] questionnaire including both quantitative and qualitative questions.

Participants were welcome to ask questions at any point during the introduction and exploration phases. The questionnaires were filled out in private without any further contact with the presenter. Before the presenter disconnected, the raters were given access to the Google Forms questionnaire as well as access to a *cheat sheet* that visually summarizes and labels all of the features from the prototype (see Table 1 in Appendix).

6.2.2 Data Collection and Analysis

We will collect both quantitative and qualitative data. Quantitative data will come from participants' ratings of the 21 heuristics across the four ICE-T dimensions: *Insight, Confidence, Essence,* and *Time*. These heuristic scores will first be averaged to arrive at the scores for the individual guidelines, then by components, and finally into an overall score representing the tool's overall assessment score.

Qualitative data will be gathered from the four open-ended questions in the Google Forms questionnaire. These responses will be analyzed for recurring themes, suggestions, and critiques, allowing us to assess the effectiveness of the *inspect structural change* feature.

[1] Google Forms: `google.com/forms/`. Accessed on September 25, 2024.

6.3 Results

In this section, we present both quantitative and qualitative findings from our evaluation of the prototype. The results are divided into two parts: first, an assessment using the ICE-T framework, followed by a qualitative analysis of two different approaches for observing temporal network changes.

6.3.1 Value-Driven Assessment of dome-insights

The quantitative analysis evaluates the visualization using the ICE-T framework, which focuses on four key components: *Insight*, *Confidence*, *Essence*, and *Time*. This subsection provides an overview of the overall results and a detailed breakdown of the scores for each component, based on the participants' feedback.

Survey Overview

The survey results from five participants (P1 to P5) (see Table 6.1) reveal a generally positive perception of the visualization, with consistently high scores across the four ICE-T categories: *Insight*, *Confidence*, *Essence*, and *Time*. Across the board, participants rated the visualization favorably, with most scores ranging between 5 and 7, indicating agreement to strong agreement on the visualization's utility. *Insight* and *Essence* stood out with the highest scores, averaging 6.45 and 6.25 respectively, respectively. These results suggest that the visualization is highly successful at generating both intentional and incidental insights, helping users uncover important patterns and unexpected findings. Additionally, the visualization effectively conveys the overall essence of the data, providing a clear and holistic overview while maintaining contextual relevance.

Time and *Confidence* received slightly lower, though still positive, ratings. *Confidence*, averaging 5.84, reflects a moderate level of trust among participants in the accuracy and completeness of the data representation. While users felt reasonably confident in their understanding of the data, the results indicate some hesitation pertaining to the reliability of the dataset and how well it was represented. Similarly, *Time*, with an average score of 6.01, suggests that while the visualization generally facilitates efficient exploration and browsing of the data, there remains potential for improvement in terms of speed and ease in accessing specific information. Notwithstanding, the aggregated score across all categories resulted in an average of 6.14, highlighting the visualization's effectiveness.

The following tables present survey results for each component: *Insight*, *Confidence*, *Essence*, and *Time*. For each component, a table displays the participants' ratings for each component's respective guidelines and the questions derived from them.

Table 6.1: Survey results from five participants (P1 to P5), each of whom responded to 21 different questions (Q1 to Q21) [WAM+18] on a Likert scale from 1 (Strongly disagree) to 7 (Strongly agree). The responses are color-coded: red tones indicate disagreement while green tones indicate agreement.

	Insight								**Confidence**				**Essence**				**Time**				
	G1			**G2**		**G3**			**G4**		**G5**	**G6**	**G7**		**G8**		**G9**		**G10**		
Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P1	6	6	6	6	6	6	6	6	6	5	6	–	6	6	6	5	6	6	5	6	6
P2	7	7	7	7	6	6	6	7	7	6	6	5	6	6	6	7	6	5	6	6	7
P3	6	7	6	7	7	7	6	6	6	5	6	4	7	6	6	6	6	6	6	6	7
P4	6	6	6	7	7	7	5	6	6	6	6	–	7	7	6	6	4	6	6	7	7
P5	5	7	7	7	7	7	7	7	7	6	7	7	6	7	7	6	6	6	7	6	7
μ	6.0	6.6	6.4	6.8	6.6	6.6	6.0	6.4	6.4	5.6	6.2	5.3	6.4	6.4	6.2	6.0	5.6	5.8	6.0	6.2	6.8
μ	6.33			6.70		6.33			6.00		6.20	5.30	6.40		6.10		5.70		6.33		
μ	6.45								5.84				6.25				6.01				
μ	6.14																				

The symbol '–' indicates "Not applicable".

Insight

The results from the *Insight* component (see Table 6.2) indicate that the visualization effectively supports users in engaging with the data, addressing specific tasks, and uncovering unexpected insights. Notably, the *Insight* component was rated the highest among all four components, with all questions receiving scores above 4. This high overall score can be attributed, in part, to the overwhelmingly positive ratings for the questions derived from the guideline focusing on providing a new or better understanding of the data.

Concretely, the question about helping users generate data-driven questions received the highest rating within the entire component, with an average score of 6.80. Most participants gave this question a rating of "strongly agree," indicating the visualization's effectiveness in prompting deeper exploration and forming new inquiries. Similarly, the question evaluating the identification of unusual or unexpected data characteristics scored an average of 6.60, reflecting how the visualization also excelled in encouraging users to uncover new, unexpected insights.

Interestingly, across the entire *Insight* component, only two responses fell into the "somewhat agree" range, while the vast majority reflected stronger agreement. This suggests a high level of consensus among participants regarding the tool's ability to support intentional insights and serendipitous discoveries. Additionally, in the case of the question related to generating data-driven inquiries, all but one participant unanimously rated it as "strongly agree," further reinforcing the tool's perceived value in aiding structured, data-driven exploration. These results demonstrate the visualization's overall strength in supporting both intentional and incidental insights and the significant contribution of this component to the evaluation's positive outcomes.

Table 6.2: Survey results on *Insight*, showing participant ratings on how the visualization facilitates obtaining insights through targeted search and incidental insights through browsing [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization facilitates answering questions about the data	The visualization exposes individual data cases and their attributes	6	7	6	6	5	6.00
	The visualization facilitates perceiving relationships in the data, like patterns & distributions of the variables	6	7	7	6	7	6.60
	The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	6	7	6	6	7	6.40
The visualization provides a new or better understanding of the data	The visualization helps generate data-driven questions	6	7	7	7	7	6.80
	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	6	6	7	7	7	6.60
The visualization provides opportunities for serendipitous discoveries	The visualization provides useful interactive capabilities to help investigate the data in multiple ways	6	6	7	7	7	6.60
	The visualization shows multiple perspectives about the data	6	6	6	5	7	6.00
	The visualization uses an effective representation of the data that shows related and partially related data cases	6	7	6	6	7	6.40

Confidence

Albeit noticeably lower than the average of the other components' averages, the results for the *Confidence* component (see Table 6.3) indicate a generally favorable perception of the visualization's ability to support accurate interpretation and understanding of data quality with some notable areas of disagreement between the raters. The highest-rated question in this component, with an average of 6.40, assessed whether the visualization uses meaningful and accurate visual encodings to represent the data. However, the second question under the guideline "The visualization avoids using misleading representations," focused on avoiding making incorrect inferences, and received a noticeably lower average score of 5.60. This suggests some concern among the evaluators regarding the potential for misrepresentation, leading to more varied ratings between 5 and 6.

Related to the guideline on facilitating learning about the domain of the data, the results were very consistent, with an average score of 6.20. Although the responses were largely in "agreement", participants did not express "strong agreement", with ratings clustering around 6, except for one participant who rated this question a 7. Given the consistency in the results, this suggests a shared reason for dissatisfaction between the raters.

A point of significant variability emerged in the question about the visualization's ability to help understand data quality issues. The average score of 5.33 masks a wide range of responses, with one participant rating this aspect highly at 7, while others rated it as low as 4. Interestingly, two participants found this question not applicable, indicating that some users did not feel this aspect was relevant to their interaction with the visualization or that the visualization did not adequately address these issues for them. This significant disagreement highlights an area where the visualization may not have been as effective or consistent.

Table 6.3: Survey results on *Confidence*, presenting participant ratings on how the visualization supports accurate interpretation and understanding of data quality [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization helps avoid making incorrect inferences	The visualization uses meaningful and accurate visual encodings to represent the data	6	7	6	6	7	6.40
	The visualization avoids using misleading representations	5	6	5	6	6	5.60
The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	6	6	6	6	7	6.20
The visualization helps understand data quality	If there were data issues, like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues	–	5	4	–	7	5.33

The symbol '–' indicates "Not applicable".

Essence

The results for the *Essence* component (see Table 6.4) show that the visualization performs consistently well in providing a big-picture perspective of the data and facilitating generalizations beyond individual data cases. The scores for this component reflect a generally strong performance, with minimal variation across participants and the overall score for the component ranks only marginally behind the *Insight* component.

The first guideline, "The visualization provides a big picture perspective of the data," received an average score of 6.40 for both questions. Participants agreed that the visualization offers a comprehensive and accessible overview of the data, with ratings exclusively in the 6 to 7 range. This indicates that the visualization effectively presents the data in a way that enables users to grasp the overall structure and meaning of the dataset, promoting a high-level understanding.

The second guideline, "The visualization provides an understanding of the data beyond individual data cases," received slightly lower average scores, ranging from 6.00 to 6.20. The question regarding the facilitation of generalizations and extrapolations of patterns and conclusions scored 6.20 on average, reflecting that most participants agreed that visualization helps users move beyond individual data points to draw broader insights. However, the question about understanding how variables relate to accomplishing different analytic tasks received the lowest average score within this component, at 6.00. One participant rated this aspect a 5, and despite the relatively good average of 6.00, compared to the other questions, it arguably suggests that there may be some room for improvement.

Table 6.4: Survey results on *Essence*, summarizing participant ratings on the visualization's ability to provide a big-picture perspective and facilitate generalizations beyond individual data cases [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization provides a big picture perspective of the data	The visualization provides a comprehensive and accessible overview of the data	6	6	7	7	6	6.40
	The visualization presents the data by providing a meaningful visual schema	6	6	6	7	7	6.40
The visualization provides an understanding of the data beyond individual data cases	The visualization facilitates generalizations and extrapolations of patterns and conclusions	6	6	6	6	7	6.20
	The visualization helps understand how variables relate in order to accomplish different analytic tasks	5	7	6	6	6	6.00

Time

The results for the *Time* component (see Table 6.5) indicate that the visualization generally supports efficient browsing and the retrieval of specific information, though there is some variability in participants' ratings.

Under the first guideline, "The visualization affords rapid parallel comprehension for efficient browsing," the visualization received mixed ratings. The question evaluating whether the visualization provides a meaningful spatial organization of the data averaged 5.60, with one participant rating it as low as 4 while the other participants rated it with a 6. This outlier indicates that while most users found the spatial organization helpful, one participant seemed to feel it could be improved and thus rated it as "Neither Agree nor Disagree". The question regarding whether the visualization shows key characteristics of the data at a glance scores marginally better with an average of 5.80, with ratings between 5 and 6, reflecting that the visualization generally helps users quickly grasp essential information.

The second guideline scored noticeably better than the first one but with relatively great variance within its individual questions, with average scores ranging between 6.00 and 6.80. The question regarding the interface's ability to support the reorganization of the visualization based on different data attributes scored 6.00 on average. However, the question concerning the visualization's ability to avoid complex commands and textual queries by providing direct interaction with the data representation received a 6.80, with no rating lower than 6, indicating that participants found the visualization to be very effective in enabling direct interaction with the data. This is the best score for a heuristic in the entire evaluation, with only one other question, "The visualization helps generate data-driven questions," also scoring a 6.80.

Table 6.5: Survey results on *Time*, reflecting participant ratings on how efficiently the visualization allows for rapid browsing and the retrieval of specific data [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	6	6	6	4	6	5.60
	The visualization shows key characteristics of the data at a glance	6	5	6	6	6	5.80
The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance	5	6	6	6	7	6.00
	The visualization supports smooth transitions between different levels of detail in viewing the data	6	6	6	7	6	6.20
	The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	6	7	7	7	7	6.80

6.3.2 Qualitative Assessment of Traditional and Novel Approaches for Observing Network Changes

As part of the questionnaire, the five evaluators also answered four open-ended questions. These questions were designed to gather qualitative feedback on two distinct approaches for observing changes within an artist's co-exhibitor network over time. The *traditional approach* required participants to manually switch between different years to visually assess changes in an artist's ego network. In contrast, the *novel approach* employed the dynamic CM *change centrality* [FPA+12] to calculate a single value expressing the change in an artist's co-exhibitor network between two years.

To explore the efficacy of these two approaches, the evaluators were asked four questions, each probing the comparison from a different point of view (see answers in the Appendix 8.2). The following subsections summarize the qualitative feedback given for each question:

Highlighting Important Changes

Q1: When observing a network over time, do you find that the novel approach highlights important changes more clearly than the traditional approach?

Participants generally felt that the novel approach offered clearer and faster insights into changes within dynamic networks, enabling them to track significant shifts more effectively than traditional methods. Many emphasized that this approach enhanced the overall efficiency of visualizing complex data over time.

However, numerous respondents pointed out areas that need improvement, particularly to better distinguish color encodings conveying change from the ones indicating an artist's nationality. Participants stated that simultaneously using different color encodings within the same diagram made it difficult to interpret the data accurately, even though participants appreciated the underlying concept.

Understanding the Evolution of Key Nodes

Q2: How well does the novel approach help you understand the evolution of key nodes in a network? Does it offer insights that the traditional approach might miss?

Participants mostly agreed that the novel approach provided a more reliable way to track the evolution of key nodes compared to traditional methods. Many remarked that the tool helped make changes more visible, reducing the risk of overlooking important data points or missing subtle shifts that static methods might not capture.

Several respondents highlighted that the novel approach was particularly effective at revealing subtle yet significant transformations in the network. For example, one participant noted that it helped them identify the evolving role of an artist in a specific group, a detail that would likely be missed when manually switching between time periods using traditional approaches. Additionally, one participant specifically mentioned that the novel approach allowed them to address broader historical questions, such as analyzing the differences in networks before and after the outbreak of World War I. However, one participant expressed difficulties in understanding the novel approach suggesting that it may contribute to a steeper learning curve.

Combining Both Approaches

Q3: Do you think using the novel approach alongside the traditional approach provides a more complete understanding of network changes, or do they tend to overlap too much in the insights they provide?

Most participants felt that combining both the novel and traditional approaches offered a more complete understanding of network changes. They indicated that the two methods complement each other, with minimal overlap in the insights they provide. According to several evaluators, the traditional approach remains valuable for detailed analysis of individual data points, while the novel approach excels at capturing broader patterns over time.

Some participants remarked that the novel method uncovers specific insights, such as changes in centrality and subtle network dynamics, which might be missed with the traditional approach. One participant emphasized that the novel approach provided a more complex and enriched understanding of the network's evolution, suggesting that using both approaches together yields a fuller and more nuanced picture of network changes.

However, one respondent offered more critical feedback, noting that, for the sake of clarity, the novel and traditional approaches should be visualized separately, as their current integration may cause confusion. Furthermore, the rater expressed concerns about how the visualization handles cases where artists do not exhibit every year, which is a frequent scenario in the dataset. The participant pointed out that the current method of displaying changes over a 10-year span often results in no meaningful centrality being displayed for artists who did not exhibit during either of the years from the selected period. Even when they exhibit intermittently, the current approach essentially excludes artists with gaps in their exhibition history. The rater suggested adopting a more flexible approach to address this issue to allow more artists to be included in the analysis.

Key Advantages of the Novel Approach

Q4: If you were to explain the benefits of the novel approach to a colleague, what would you highlight as its main advantage over the traditional approach?

Participants identified several key advantages of the novel approach, particularly its speed, flexibility, and ability to generate new insights. Many emphasized that it allows for a quick, high-level overview of larger patterns, making it easier to grasp trends across the dataset. One participant even noted that the convenience of the tool made it something that every dataset should have.

Additionally, several participants pointed out that the novel approach encourages deeper exploration by prompting new research questions. For example, one respondent appreciated how the tool's ability to reveal subtle dynamics which traditional year-by-year methods often miss. In general, participants praised the tool for offering insights that static methods alone would struggle to reveal.

CHAPTER 7

Discussion

This chapter aims to answer the research question of this thesis:

"Can centrality measures enhance visual analytics approaches for dynamic network visualization to improve insight-based analysis and help identify key actors during the analysis?"

To provide a comprehensive response, this chapter is structured around the following three SRQs that were formulated at the beginning and guided the research effort (see Section 1.2). Each following section of this chapter will address an SRQ by relating the results from the evaluation and elaborating on the implications:

7.1 Visualization Methods for CMs in Network Analysis

SRQ1: Which visualization methods are used to communicate insights into network analysis using centrality metrics?

7.1.1 Visualization Methods

Recognizing that centrality metrics are abstractions of network structures [Rod19b] is essential for effective visualization. Metrics such as degree, betweenness, and eigenvector centrality quantify the significance of nodes based on their relationships. Consequently, centrality metrics offer an abstract representation of a node's connections, capturing its role within the network without needing to explicitly display all underlying relationships. This abstraction enables new opportunities for visualization, allowing the essence of a node's connections to be conveyed without having to visualize them explicitly.

While node-link diagrams are commonly used in network visualizations [CM11], they are not always the most effective way to represent centrality metrics. While node-link diagrams are commonly seen as the go-to method for displaying the relationships between nodes, since centrality metrics already quantify these relationships by measuring a node's relative importance, visualizing the entire network with its many relationships isn't necessary or even helpful for the purpose of conveying a particular insight. As a consequence, alternatives to the node-link visualization with different strengths become viable. This abstraction thus awards flexibility.

The literature review highlights this flexibility by showcasing the diversity of visualization tools used for conveying CMs (see Table 3.1). This table summarizes the numerous different methods employed as part of the reviewed paper of the literature review. This demonstrates the broad range of approaches that can effectively communicate insights from CMs. Furthermore, the discussion in the state-of-the-art literature review (see Section 3.3.1) highlights the flexibility and diversity of visualization tools used for CMs, empowering researchers to use the right tool for the job depending on their analytical goals.

7.1.2 Temporal Changes

As previously shown in the discussion section of the literature review (see Table 3.1), which compares the frequencies of visualization methods in the literature, 75% of the papers dealing with dynamic network data did not use dynamic visualizations, i.e. visualizations explicitly depicting time. This suggests that time-based visualizations are not always indispensable for conveying insights from dynamic network data. Many researchers, despite working with dynamic network data, opted for visualizing metrics without a dynamic context. Instead, temporal aspects were often referenced externally in tables or charts, complementing the visualizations of centrality metrics. Ultimately, the decision to include temporal elements should be guided by the value and clarity they add to the research. According to the reviewed literature, this means that in many cases simpler, static methods may be more effective.

7.1.3 Conclusion

In conclusion, understanding that centrality metrics are abstractions of network structures increases the pool of available visualization techniques that effectively communicate insights from network data. This flexibility lets researchers move beyond traditional node-link diagrams and utilize alternative methods, like scatter plots and bar charts that are not usually associated with visualizing dynamic network data. Also, the literature highlights that many opt for static visualizations instead of more complex dynamic visualizations. While temporal changes are inherent in dynamic networks, time-based visuals are not always necessary and should be included only when they add value. Ultimately, the choice of what visualization method to use and whether to also convey time should align with the specific analytical goals, and the most effective visuals are those that clearly and concisely convey the data.

7.2 Static vs. Dynamic CMs for Temporal Analysis

SRQ2: Do novel dynamic centrality measures more effectively capture temporal changes in networks compared to their static counterparts?

CMs have proven effective in our prior analysis for gaining insights and identifying key actors within networks (see Table 6.1). However, capturing temporal changes in networks using this prototype still presents challenges, particularly due to issues like change blindness [ELMS91] when comparing different points in time. On this issue, extensive research [BBDW17] has explored various methods, such as animation, layering, transitions, and superimposition, to effectively convey changes across time, but comparing snapshots remains cognitively demanding. While this research effort devised *Ego-View* using various techniques to address these challenges, it still doesn't fully alleviate the difficulties of detecting temporal changes.

In an attempt to address this challenge and further explore whether dynamic CMs can enhance visual analytics in dynamic contexts, we explored various ways CMs can capture temporal changes within a network and tested two approaches. The first, a conventional method, relies on a slider for manual comparison of network snapshots in *Ego-View* (see Section 4.2.1), forcing users to visually track and remember changes over time, a process that can be mentally taxing. The second, novel approach, employs *change centrality* (see Section 4.3.5), a dynamic CM that quantifies temporal changes into a single value. This method simplifies making comparisons by reducing complex visual changes to a number [FPA+12].

Evaluator feedback highlighted the advantages of the novel approach. Many found that it allowed them to perceive changes more quickly and clearly than the traditional method, as outlined in the results section of the qualitative evaluation (see Section 6.3.2). One participant noted, "Absolutely, changes are visible much faster and in a clearer way." This reflects the benefit of having a single metric to express the amount of change that has occurred thus reducing the cognitive load associated with manual comparisons. Furthermore, due to this, the novel approach was praised for uncovering subtle but important transformations that might otherwise be missed. An evaluator commented, "By utilising centrality measures, the novel method enables multi-layered comparisons, uncovering subtle dynamics and patterns over time that are difficult to detect through the traditional, year-by-year visual inspection." This highlights the potential of dynamic CMs to provide a deeper understanding of network evolution, revealing complex dynamics that static snapshots might overlook.

A key takeaway from the feedback section (see Section 6.3.2) on the novel approach is the paradigm shift introduced by quantifying changes through dynamic CMs. This reduces the cognitive burden on users. Instead of manually tracking and recalling visual changes across time points, users can rely on a single metric to identify significant transformations, streamlining the analysis process. This shift from manual comparison to algorithmic calculation allows users to focus on the magnitude of change rather than the chore of comparing two images.

Nevertheless, the traditional approach remains valuable for detailed analysis. One participant emphasized, “For more in-depth information about the individual data presented, traditional approaches will stay relevant, probably without creating too much of an overlap.” This highlights that while the novel approach excels at providing a broad overview and identifying broader changes, traditional methods are crucial for capturing more nuanced developments.

Evaluators suggested that utilizing both approaches in conjunction yielded the most insightful results when exploring. They noted that, despite some loss of information with the novel approach, it excels at guiding users to the most relevant areas by highlighting where the largest changes occur. Complementing this, the traditional method can then be employed for a closer examination of specific time points or relationships. This hybrid approach leverages the strengths of both methods, offering a more comprehensive understanding of network evolution.

7.2.1 Conclusion

Overall, the comparison study demonstrates that while no single approach is definitively superior, as each method offers unique strengths depending on the context, the quantification of change through the novel approach presents a valuable complement to the traditional approach. Ultimately, the combination of both approaches facilitates a more comprehensive understanding of dynamic networks.

7.3 Effectiveness of a Visual Analytics Prototype Using CMs

SRQ3: How well can a visual analytics approach for dynamic network visualization facilitate deeper insights and assist in identifying key actors using different centrality measures?

A key element in assessing how well a visual analytics approach based on CMs can facilitate deeper insights was the development of a functional prototype[1]. Creating a prototype allowed us to evaluate the effectiveness of various strategies employed to leverage CMs to enhance the analysis of dynamic network data.

The prototype demonstrated strong efficacy, as indicated by consistently high scores in *Insight*, *Essence*, and *Time* (see Table 6.1). These findings confirm that a visual analytics system built around CMs not only aids in identifying key actors but also promotes deeper insights into dynamic networks (see Section 6.3.2). Below, we discuss two key factors that arguably contributed to the success of the prototype by employing CMs.

[1]dome-insights: `github.com/tridelt/dome-insights`. Accessed on October 13, 2024.

7.3.1 Promoting Intentional and Incidental Insights

The *Insight* component, with the highest average score of 6.45 (see Table 6.2), underscores the system's aptitude to foster both intentional, task-driven insights and incidental, serendipitous discoveries. Participants frequently reported that the system encouraged them to explore beyond their initial objectives. For instance, one rater highlighted how the prototype's design was the reason to investigate relationships that hadn't previously been considered, showcasing the system's role in facilitating a broad range of insights.

A major contributing factor to this outcome was the system's ability to switch between different CMs allowing users to inspect nodes' changing centralities from different perspectives. This added nuance and enriched the discovery process. Furthermore, the system struck a balance between providing a macro-level overview and allowing micro-level analysis. This fluidity between various levels of detail, reflected in the *Time* component's score of 6.01 (see Table 6.5), enabled users to seamlessly shift their focus from the overall network structure to specific nodes of interest.

7.3.2 Tracking the Evolution of Key Actors

Another key strength of the prototype was its capability to monitor the evolution of key actors and their centralities across time. On a broader level, the *Overview* empowered users to effortlessly follow the dynamic transformations within the network as a whole by showing how node sizes and dominance of nationalities evolved from year to year.

At a more granular level, the *Ego-View* enabled two complementary ways to track changes between nodes. Participants highlighted how the ability to visualize gradual changes using a year-by-year slider, alongside quantifying absolute changes between two points in time through *change centrality* significantly deepened their understanding of evolving relationships.

7.3.3 Limitations of dome-insights

Visualization Approach

A critical aspect of the prototype is its ability to visualize temporal changes and pattern shifts in an artist's co-exhibitor network. The *Ego-View* facilitates this by displaying the top 200 artists who have co-exhibited with the ego. However, this approach has an inherent limitation: it selects the top 200 co-exhibitors based on their centrality score, under the assumption that individuals with the highest metrics are of greater interest. However, this method does not reflect the closeness or frequency of relationships between artists, which is arguably particularly important in the context of co-exhibition networks. For instance, an artist may have exhibited only once with famous artists at a large exhibition while frequently co-exhibiting with lesser-known artists. In such cases, the algorithm prioritizes co-exhibitors with high centrality scores, potentially omitting frequent and close co-exhibitors who have lower centrality scores but may be of higher interest to users seeking to understand strong, close relationships. This limitation arises partly due to practical considerations: for many artists, there are simply not enough co-exhibitions to reliably build a top 200 co-exhibitor network based solely on the number of co-exhibitions.

Centrality Measures

At the end of the day, CMs are abstractions because they reduce complex relational data into numerical values. And while they provide insights into network dynamics, they don't always align perfectly with real-world interpretations of concepts, like "influence" or "importance." For instance, even though a node with high eigenvector centrality is commonly associated with being "influential", it might not always align with what domain-experts consider to be an "influential" artist. As a result, to provide a more holistic and nuanced understanding of a node's importance, users are well advised to use multiple CMs in conjunction. Rodrigues outlines how different CMs can calculate different centralities for the same node based on the CM and suggests using multiple CMs may be more effective in characterizing a network [Rod19b].

Notwithstanding, the problem with that however is that working with CMs requires the user to know their meaning and learn what they capture. And while CMs, like degree centrality, are relatively straightforward to understand, others, such as eigenvector centrality and betweenness centrality [Rod19b], are less trivial abstractions and have a steeper learning curve, moreso even dynamic CMs, like *change centrality* [FPA+12], which often employ sophisticated algorithms that require some time getting acquainted with. While the use of numerous CMs allows inspection of data from different aspects of relative importance thus contributing to a more holistic picture, it also compounds the CMs the user needs to familiarize with. Consequently, new users may perceive this as a significant barrier to entry.

7.3.4 Conclusion

The high score from the evaluation (see Table 6.1) and positive qualitative evaluation (see Section 6.3.2) confirm that using CMs for visualization and interaction design of the visual analytics tool not only enhanced the tool's ability to identify key actors but also deepened users' understanding of how these actors evolved over time. While some limitations in the prototypical approach were identified, these do not undermine the overall effectiveness of the method. The positive outcomes demonstrate that integrating CMs into dynamic network visualizations can significantly enhance users' ability to explore complex network structures and track key actors over time, facilitating a deeper understanding of evolving relationships within the network.

CHAPTER 8

Conclusion

This chapter begins by outlining areas for future work and then summarizes findings and insights gained from this research effort.

8.1 Future Work

8.1.1 Modifying and Tuning CMs

The concept of importance in a network is inherently relative. While traditional CMs are effective due to their simplicity and ease of interpretation, it can be useful to modify them to capture more complex notions of importance. This research has demonstrated the potential of adjusting CMs based on domain-specific information, such as modifying CMs according to an artist's contribution in a given year, which provided valuable insights for art historians assessing an artist's success (see Section 4.2.3). Additionally, research by Taylor et al. explored the use of hyperparameters in the context of an *eigenvector-based* dynamic CM, offering further flexibility for analysis. By adjusting the strength of coupling between time layers, their method provides more control over how much influence from earlier periods persists into the future, making it useful for identifying nodes with sustained centrality in a network [TMC+17b].

However, overdoing these adjustments bears the risk of introducing excessive complexity, especially when dealing with novel CMs that are already sophisticated by themselves. Consequently, adding too many hyperparameters can make CMs harder to interpret. Future research could investigate how to leverage these hyperparameters in controlled ways that still allow for clear visualization and meaningful interpretation of importance while ensuring that a balance between complexity and usability is maintained.

8.1.2 Enhanced Temporal Visualization

The use of the dynamic CM, *change centrality* [FPA+12], has proven successful in enhancing the visualization of temporal changes in dynamic networks (see Section 6.3.2). While these developments are valuable, there remain many opportunities for further improvement in visual analytics through the exploration of additional dynamic CMs. An *eigenvector-based* dynamic CM highlights nodes with sustained influence across different periods, making it interesting for detecting long-term leaders or hubs [TMC+17b]. *Temporal dynamic-sensitive centrality* is particularly useful for understanding spreading dynamics in evolving networks because it accounts for both the network's structure and the timing of interactions [HY17]. Lastly, *dynamic centrality* takes a different approach by only considering other nodes over time if there exists a path that connects them, based on the intuition that influence requires an actual connection between nodes [LGK10]. Future research could take advantage of these dynamic CMs and explore their potential for enhancing visual analytics.

8.1.3 Versatile Data Handling

While dome-insights was evaluated on a specific dataset, it has demonstrated the potential to be applied across a wide range of dynamic social networks, much like Palladio[1]. As a result, a goal for future research could entail exploring effective ways to make the tool accessible directly through the browser without the need for a dedicated backend.

However, to achieve this, the tool must first support various data formats, enabling users to import different types of dynamic network datasets easily. Moreover, it would entail developing an intuitive, user-friendly interface that ensures that users can upload and work with their data without needing any technical knowledge. Furthermore, since such a tool would have to function without a backend, all computations, including the computations of CMs, need to run entirely in the frontend. This presents technical challenges, as real-time algorithmic computations of CMs would have to be performed efficiently within the browser to ensure a smooth user experience. Hence, future development could also focus on optimizing these computations to handle complex datasets while maintaining responsiveness and usability, ensuring that the tool remains user-friendly.

8.2 Summary

This research set out to determine whether CMs could enhance visual analytics for dynamic network visualization, aiming to improve insight-based analysis and aid in identifying key actors within evolving networks. To achieve this, we started by conducting a comprehensive state-of-the-art literature review, which directly informed the design and development of a centrality-based [VHW08] visual analytics prototype. Drawing on key insights from the reviewed literature, the design incorporated multiple CMs, leveraging

[1] Palladio: `hdlab.stanford.edu/palladio/`. Accessed on October 4, 2024.

their strengths to create a tool that allowed users to effectively explore network dynamics at both macro and micro levels.

The prototype developed in this research allows users to switch seamlessly between broad network overviews and detailed, node-specific analysis. A key feature is its ability to track the evolution of key actors over time, helping users identify shifts in node importance and detect structural changes. Additionally, by integrating a dynamic CM, the prototype provided users with a new way to detect shifts in network evolution through quantification.

Ultimately, the results of this study (see Section 6.3) demonstrate that the centrality-based prototype developed in this research successfully fulfilled the research objective of improving visual analytics for dynamic networks. While future work could explore further refinements, such as incorporating additional static and dynamic CMs and adding usability features to generalize the applicability of the application, this research nonetheless succeeded in demonstrating the significant role CMs can play in enhancing network visualization.

List of Figures

List of Tables

Bibliography

[AIP+18] Ademar T Akabane, Roger Immich, Richard W Pazzi, Edmundo RM Madeira, and Leandro A Villas. Distributed egocentric betweenness measure as a vehicle selection mechanism in vanets: A performance evaluation study. *Sensors*, 18(8):2731, 2018.

[BBDW17] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer graphics forum*, pages 133–159. Wiley Online Library, 2017.

[BEJ18] S.P. Borgatti, M.G. Everett, and J.C. Johnson. *Analyzing Social Networks*. SAGE Publications, 2018.

[BEPK24] Thomas Banafa, Susa Eräranta, Lasse Peltonen, and Marko Keskinen. Social network analysis of eu flood risk management plans: Case finland. *Journal of Flood Risk Management*, 17(1):e12945, 2024.

[BMB+20] Christina Bartosch, Nirmalie Mulloli, Daniel Burckhardt, Marei Döhring, Walid Ahmad, and Raphael Rosenberg. *The database of modern exhibitions (DoME): European paintings and drawings 1905-1915* , chapter 30, pages 423–434. Routledge, 2020.

[CM11] Carlos D Correa and Kwan-Liu Ma. Visualizing social networks. *Social network data analytics*, pages 307–326, 2011.

[CU23] Nazim Choudhury and Shahadat Uddin. Evolutionary features for dynamic link prediction in social networks. *Applied Sciences*, 13(5):2913, 2023.

[ELMS91] Peter Eades, Wei Lai, Kazuo Misue, and Kozo Sugiyama. Preserving the mental map of a diagram. Technical report, Technical Report IIAS-RR-91-16E, Fujitsu Laboratories, 1991.

[FPA+12] Paolo Federico, Jürgen Pfeffer, Wolfgang Aigner, Silvia Miksch, and Lukas Zenk. Visual analysis of dynamic networks using change centrality. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 179–183. IEEE, 2012.

[FSIK19] Takayasu Fushimi, Kazumi Saito, Tetsuo Ikeda, and Kazuhiro Kazama. Estimating node connectedness in spatial network under stochastic link disconnection based on efficient sampling. *Applied Network Science*, 4:1–24, 2019.

[HS12] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[HY17] Da-Wen Huang and Zu-Guo Yu. Dynamic-sensitive centrality of nodes in temporal networks. *Scientific reports*, 7(1):41454, 2017.

[Ing05] Stephen F Ingram. An interactive small world graph visualization. *Technical Report*, 2005.

[JC24] Charity S Jacobs and Kathleen M Carley. # whatisdemocracy: finding key actors in a chinese influence campaign. *Computational and Mathematical Organization Theory*, 30(2):127–147, 2024.

[KMSZ08] Daniel A Keim, Florian Mansmann, Andreas Stoffel, and Hartmut Ziegler. Visual analytics, 2008.

[LGK10] Kristina Lerman, Rumi Ghosh, and Jeon Hyung Kang. Centrality metric for dynamic networks. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, pages 70–77, 2010.

[LS23] Yose Lee and Ducksu Seo. Identifying relationship between regional centrality and poi facilities: A case study of seoul metropolitan area. *ISPRS International Journal of Geo-Information*, 13(1):12, 2023.

[LV20] Stephanie C Leone and Paul A Vierthaler. Innocent x pamphilj's architectural network in rome. *Renaissance Quarterly*, 73(3):897–952, 2020.

[MHZ23] Guojie Ma, Jinxing Hu, and Riquan Zhang. Spatial-temporal distribution and coupling relationship of high-speed railway and economic networks in metropolitan areas of china. *Land*, 12(6):1193, 2023.

[MRBC19] Cornelia Metzig, Oliver Ratmann, Daniela Bezemer, and Caroline Colijn. Phylogenies from dynamic networks. *PLoS computational biology*, 15(2):e1006761, 2019.

[New03] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[Nor06] Chris North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.

[Oko15] Chitu Okoli. A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37, 2015.

[PP24] Wichaya Peechapat and Nattapong Puttanapong. Collaboration and competition: A social network analysis of thailand's music industry. *Economies*, 12(2):45, 2024.

[PR08] Mark Petticrew and Helen Roberts. *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons, 2008.

[Rod19a] Francisco Aparecido Rodrigues. Network centrality: an introduction. *A mathematical modeling approach from nonlinear dynamics to complex systems*, pages 177–196, 2019.

[Rod19b] Francisco Aparecido Rodrigues. Network centrality: an introduction. *A mathematical modeling approach from nonlinear dynamics to complex systems*, pages 177–196, 2019.

[RT81] Edward M. Reingold and John S. Tilford. Tidier drawings of trees. *IEEE Transactions on software Engineering*, pages 223–228, 1981.

[SBPS23] Lea F Stöber, Marius Boesino, Andreas Pyka, and Franziska Schuenemann. Bioeconomy innovation networks in urban regions: The case of stuttgart. *Land*, 12(4):935, 2023.

[SFS+19] Martin Sterchi, Céline Faverjon, Cristina Sarasua, Maria Elena Vargas, John Berezowski, Abraham Bernstein, Rolf Grütter, and Heiko Nathues. The pig transport network in switzerland: Structure, patterns, and implications for the transmission of infectious diseases between animal holdings. *PLoS One*, 14(5):e0217974, 2019.

[SFTN18] Mohammed Saqr, Uno Fors, Matti Tedre, and Jalal Nouri. How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PloS one*, 13(3):e0194777, 2018.

[Shn03] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pages 364–371. Elsevier, 2003.

[SIA23] Alexandros Z Spyropoulos, Evangelos Ioannidis, and Ioannis Antoniou. Interoperability and targeted attacks on terrorist organizations using intelligent tools from network science. *Information*, 14(10):580, 2023.

[SND05] Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, 11(4):443–456, 2005.

[SSM+23] Alexandros Samalis, Alexandros Z Spyropoulos, Georgios C Makris, Charalampos Bratsas, Andreas Veglis, Vassilis Tsiantos, Anthoula Baliou, Emmanouel Garoufallou, and Anastasios Ventouris. Data journalism and network theory: A study of political communication through x (formerly twitter) interactions. *Journalism and Media*, 4(4):1141–1168, 2023.

[Sta14] John Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 46–53, 2014.

[SZC18] Bilong Shen, Weimin Zheng, and Kathleen M Carley. Urban activity mining framework for ride sharing systems based on vehicular social networks. *Networks and Spatial Economics*, 18:705–734, 2018.

[TEWG23] Shantanu Tilak, Marvin Evans, Ziye Wen, and Michael Glassman. Social network analysis as a cybernetic modelling facility for participatory design in technology-supported college curricula. *Systemic Practice and Action Research*, 36(5):691–724, 2023.

[TMC+17a] Dane Taylor, Sean A Myers, Aaron Clauset, Mason A Porter, and Peter J Mucha. Eigenvector-based centrality measures for temporal networks. *Multiscale Modeling & Simulation*, 15(1):537–574, 2017.

[TMC+17b] Dane Taylor, Sean A Myers, Aaron Clauset, Mason A Porter, and Peter J Mucha. Eigenvector-based centrality measures for temporal networks. *Multiscale Modeling & Simulation*, 15(1):537–574, 2017.

[TS19] Zhasmina Tacheva and Natalie Simpson. Social network analysis in humanitarian logistics research. *Journal of Humanitarian Logistics and Supply Chain Management*, 9(3):492–514, 2019.

[VHW08] Frank Van Ham and Martin Wattenberg. Centrality based visualization of small world graphs. In *Computer Graphics Forum*, volume 27, pages 975–982. Wiley Online Library, 2008.

[WAM+18] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):491–500, 2018.

[Was94] Stanley Wasserman. Social network analysis: Methods and applications. *The Press Syndicate of the University of Cambridge*, 1994.

[YFDH01] Ka-Ping Yee, Danyel Fisher, Rachna Dhamija, and Marti Hearst. Animated exploration of graphs with radial layout. In *Proc. IEEE InfoVis 2001*, pages 43–50, 2001.

[YHC+22] Zhenkai Yang, Yixin Hua, Yibing Cao, Xinke Zhao, and Minjie Chen. Network patterns of zhongyuan urban agglomeration in china based on baidu migration data. *ISPRS International Journal of Geo-Information*, 11(1):62, 2022.

[YKSJ08] Ji Soo Yi, Youn-ah Kang, John T Stasko, and Julie A Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 1–6, 2008.

[YMC21] Yanan Yu, Marguerite Moore, and Lisa P Chapman. Social network analysis of an emerging innovation: direct-to-garment printing technology. *Journal of Fashion Marketing and Management: An International Journal*, 25(2):274–289, 2021.

[YZGW22] Zhangbo Yang, Jiahao Zhang, Shanxing Gao, and Hui Wang. Complex contact network of patients at the beginning of an epidemic outbreak: an analysis based on 1218 covid-19 cases in china. *International Journal of Environmental Research and Public Health*, 19(2):689, 2022.

Appendix

Appendix A.1: Heuristic-Based Evaluation Questions

Table 1: Heuristic-Based Evaluation Questions from Wall et al.'s Value-Driven Visualization Evaluation [WAM+18]

Insight: the insights and insightful questions it spurs	
G1	**The visualization facilitates answering questions about the data**
Q1	The visualization exposes individual data cases and their attributes
Q2	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables
Q3	The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases
G2	**The visualization provides a new or better understanding of the data**
Q4	The visualization helps generate data-driven questions
Q5	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values
G3	**The visualization provides opportunities for serendipitous discoveries**
Q6	The visualization provides useful interactive capabilities to help investigate the data in multiple ways
Q7	The visualization shows multiple perspectives about the data
Q8	The visualization uses an effective representation of the data that shows related and partially related data cases
Confidence: the confidence it inspires about the data and its domain	
G4	**The visualization helps avoid making incorrect inferences**

Q9 The visualization uses meaningful and accurate visual encodings to represent the data

Q10 The visualization avoids using misleading representations

G5 The visualization facilitates learning more broadly about the domain of the data

Q11 The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes

G6 The visualization helps understand data quality

Q12 If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues

Essence: the overall essence of the data it conveys

G7 The visualization provides a big picture perspective of the data

Q13 The visualization provides a comprehensive and accessible overview of the data

Q14 The visualization presents the data by providing a meaningful visual schema

G8 The visualization provides an understanding of the data beyond individual data cases

Q15 The visualization facilitates generalizations and extrapolations of patterns and conclusions

Q16 The visualization helps understand how variables relate in order to accomplish different analytic tasks

Time: the time savings the visualization provides

G9 The visualization affords rapid parallel comprehension for efficient browsing

Q17 The visualization provides a meaningful spatial organization of the data

Q18 The visualization shows key characteristics of the data at a glance

G10 The visualization provides mechanisms for quickly seeking specific information

Q19 The interface supports using different attributes of the data to reorganize the visualization's appearance

Q20 The visualization supports smooth transitions between different levels of detail in viewing the data

Q21 The visualization avoids complex commands and textual queries by providing direct interaction with the data representation

Appendix A.2: Evaluation Results Overview

Table 2: Survey results from five participants (P1 to P5), each of whom responded to 21 different questions (Q1 to Q21) [WAM+18] on a Likert scale from 1 (Strongly disagree) to 7 (Strongly agree). The responses are color-coded: red tones indicate lower ratings and green tones indicate higher ratings.

	Insight								**Confidence**				**Essence**				**Time**				
	G1			**G2**		**G3**			**G4**		**G5**	**G6**	**G7**		**G8**		**G9**		**G10**		
Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P1	6	6	6	6	6	6	6	6	6	5	6	–	6	6	6	5	6	6	5	6	6
P2	7	7	7	7	6	6	6	7	7	6	6	5	6	6	6	7	6	5	6	6	7
P3	6	7	6	7	7	7	6	6	6	5	6	4	7	6	6	6	6	6	6	6	7
P4	6	6	6	7	7	7	5	6	6	6	6	–	7	7	6	6	4	6	6	7	7
P5	5	7	7	7	7	7	7	7	7	6	7	7	6	7	7	6	6	6	7	6	7
μ	6.0	6.6	6.4	6.8	6.6	6.6	6.0	6.4	6.4	5.6	6.2	5.3	6.4	6.4	6.2	6.0	5.6	5.8	6.0	6.2	6.8
μ		6.33		6.70			6.33		6.00		6.20	5.30	6.40		6.10		5.70			6.33	
μ					6.45						5.84				6.25				6.01		
μ											6.14										

The symbol '–' indicates "Not applicable".

Appendix A.3: Evaluation Results Components

Table 3: Survey results on *Insight*, showing participant ratings on how the visualization facilitates obtaining insights through targeted search and incidental insights through browsing [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization facilitates answering questions about the data	The visualization exposes individual data cases and their attributes	6	7	6	6	5	6.00
	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables	6	7	7	6	7	6.60
	The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	6	7	6	6	7	6.40
The visualization provides a new or better understanding of the data	The visualization helps generate data-driven questions	6	7	7	7	7	6.80
	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	6	6	7	7	7	6.60
The visualization provides opportunities for serendipitous discoveries	The visualization provides useful interactive capabilities to help investigate the data in multiple ways	6	6	7	7	7	6.60
	The visualization shows multiple perspectives about the data	6	6	6	5	7	6.00
	The visualization uses an effective representation of the data that shows related and partially related data cases	6	7	6	6	7	6.40

Table 4: Survey results on *Confidence*, presenting participant ratings on how the visualization supports accurate interpretation and understanding of data quality [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization helps avoid making incorrect inferences	The visualization uses meaningful and accurate visual encodings to represent the data	6	7	6	6	7	6.40
	The visualization avoids using misleading representations	5	6	5	6	6	5.60
The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	6	6	6	6	7	6.20
The visualization helps understand data quality	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues	–	5	4	–	7	5.33

The symbol '–' indicates "Not applicable".

Table 5: Survey results on *Essence*, summarizing participant ratings on the visualization's ability to provide a big-picture perspective and facilitate generalizations beyond individual data cases [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization provides a big picture perspective of the data	The visualization provides a comprehensive and accessible overview of the data	6	6	7	7	6	6.40
	The visualization presents the data by providing a meaningful visual schema	6	6	6	7	7	6.40
The visualization provides an understanding of the data beyond individual data cases	The visualization facilitates generalizations and extrapolations of patterns and conclusions	6	6	6	6	7	6.20
	The visualization helps understand how variables relate in order to accomplish different analytic tasks	5	7	6	6	6	6.00

Table 6: Survey results on *Time*, reflecting participant ratings on how efficiently the visualization allows for rapid browsing and the retrieval of specific data [WAM+18].

Guideline	Question	P1	P2	P3	P4	P5	Avg
The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	6	6	6	4	6	5.60
	The visualization shows key characteristics of the data at a glance	6	5	6	6	6	5.80
The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance	5	6	6	6	7	6.00
	The visualization supports smooth transitions between different levels of detail in viewing the data	6	6	6	7	6	6.20
	The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	6	7	7	7	7	6.80

Appendix B: Qualitative Questions & Responses

Q1. When observing a network over time, do you find that the novel approach highlights important changes more clearly than the traditional approach?

P1 Absolutely, changes are visible much faster and in a clearer way.

P2 I like the novel approach but would need to see it in the data better (now I am searching for color in the nods and get confused by colors of countries).

P3 Yes, I believe the novel approach does highlight important changes more clearly than the traditional approach. It offers a more precise and efficient way to track changes over time.

P4 Yes.

P5 Yes, although as mentioned in evaluation, the additional info on artists countries of origin etc can be distracting from the main point of the visualisation.

Q2. How well does the novel approach help you understand the evolution of key nodes in a network? Does it offer insights that the traditional approach might miss?

P1 Very well. While traditional approach are prone to error and missing data points, this novel approach seems much more reliable in making evolutions visible.

P2 I think I don't get there easily to understand the evolution of key nods.

P3 The novel approach is more effective than the traditional method in highlighting subtle but significant transformations that might otherwise be missed when manually switching between years. For example, it becomes particularly interesting to observe an artist was playing a significant role in a certain artist group in a particular year, which is difficult to discover from traditional approach. As I explore the model and visualisations, I find myself considering new questions, which leads to the discovery of fresh dynamics in the network. Overall, this new method offers a deeper understanding of network evolution, making it easier to spot key trends and emerging patterns that the traditional approach might overlook.

P4 Yes. More dynamic and quicker insights into a broad time span

P5 Allows for broader historical questions eg) looking at the differences in networks before and after the outbreak of WW1

Q3. Do you think using the novel approach alongside the traditional approach provides a more complete understanding of network changes, or do they tend to overlap too much in the insights they provide?

P1 For more in depths information about the individual data presented, traditional approaches will stay relevant, probably without creating too much of an overlap.

P2 I think it's valid to apply both at the same time. I would just need to see the change centrality measure better.

P3 To me, the new approach provides unique insights that the traditional approach do not.

P4 a more complete and complex understanding!

P5 I think for the sake of clarity they should be visualized separately, see above. Also maybe a slightly different way of taking into account each artist´s missing years.. so that displaying changes between the whole 10 years does not result in only black dots for lack of data. As it is right now, to show any meaningful change for the span of 10 years, all artists have to have exhibited every year, which is unlikely.

Q4. If you were to explain the benefits of the novel approach to a colleague, what would you highlight as its main advantage over the traditional approach?

P1 The speed at which a bird's eye view of larger pattern becomes possible and the flexibility in questioning the data set.

P2 The color coding like a heat map.

P3 To apply the new method has mainly two advantages for me: 1. Thought-provoking: Unlike the traditional method of manually switching between years, the novel approach highlights key shifts, prompting the generation of new, data-driven research questions. 2. Discovery of complex dynamics: By utilising centrality measures, the novel method enables multi-layered comparisons, uncovering subtle dynamics and patterns over time that are difficult to detect through the traditional, year-by-year visual inspection.

P4 That it is a quick and convenient way to understand a rich data set quickly (every database should have something like this!)

P5 Allows for intuitive visualization of temporal changes in the network, year by year.

Appendix C: Cheat Sheet

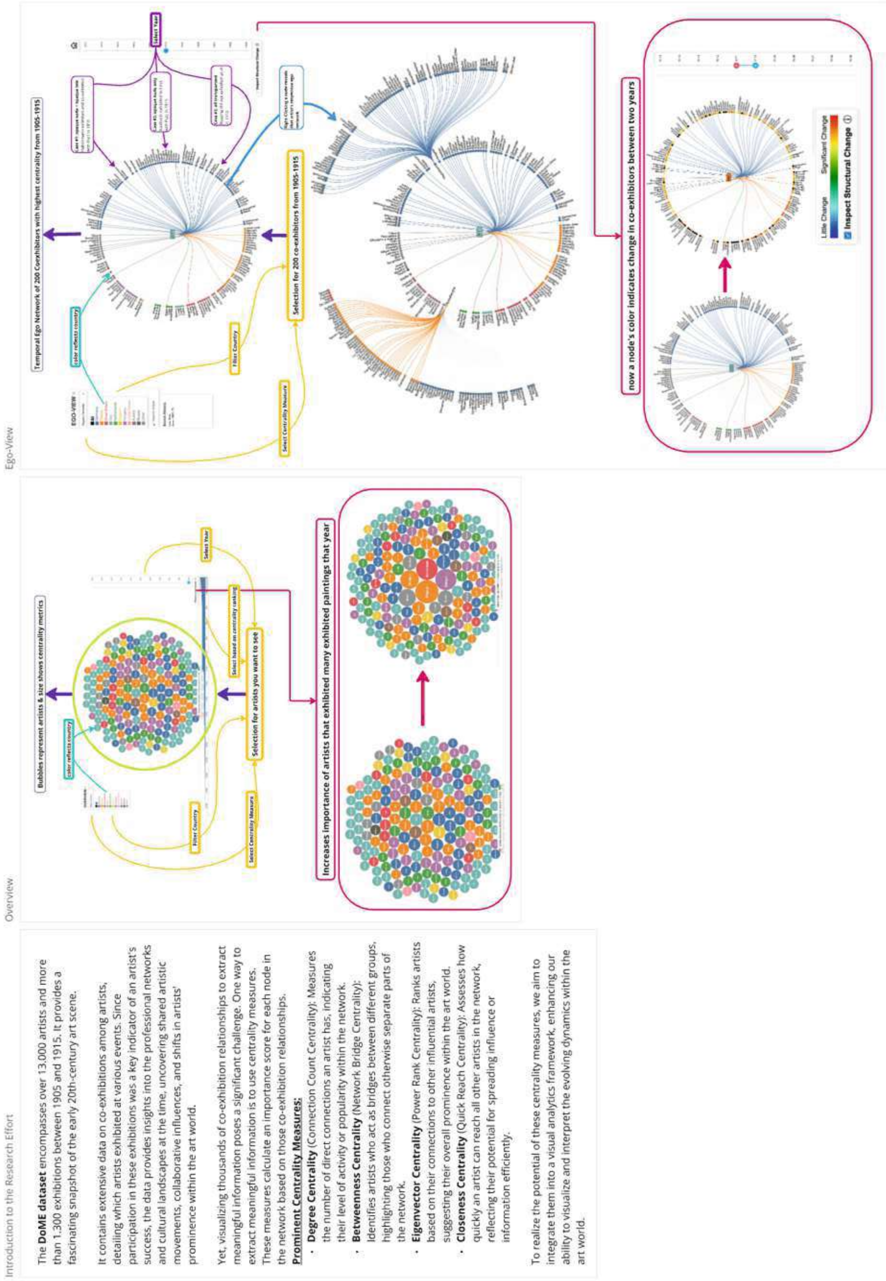

Figure 1: Cheat sheet used during the evaluation by the raters