TU WIEN Informatics

Network Analysis on the Austrian Media Corpus:

Examining measures of co-occurrence between entities in Austrian media

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Logic and Computation

eingereicht von

Gabriel Grill, Bsc.
Matrikelnummer 01025120

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Dipl.-Ing. Dr.techn. Hannes Werthner
Mitwirkung: Mag.[a]rer.nat. Dr.[in]techn. Julia Neidhardt

Wien, 31. August 2023

Gabriel Grill	Hannes Werthner

Informatics

Network Analysis on the Austrian Media Corpus

Examining measures of co-occurrence between entities in Austrian media

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Logic and Computation

by

Gabriel Grill, Bsc.
Registration Number 01025120

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Hannes Werthner
Assistance: Mag.^{a}rer.nat. Dr.^{in}techn. Julia Neidhardt

Vienna, August 31, 2023

Gabriel Grill	Hannes Werthner

Erklärung zur Verfassung der Arbeit

Gabriel Grill, 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, 31. August 2023

Gabriel Grill

Acknowledgements

First and foremost, I want to thank my advisors, Dr.$^{\text{in}}$ Julia Neidhardt and Prof. Hannes Werthner, whose support and guidance enabled me to finish this thesis. I also want to thank Dipl. Ing. Matej Ďurčo and Mag. Hannes Pirker from the Austrian Centre for Digital Humanities for their insights into handling the Austrian Media Corpus (AMC) and for helping me when needed. I also want to thank Prof. Hideaki Takeda, who provided valuable advice while I was at the National Institute of Informatics in Tokyo. Finally, I want to thank my family, who provided emotional support throughout my studies.

Kurzfassung

Quantitative Forschung zu Zeitungen kann Einblicke in Berichterstattungsmuster liefern und damit einen öffentlichen Diskurs ermöglichen. Printmedien sind in demokratischen Gesellschaften unverzichtbar, daher bleibt ihre Erforschung wichtig. Diese Arbeit untersucht österreichische Berichterstattung mit netzwerkbasierten Methoden und diskutiert die Eignung eines solchen Ansatzes für die Analyse von Zeitungen. Die Arbeit trägt zu wissenschaftlichen Debatten über die Vorteile und Risiken des Einsatzes solcher Methoden bei. Wir führen eine vergleichende Analyse von sechs Zeitungen durch (Der Standard, Die Presse, Österreich, Die Kronen Zeitung, Kurier, Die Heute) und diskutieren Unterschiede in der Berichterstattung während der österreichischen Präsidentschaftswahlen 2016.

Dies ist die erste Forschungsarbeit die einen netzwerkbasierten Ansatz für den Austrian Media Corpus (AMC) anwendet, eine vollständige und einzigartige Sammlung, die die letzten drei Jahrzehnte österreichischer Medienberichterstattung umfasst. Wir wenden Natural Language Processing Methoden an um textliche Referenzen zu Entitäten zu erkennen und daraus, welche die Knoten in den Netzwerken darstellen. Wir evaluieren mehrere Erkennungsalgorithmen für Entitäten anhand von annotierten Artikeln. Die beste Methode war eine Kombination aus einer Erkennung basierend auf Wikidata-Einträgen und einem open-source machine learning Modell. Wir konstruieren die Netzwerke indem extrahierten Entitäten und erkannte Begriffe in Beziehung gesetzt werden, wenn sie in einem Satz gleichzeitig vorkommen. Wir wenden verschiedene Algorithmen auf die resultierenden Netzwerke an, um Entitäten nach Relevanz zu ordnen und Knoten zu gruppieren, um Themen während der Wahl zu erkennen.

Wir präsentieren mehrere deskriptive Statistiken zu Veröffentlichungsmustern und dem Auftreten von Entitäten in den Zeitungen, die geschlechtsspezifische Benachteiligungen und die am häufigsten genannten Präsidentschaftskandidaten aufzeigen. Im Vergleich zu diesen Ergebnissen, zeigt unser netzwerkbasierter Ansatz andere Eigenschaften der Berichterstattung auf. Abschließend weisen wir auf Probleme netzwerkbasierender Methoden hin, wie z. B. Flexibilität bei der Parametrisierung und Unübersichtlichkeit von Netzwerkvisualisierungen. Wir argumentieren, dass diese Herausforderungen ein zweischneidiges Schwert sind, da beispielsweise Flexibilität Forschenden auch mehr Möglichkeiten geben kann für Erkundung von Daten und qualitative Interpretation.

Abstract

The quantitative study of news media can provide insights into reporting patterns and enable public discourse. Print media is essential in democratic societies, so its study remains important. This thesis examines Austrian reporting using network-based methods and unpacks the suitability of such an approach for news analysis. We seek to contribute to scholarly debates around the benefits and risks of using such methods to make sense of reporting. We conduct a comparative analysis of six news outlets (Der Standard, Die Presse, Österreich, Die Kronen Zeitung, Kurier, Die Heute) and discuss differences in reporting during the 2016 Austrian presidential elections.

This is the first research effort applying a network-based approach to the Austrian Media Corpus (AMC), a complete and unique collection encompassing the last three decades of Austrian media coverage. We use natural language processing to extract an expressive subset of named entities representing network nodes. Several entity recognition schemes are evaluated based on a set of labeled articles. An approach combining named entity linking based on a Wikidata dictionary with an open-source recognition model performed best. The networks are constructed by relating the extracted entities and certain terms when they co-occur in a sentence. We apply various algorithms to the resulting networks to rank entities according to relevance and cluster nodes to detect themes during the election.

We provide several descriptive statistics on publishing patterns and the occurrence of entities in the newspapers, revealing gender bias and the presidential candidates most mentioned. Our network-based approach reveals differences in reporting compared to results based on counting mentions. However, we also point out issues of these methods, such as flexibility in parameterization and messy visualizations. We argue these challenges are a double-edged sword as, for example, flexibility may also give researchers agency to enable more exploration and qualitative interpretation.

Contents

CHAPTER 1

Introduction

Print media has an essential role in the co-production of public discourse in democratic societies [Hab91]. Its study can aid in reflexive meta-debates on media industry practices to make them more accountable. Notably, tabloid media and sensational reporting have received considerable attention in the wake of the so-called “Brexit” [Sea16, WF22], the election of former US President Trump [Aza16, Kel16, WSP+16] and the rise of populist right-wing parties in Europe [CG17, FMS+20]. Scholars suggest that populist movements are able to get attention and coverage in the news through various communication tactics, such as making frequent provocative and sensationalist statements [Pic16]. Some studies even indicate a possible trend toward a so-called “tabloidization of news” [WF22]. This troubles traditional classifications and illustrates how populist and divisive messaging can spread across outlets also with different standards and values. Beyond broadly disseminating their opinions, these movements also aim at discrediting certain high-quality media outlets perceived as undermining their narratives. For instance, the Pegida movement frequently raised allegations of biased reporting against reputable media organizations [HH16]. These issues highlight a need for more data-oriented research into media coverage and how it portrays events, public debates, and other happenings.

We examine Austrian reporting using network-based methods and develop a proof-of-concept implementation for data exploration. Our approach seeks to identify relevant entities, topics, and connections between them in co-occurrence text networks extracted from news articles. These text networks are comprised of nodes representing lemmatized words and named entities. The relations between them are co-occurrences in a sentence. The results of our network analysis can be used to compare different news outlets and, thereby, possibly give insights into the study of media bias [Ben16]. Such an analysis may encourage media creators to reflect on their work and the public to be more informed on underlying systematic biases of news coverage.

This thesis provides a quantitive analysis of media reporting during the 2016 Austrian presidential election. In this introductory chapter, we elaborate on the motivation behind this work and give some examples of possible applications. Then, we explain the research problem, highlight our methodological approach, and give an overview of the structure of the work. We conducted this project with the support of the E-Commerce Group, Institute of Information Systems Engineering, TU Wien, and the Austrian Center for Digital Humanities (ACDH), Austrian Academy of Sciences. Moreover, during a three-month internship at the National Institute of Informatics in Tokyo (Japan), the named entity detection scheme was refined.

Network analysis is a powerful method to get insights into structural properties of complex systems, such as the centrality of a node in a network. It has been applied in many scholarly contexts such as social science, biology, finance, medicine, and more [New03]. In this work, we apply network analysis techniques to examine relational aspects of entities extracted from articles published by Austrian media. This thesis is the first research effort applying a network-based approach to the Austrian Media Corpus (AMC) [DMPR14], a complete and unique collection encompassing the last three decades of Austrian media coverage. We further want to compare different modeling approaches and discuss their implications. We use data mining, Natural Language Processing (NLP), and network analysis techniques to illustrate an approach to news analysis. This work is situated in the fields of network science, data mining, digital humanities, and news analytics.

1.1 Problem statement

The contributions of this thesis belong to two levels. We gain insights into the benefits and issues of using co-occurrence networks for news analytics and obtain concrete statements about a news event by conducting an empirical analysis. We illustrate this approach through a proof of concept implementation. The empirical analysis is conducted through a case study encompassing reporting on the 2016 Austrian presidential elections. We specifically focus on the following research questions:

(1) To what extent can we make sense of the quality of extracted named entities and relationships used to construct co-occurrence networks?

(2) Which methods are appropriate to detect relevant entities and topics in the extracted networks?

(3) What can a co-occurrence network-based approach reveal about reporting on the 2016 Austrian presidential election?

(4) What are the benefits and risks of a co-occurrence network-based approach to news and text analytics?

This thesis discusses these questions, provides conclusions, and shows possibilities for future work.

1.2 Methodological approach

We conducted this project using the Design Science Research Framework [HMPR04]. Our artifacts are methods and algorithms. Problem relevance is given by the importance of better understanding the news reporting to enable public and reflexive discourse on media. Research rigor is given by its grounding in network and computer science theories. The models and methods are evaluated through quantitative testing and qualitative interpretation of results. For example, using appropriate quantitative measures such as modularity to substantiate the strength of identified communities. The methodological approach was comprised of the following steps:

1. **Literature review and tool search**
   First, a survey of the literature on news analysis and related NLP and network analysis techniques was conducted. We also identified and tried out different available open-source tools and libraries that could be employed in this project for cleaning, processing, storing, and analyzing textual and network data.

2. **Extraction of entities and co-occurrence relations**
   Next, articles of the AMC corpus were cleaned and parsed using NLP techniques to extract named entities and different types of co-occurrence relationships. We devised and tested different named entity detection schemes to identify Austrian key political actors in the texts. The Austrian Center for Digital Humanities provided a human-labeled set of articles part of the AMC to aid in our evaluation of the named entity detection schemes. The consequent tests enabled us to choose between different frameworks, methods, and configurations. We ultimately decided on an approach that combines a newly developed entity matching scheme that uses Wikidata as a data source with an available German language open-source entity detection model. This step required multiple iterations. We devised this approach because various names and synonyms of influential Austrian politicians, like Alexander Van der Bellen, were challenging to identify with available open-source models often trained on German and not Austrian textual data.

3. **Descriptive corpus statistics**
   We then calculated various descriptive statistics on the corpus to make sense of its composition. These gave insights into reporting frequency on certain issues, thematic areas, and key actors.

4. **Network modeling and analysis**
   The extracted named entities and relationships were then stored in a graph database to enable efficient processing of the networks. Furthermore, we set up a service to generate networks based on various parameters and used it to analyze different co-occurrence networks. In these networks, the nodes represent extracted named entities or words from articles and the relations co-occurrence in a sentence. We used various algorithms and measures in our analysis of the resulting networks.

5. **Exploration of networks and empirical analysis**
   We developed a web application and a Gephi connector to enable fast querying and visualization of co-occurrence networks and related statistical measures. We used these tools to examine reporting on the 2016 Austrian presidential election across news outlets and timeframes. This process involved tweaking parameters to generate and explore different visualizations and choose those deemed most expressive. We employed a community detection algorithm to identify themes in news coverage and a node centrality algorithm to rank political actors according to a measure of relevance. We further refined the pipeline based on the interpretation of the results and small tests, restarting this iterative process at step two when we noticed issues.

1.3 Structure of the work

The next chapter provides an overview of the state of the art in fields intersecting with this work, focusing on existing approaches. The following chapters explain methods, models, and concepts used in this work, present the suggested solution, provide a critical reflection on the results, list open issues and conclude with a summary of this work and discuss possible future work. We provide a short outline of each chapter below:

- **Literature review**
  This chapter briefly surveys research on computational approaches to media analysis. Additionally, we provide a short review highlighting important entity and relationship extraction and network analysis methods.

- **Methodological approach**
  We first list and shortly explain programming languages, frameworks, and tooling applied in this work. We illustrate the specific implementation details and discuss underlying design decisions. The second part of the chapter outlines the NLP and network analysis pipeline and briefly introduces the methods used.

- **Evaluation of named entity recognition**
  We present an evaluation of our named entity recognition scheme based on several coded articles.

- **Austrian Media Corpus**
  In this chapter, we introduce the AMC and present descriptive statistics to illustrate the structure and content of articles as part of the time frame for our case study.

- **Network analysis of the Austrian presidential election**
  We present the empirical results of the network-based media analysis of the 2016 Austrian presidential election.

- **Conclusion and future work**
  In this chapter, we conclude with a summary of the contributions and results of this work, discuss implications, and provide an outlook for future work.

CHAPTER 2

Literature review

In this chapter, we discuss state-of-the-art literature related to the computational analysis of news articles. The review focuses on network-based approaches to news analytics but also discusses more general approaches in the realms of text network analysis and natural language processing applicable to the domain of interest of this thesis.

2.1 News analysis

There are various quantitive approaches to studying media with longstanding, scholarly histories and traditions [BB12, Jen20]. The focus of this literature review is scoped around recent computational approaches. More specifically, the field of news analytics and efforts to measure properties of aggregated news stories. The objects of analysis for this approach are usually unstructured texts, which can be enriched with semantic information, for example, through named entity recognition. Consequently, text mining and information retrieval literature are also relevant to this project.

Our 2016 Austrian presidential election analysis covers September 2015 to the end of December 2016. The framing of Europe's 2015 refugee crisis in six Austrian media outlets, "three national quality papers (Der Standard, Die Presse, Salzburger Nachrichten), and three tabloid newspapers with a regional focus on Eastern Austria (Kronenzeitung, Kurier, Heute)," has been examined in [GB17]. The work encompasses an overlapping timeframe as articles were studied between January 2015 and January 2016. The authors identified news articles covering the refugee crisis, used various methods to filter words and count co-occurrences in those articles, and then applied Principle Component Analysis (PCA) [Jol86] to identify clusters, which were understood as representing specific news frames. The coverage of these frames in different news outlets was then analyzed over time. Various computational methods for frame analysis have been explored in different outlets and contexts. For example, news coverage of nuclear power issues was examined using cluster and sentiment analysis [BVd16]. The clusters were identified through the k-means method [Scu10]. Frame analysis was also used to show how fossil fuel companies use public communication to minimize risks of climate change [SO21]. The authors used a mixed method approach combining quantitative approaches like Latent Dirichlet Allocation (LDA) [BNJ03] with qualitative analysis. Such interdisciplinary approaches are promising and are actively being developed and refined, such as person-oriented framing analysis [Ham23]. More recent research efforts are concerned with evaluating the reliability and tradeoffs of different computational approaches to frame analysis [EHLB23].

Various computational techniques were used for all kinds of analysis on reporting. For example, researchers aimed to model and predict the popularity of articles based on a few parameters like title and a short description of the article [HFC12]. The issue of media bias has also received considerable attention [NSZ+15, HDG19, DSMM23]. Differences in reporting between online and print news were examined through techniques such as sentiment analysis and named entity detection [BT17]. The media tonality during the final weeks of the 2013 Austrian national parliamentary campaign was explored in [HJ17]. The authors found a slightly more negative tone whenever a political actor is mentioned in contrast to a party but stated that further study would be required to constitute this result as definitive evidence. Topic modeling techniques, such as LDA, have been applied to get insights into trends and patterns in news corpora [JvW16].

A study [FAL+13] based on 500 different news outlets, quantified readability, subjectivity, and gender imbalances in reporting through a specifically trained Support Vector Machine (SVM) [CS00]. The authors also examined relations between identified topics and the popularity of articles. These measures were also used to compare a selection of news outlets and classify them into tabloids and broadsheets based on the topics they cover. Gender imbalance in news articles was also examined in [JLS+16]. Measuring the amount mentions of entities in the texts, they found a persistent gender gap in various topics. Their second finding was that the gap is lower when analyzing the number of women shown in images but still persistent. In [HMW14], the mentions of entities such as politicians, countries, and organizations were visualized over time to highlight their coverage. The entities were linked to publicly available knowledge bases. Recently, approaches using word embeddings have also received considerable attention. For example, researchers used word embeddings to compare the portrayal of different groups by media [WF22] or to uncover stereotyping in reporting [KTR21].

Various frameworks for analysis of large amounts of news articles have been developed over time, such as Lydia [LKS05], the European Media Monitor [AV09] and Noam [FAT+11]. In particular, the Global Data on Events, Location and Tone (GDELT) Project [LS13], which monitors news worldwide, has been considerably adopted [VPM+20, HNI21, OAO22, HNA23] as a tool for retrieving information about entities and important events. A recent survey highlights various methods developed to identify events and narratives surrounding them [KNMN23].

Researchers also used co-occurrence networks in news analysis [SG18, KRK20, CHL22, MA23]. Various network analysis methods were applied in [ÖCB08] to examine a network of manually identified persons extracted from the Reuters-21578 corpus. The authors also used node ranking techniques to measure the influence of persons, and the results were compared with prominence measures from Wikipedia. Similarly, in [TRv15], named entity co-occurrence networks were examined to identify elites. Another approach to co-occurrence network-based analysis was published by the ECB [RS16] to identify and study connections between banks and evaluate system risk. In [DCT12], co-occurrences between people, organizations, locations, resources, and conflicts were extracted from the news outlet Sudan Tribune to construct networks. In the next step, the change of various network metrics was plotted over time, and rankings of important entities were generated. They found that a set of key entities was staying robust on top over time in the rankings. The structure of media occurrences was studied based on entity co-occurrence networks in [TRv16]. Their results indicate that reporting focuses on people already prominent in the news.

Networks with more complex relations were also used in investigations of news reporting [Seg21, SVRS22]. In [SVC15, SDFFC15], an approach to automate Qualitative Narrative Analysis was used to make sense of the contents of articles written during the 2012 US elections. Narrative networks were generated by extracting Subject-Verb-Object Triplets, whereas subjects and objects in a sentence represented nodes and tonality verbs links between named entities. These semantic links were split into positive and negative connections, forming two network types. Since the nodes were named entities, social network analysis methods were applied. In [LSTC17] similarly, narrative network analysis was applied to get insights into institution and person networks extracted from local British newspapers between 1800 and 1950. Another application area of network-based approaches is the analysis of news citations [SG15, SSG18]. Researchers also analyzed news spread across different outlets and social media, giving insights on news cycles [TFA16].

Beyond an interest in journalism and media studies, a significant body of work on news analytics is centered on applications to the finance industry to better and faster understand relations between news and markets [AR23]. The goal is often to predict market movements through indicators such as sentiment. A handbook on news analytics in finance was published over a decade ago [MM11]. Nowadays, analyzing news texts is considered a normal part of the repertoire of data scientists in finance [CRRS21]. In the last couple of years, a lot of attention was also given to the automated detection of "fake news" online as highlighted by numerous works such as [AG17, CRC15, GMK$^+$18, Wan17, ZZ20, MSA22]. Ultimately, news analytics is a vibrant field where network-based approaches also received significant attention.

2.2 Network-based text analysis

In this section, we survey literature on graph or network-based approaches to text analysis. The previous section already covered several efforts in the realm of news analytics. In [SK14], a survey of graph-based text representations is given. Consolidating this review with our own, we distinguish networks in this section based on the choice of features extracted from the texts and used to construct nodes and relations.

In text networks, nodes usually represent a filtered set of words occurring in texts, for instance, narrowed through a stop-word list. Words are often chosen based on their assigned part-of-speech tags [MT04] or whether they were recognized as named entities [LSTC17, TRv16]. Furthermore, nodes often represent different words when they refer to either the same or similar concepts [Hen04, DC05, JS07, YSJ14, Par19, THH20] or (named) entities [DCT12, SVC15, SERS22]. Thus, NLP approaches such as co-reference resolution, entity linking, anaphora resolution, lemmatization, and stemming are applied to construct nodes. The performance of these methods influences readability, as networks with fewer nodes are easier to grasp and understand. The quality of the results of later network analysis methods can also improve since well-consolidated nodes yield a more concise and semantically rich representation of the underlying texts. Previous work has shown that aggregating synonyms can improve the performance of topic [PK15] and keyword extraction [Alr21]. In text networks, nodes can also represent complete sentences [YKY08], documents [YKY08] or topics [GOB+12]. The relationships between them are often measures of similarity, association, interlinking, or proximity. The provided list of approaches to node extraction from texts in this paragraph is not exhaustive.

The next step in network construction is relating the extracted nodes to each other. Various approaches have been studied to conceptualize relatedness between words, concepts, and entities in textual data. One of the most basic and widely studied models for relationships between words is co-occurrence in a specific context, i.e., a bag-of-words or n-gram-based representations [PNPV23]. A connection is formed between a word and all words before and after the word until a specified distance threshold has been reached. This can be either a fixed window size [CCV+15] or a specific semantic context such as an occurrence in the same sentence [CLA+06, SPW15] or document [YSZ+18, PNPV23]. These notions can also be combined as highlighted in [DCT12], where a window size of seven was chosen, which is restricted by a rule allowing the window not to exceed adjacent sentences. In this work, we refer to this category of text networks as co-occurrence networks. The extracted edges between nodes in these networks are often undirected. This method is unsupervised and based on the two assumptions. First, words and concepts more closely positioned in a text are also more related. The second assumption is that using many documents in analysis makes meaningful relationships more salient.

Other types of extracted relations include similarity or syntactic dependency between words [iSK04, OB20, PNPV23]. More recently, word embeddings and deep learning were also used in relationship extraction [ZSLZ22]. Three broad types of semantic relationship extraction between entities are summarized in [EBSW08, Kon14]. In knowledge-based methods, the extraction is based on manually crafted rules identifying a particular set of relations. This approach is usually applied for domain-specific tasks, thus requiring expert knowledge in many cases. It may yield rules that struggle to grasp the messiness of natural languages, which may entail low recall and high precision. Supervised methods require a labeled training set to train a machine learning model, which is applied as a classifier to extract relations. A significant amount of correctly labeled data is required to ensure accuracy. Self-supervised methods involve bootstrapping a small set of labeled samples to train a classifier. It is then applied to label unstructured data, which are used to train another supervised learning algorithm for relation extraction. According to this classification, the co-occurrence relation extraction discussed in the previous paragraph could be considered a knowledge-based or unsupervised method.

To analyze text networks, algorithms are applied for visualization and providing insightful statistics. Visualizations enable scholars to interpret texts' underlying content and structure through a unique perspective [GP15]. Similar to word clouds, with well-chosen parameters, they can be read as summarizations of texts. For example, in prior work [SAG17], a framework to utilize small co-occurrence networks of entity mentions was developed to support exploratory navigation of text collections. The visualization and construction of text networks influence their narrative capabilities [VBJG16]. For example, prior work [ATMM17] suggests enhancement of readability for humanities scholars can be possible through displaying semantic categories of entities next to their name, i.e., the name "Elfriede Jelinek" would be shown with a corresponding semantic category like author. The readability of networks is also strongly determined by layout algorithms employed, as just randomly placing nodes in a two-dimensional space would be chaotic and confusing. The Force Atlas algorithm [BHJ09] is suggested as a well-rounded layout choice for text networks [Par11].

A great variety of statistical methods to study the properties of networks are well summarized in [WF94, New03]. Many of those are employed to derive meaning beyond the interpretation of visualized text networks. We found a substantial body of research on applying various centrality and ranking measures to derive keywords or important entities from co-occurrence networks, e.g. [MT04, KMK16, BBS18]. Several surveys on different graph-based keyword extraction methods give insight into various approaches [BMM15, Gar21]. Centrality measures in text graphs have also been used to extract sentences for text summarization [ER04], retrieve representative words for sentiment analysis [CCV+15], and rank documents for information retrieval [BL12]. Another widely applied method is community detection to retrieve topics in co-occurrence networks representing a collection of texts. In particular, the KeyGraph algorithm [OBY98] was shown to have high accuracy and good runtime performance [SR13, YSZ+18]. Furthermore, [SPW15] highlighted how policy frames can be extracted by retrieving topic clusters in specifically filtered co-occurrence networks to illustrate differences in political discourse on nuclear energy across six countries. Ultimately, there are a variety of ways text networks can be constructed, visualized, and analyzed, and they are still being researched while also new methods are developed. It depends on requirements and priorities which approaches are most suited for a specific study.

2.3 Our approach

After reviewing the literature and evaluating available tools and models, we focused on co-occurrence networks extracted from news articles through various NLP techniques. The nodes of these networks are lemmatized nouns, adjectives, and named entities, and they are related when they co-occur in the same sentences. This approach is inspired by prior work that uses network-based methods for analyzing texts and news [MT04, SR13, SAG17]. We discuss in section 3.2.7 various other discarded node and relation extraction approaches, which we also wanted to use but found to not perform well on our dataset after some initial testing. We chose the Force Atlas 2 layout algorithm for visualizing the networks [JVHB14], which implements various improvements compared to the predecessor based on user feedback. As our review indicates, it is commonly used to visualize text networks. We use language-agnostic, unsupervised methods to explore and analyze the co-occurrence networks. First, we use the Textrank algorithm [MT04] to rank entities according to relevance (see section 3.3.2). Our review indicates that this approach is established and tested, and much research builds upon it [Gar21, ZSLZ22]. Second, we use a community detection algorithm to identify themes and topics in the networks. Our approach is inspired by the KeyGraph algorithm [OBY98], which performs well according to our review. It is further discussed in section 3.3.1. Our methodological approach is described in more detail in chapter 3. More recently, approaches based on word embeddings have received significant attention, but we decided against adopting those because of concerns about the interpretability of the underlying models. We think that text networks provide many opportunities for interdisciplinary, qualitative interpretation as they are expansive representations of many texts, which can enable domain experts to conduct close readings and possibly add more contextualized knowledge to analysis. Furthermore, some of the subcorpora are comparatively small (e.g., Heute only has 4001 articles on domestic politics according to our classification), which can make a word embedding approach instable [AM18]. Graph-based representations are also used as inputs for deep learning techniques, as we have illustrated, further illustrating the need to improve such representations even when deep learning receives much attention at this moment. Thus, future work could also use our network construction scheme as input for further machine learning-based news analysis.

CHAPTER 3

Methodological approach

One primary outcome of this work was a proof-of-concept implementation for network-based news analysis. We thus built a multi-layered software pipeline (see fig. 3.1) that encompasses a NLP & semantic annotation, network extraction and storage, and a network and text analysis layer.

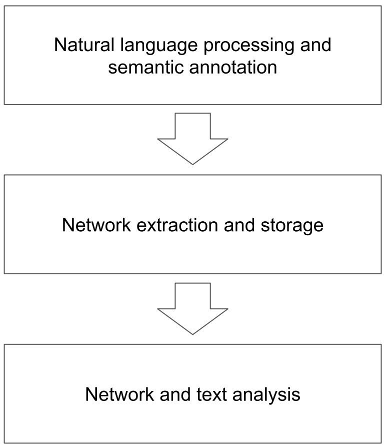


Figure 3.1: Layers of the news analysis tool.

A network-based approach to data exploration requires selecting and tweaking various parameters to gain insights from the data. We discuss the benefits and limitations of this flexibility further in the conclusion. We implemented a web application on top of this stack for quick querying of the data and visualization of results, which was presented at a public science event in Vienna (at the “Lange Nacht der Forschung 2018”). This tool was envisioned as a proof-of-concept implementation meant to aid digital humanities scholars in quickly conducting an exploratory analysis of text corpora by tweaking parameters and constructing expressive network visualizations. In the remainder of this chapter, we discuss programming languages, frameworks, and other tools used to create this thesis. We describe the inner workings of our news analysis software stack, briefly explain natural language processing and network analysis methods applied in this work, and unpack design decisions.

3.1 Implementation

Our primary concern for choosing NLP models and tools was their performance and quality. However, we also considered execution speed in our decision-making due to the great amount of textual data. The entire news corpus encompasses about 6 Giga Byte (GB) of texts without annotations and 23 GB with annotations after processing. We extracted about 16.775.932 nodes and 192.527.297 relations from the articles made available to us in the XML-based FoLiA format [vGR13]. These were stored in a graph database to enable efficient querying and filtering. The NLP processing, network construction, and querying, and most of the later analysis were executed solely on a dedicated virtual machine at the ACDH with an Intel(R) Xeon(R) CPU E5-2630 using eight cores with 2.30GHz per core and about 43 GB of Random-access memory (RAM). To utilize the available cores, we prioritized tools and frameworks that support parallelization. We found that the available open-source NLP models and tools at the time we devised the pipeline were not able to deal with the names of some prominent Austrian politicians, organizations, and locations. We expect that most systems were likely trained on texts or use dictionaries focusing on Germany, which leads to the biased classification. Thus, the tokenization and named entity recognition schemes were adapted using a rule-based approach to remedy these deficiencies in an Austrian context. We also wanted that the resulting networks should be expressive and visually appealing. Thus, we chose tooling that also supports the tweaking of visualizations and was designed with network exploration in mind. In the remainder of this section, we explain the concrete tooling choices.

The primary programming language used for this project is Python, as it offers a variety of frameworks and libraries for data science, network analysis, and NLP. It is considered [Oli07, MA11] to be state-of-the-art in these areas, and according to a recent survey [Moo22], it is the most used programming language by data scientists. We used mainly Python to devise our NLP and network analysis pipeline. The initial step of the NLP pipeline involved tokenization and was conducted through a tool developed by the ACDH based on unitok [MPS14]. The spaCy [1] NLP library [HM17] was used for sentence segmentation, part of speech tagging, and as a basis for Named Entity Recognition (NER). It was chosen due to its good performance on benchmarks, emphasis on speed in processing, and support for parallel computation. We supplemented the spaCy NER scheme with a rule-based approach to better fit the Austrian context. We crafted a dictionary of Austrian-named entities, encompassing politicians, organizations, and locations, by retrieving semantic data from Wikidata [VK14] via SPARQL-queries. We used this dictionary to create rules to match prominent named entities. This approach also

[1] SpaCy 1.10.1 - `http://spacy.io/`

made it possible to use synonyms for a simple co-reference resolution scheme and to acquire additional metadata on these entities, such as gender, party affiliation, or occupation, for further analysis. Finally, lemmatization was applied using Inverse Wiktionary for Natural Language Processing (IWNLP) [LC15] and spaCy. In section 3.2.7, we discuss several failed attempts at extracting additional semantic information from the articles to construct more expressive networks.

The information extracted through the application of NLP tools was used to annotate articles from which different co-occurrence networks were constructed and stored in a graph database. This process and the network analysis are explained in more detail in section 3.3. We used PySpark [2] for efficient network extraction from the texts. PySpark is a wrapper library making the functionality of the large-scale data processing framework Spark [ZCF+10] accessible in Python. Neo4j [Web12] was chosen as our graph database due to its intuitive SQL-inspired query language Cypher, support for parallel processing, and its high adoption in the open-source community. We used the interactive interpreter IPyhon [PG07] and the web editor and scripting environment Jupyter Notebook [KRP+16] for data analysis tasks. A proof-of-concept web application was written in JavaScript, HTML 5, and CSS 4 to offer an interface to perform analysis tasks. JQuery[3] and JQuery UI[4] were used to make this application interactive. The following Python libraries and frameworks were used for text and network analysis:

- SciPy [JOP01] was used as a basic scientific computing library.
- NumPy [Oli06] and Pandas [McK10] libraries were employed for statistical analysis.
- Matplotlib [Hun07] was used to create visualizations shown in the dissertation.
- Networkx [HSSC08] and igraph [CN06] were employed for a variety of network analysis tasks.

The visualization of networks and clustering was conducted using the gephi-toolkit [BHJ09], which is only available in Java. Therefore a wrapper class was written in Java to make it accessible in Python via Py4J[5].

3.2 Natural language processing pipeline

The basis for our analysis is a corpus of Austrian news articles. To construct expressive networks, we use NLP techniques to extract named entities, words with part of speech tags, and relations from them. This section provides an overview of NLP methods used to create this thesis. NLP is concerned with the computer-supported processing,

[2] https://spark.apache.org/
[3] https://jquery.com/
[4] https://jqueryui.com/
[5] https://www.py4j.org/

transformation, and analysis of texts written in natural languages. The texts in the AMC corpus are mostly unstructured except for paragraphs, which are delimited via XML tags. The articles also have metadata (i.e., article properties) provided by the news outlets, such as the genre of an article. However, we found them to be unreliable and inconsistent across different outlets (see section 5.2). We provide a schematic overview of our adopted text processing pipeline in fig. 3.2. Every step mentioned in the figure is explained in more detail later in this section.

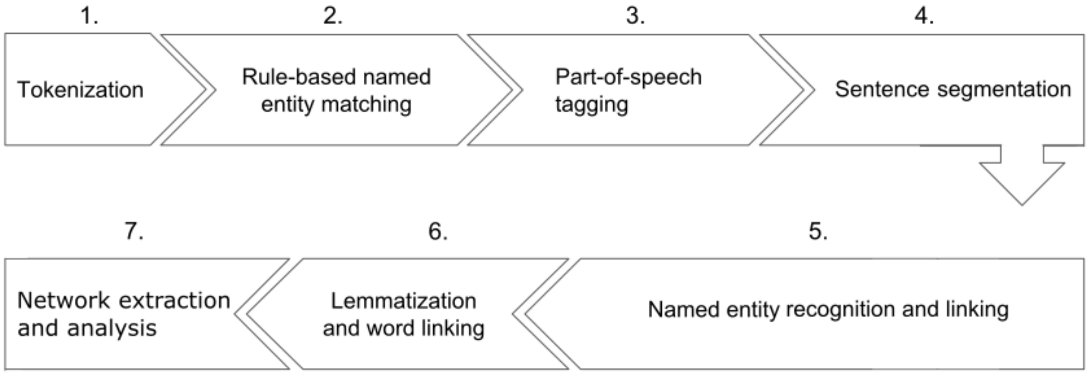


Figure 3.2: The NLP pipeline.

The tokenization (see section 3.2.1) is done via an adapted version of unitok [MPS14]. This tool is fast, so it was unnecessary to implement a scheme for parallelization. The result of this step are files containing tokenized articles. These are passed to a routine that searches for mentions of named entities previously retrieved from Wikidata (see section 3.2.2). Furthermore, full names of identified persons comprised of more than two tokens are transformed into a form to only consist of two tokens. We found that the named entity recognition model of spaCy 1 has issues with names of persons comprised of more than two tokens, like the Austrian president "Alexander Van der Bellen." Thus, this transformation into two tokens is applied before the texts are passed to spaCy (see section 3.2.2). We executed this process in batches to enable parallel processing to reduce the time required for the other NLP tasks, including part-of-speech tagging (see section 3.2.3), named entity recognition (see section 3.2.5) and sentence segmentation (see section 3.2.4). After this processing, the shortened names are reverted into their original form, and an entity linking scheme (see section 3.2.5) based on synonyms retrieved from Wikidata and stemming is applied. For all words not identified as named entities, a lemmatized (see section 3.2.6) form is looked up, and words with the same base form are linked together. These post-processing steps are not parallelized because the bottleneck in our pipeline was spaCy. The lemmatized nouns, adjectives, and detected named entities represent the nodes in the co-occurrence networks we construct for our analysis. The presented NLP pipeline can be considered an Information Extraction (IE) system. The annotated articles are stored in text files on disk after this process. The networks are then extracted from these files and stored in a graph database (see section 3.3).

3.2.1 Tokenization

Usually, the first step in NLP processing and our pipeline is tokenization [WK92]. Tokenization involves splitting unstructured text into lists of character chunks, so-called tokens. They function as the basic building blocks for processing in the later pipeline steps. The tokens are usually crafted through text splitting based on pattern-matching rules, trained machine-learning models, or a combination of both. The most basic form of rule-based tokenization is splitting based on a list of delimiters. Depending on the implementation, punctuations, spaces, or other special characters may be omitted entirely in the resulting token list or considered separate tokens.

Example:

Input:	Dr.	Romano	liest	ein	Buch.	
Output:	<u>Dr.</u>	<u>Romano</u>	<u>liest</u>	<u>ein</u>	<u>Buch</u>	<u>.</u>

Table 3.1: The text is split into six tokens because the period after "Dr" is identified as part of an acronym.

At first sight, simple delimiter-based rules are sufficient, but language ambiguities pose problems. The example 3.1 illustrates how simply using spaces and punctuation as delimiters would result in "Dr." not being recognized as a cohesive token. A good tokenization scheme should be useful from a methodological perspective and adhere to linguistic rules. [WK92, Tri13] Accuracy is important because errors propagate to upper levels of the NLP pipeline and cause higher-order problems. State-of-the-art systems are already very sophisticated, like the SoMaJo [PU16] tokenizer for web and social media texts in German. It was able to achieve an F1-score of 99.57 % for the EmpiriST 2015 shared task. The approach was rule-based and may not generalize well to other domains. The authors also noted that a simple split on white space and common punctuation marks yields an F1-score of 96.73 % on the same test data. This result highlights the immense difficulty of building a system with an accuracy close to or equal to 100 %.

Due to ambiguities, several possibilities for correct tokenization are available in some cases, and depending on the objectives, appropriate behavior has to be chosen. For instance, the name of the Austrian president, "Alexander Van der Bellen", may be either split into four tokens: "Alexander", "Van", "der", "Bellen" or two, the first name and the last name: "Alexander", "Van der Bellen". The tokenization for both cases is valid, but both have different meanings. Therefore, the semantics of this level of the NLP pipeline must be considered when implementing higher-up levels.

Similar to the previous example, the hyphenated word "Erste-Hilfe-Kurs", meaning first aid course, can be tokenized differently. It may be identified as a single term, "Erste-Hilfe-Kurs", or as a compound, like "Erste-Hilfe" (first aid), "Kurs" (course), or "Erste" (first), "Hilfe" (aid), "Kurs" (course). In German, such compound words are common and, in many cases, not even separated by dashes, like in the case of "Autofahrerin", which translates to a woman car driver. Nouns can also be gendered. The word car driver may be translated to a male “Autofahrer”, or a female “Autofahrerin”, version. Gender-neutral forms are not available in this case, making it difficult to infer if the gender-neutral concept, the female or male version, was referenced. Due to this ambiguity, a variety of gender-neutral forms have emerged, like “AutofahrerIn”, “AutofahrerIN”, “Autofahrer-in”, “Autofahrer*in”, “Autofahrer und -in”, “Autofahrerin und Autfahrer”. Since language is not static but ever-changing, different forms may emerge. In turn, tokenization systems must periodically incorporate new language use knowledge to enable high accuracy.

The tokenization scheme applied in this work is based on unitok [MPS14] and was amended by an expert of the ACDH to fit the Austrian context better. We initially wanted to use spaCy’s built-in tokenization but found it to have issues we some widely used Austrian acronyms, such as “Mag.”. Therefore, a dictionary fitting the Austrian context was crafted by the ACDH and integrated into unitok. Still, we found issues while adopting this scheme in our analysis. For instance, “3D” is split into the tokens “3” and “D”. It can be desirable to separate numerical tokens, but for our aggregation-based analysis, it is not. However, we consider the counterexamples we encountered during testing only to have minimal impact on our main analysis of election coverage. Therefore, some of these problematic cases were not addressed during the tokenization and should be considered in future work but were out of scope for this thesis. Also, conducting an in-depth evaluation of this tokenization scheme with a big test set would be desirable to enable future work on the corpus. The output after this transformation keeps all the tags of the original raw XML article files. Every line in the resulting file corresponds to either a tag or token. Every token is assumed to be followed by a trailing space if no “</g>” is found in the next line. This representation is called verticals format [6].

3.2.2 Rule-based named entity matching

Named entity recognition is concerned with identifying names in a text and assigning them a corresponding type [NS07, Eis19]. It is assumed the names refer to an instance of the type such as concrete persons or objects in the world. The variety of available types is a design decision, and therefore, implementation depended. We are interested in entities of the types: person, organization, and location. We illustrate the application of NER to a sentence in table 3.2, where three entities were identified.

Example:

Input:	Christian	Kern	fliegt	am	13.	Februar	nach	Brüssel.
Output:	Christian	Kern	fliegt	am	13.	Februar	nach	Brüssel.
	PERSON				DATE			LOCATION

Table 3.2: In the text the named entities “Christian Kern”, “13. Februar” and “Brüssel” are identified.

[6] https://www.sketchengine.eu/my_keywords/vertical/

The recognition of a named entity is often not unambiguous, as, for example, the German noun “Grill” may refer to the last name of the author of this work or a grill used for cooking. Another example more relevant to our case study is the noun “Kern”, which may refer to Austria’s former chancellor or the word “core”. Consequently, NER is context- and domain-dependent, as names can be specific to certain domains, locations, or cultures. A solely rule-based approach may require a great number of rules or dictionary entries used for matching to achieve satisfactory recall and may not be trivially transferable to other contexts. However, crafting many rules can also create matching conflicts instead of solving the issue of ambiguity, e.g., rules matching “Austria” may identify the country or the famous football club. Context information is crucial for accurate recognition but is hard to grasp in rule-based approaches. In contrast, a supervised machine learning approach may require a lot of training data, but the resulting model may have learned context information from the training examples. Also, no lookups in a big database or dictionary are necessary. According to [JBL16], most implementations either employ a rule- or a machine learning-based approach or a combination of both (hybrid).

The quality of detecting named entities, especially politicians and political parties, is central to our analysis of the 2016 Austrian presidential elections. Therefore we devised a set of sentences to test various state-of-the-art frameworks at the time in their ability to detect relevant actors such as “Norbert Hofer” and “Alexander Van der Bellen”. In this test, we found that the tools were not able to detect the winning candidate “Alexander Van der Bellen” correctly in several of test sentences (see section 4). Thus, we decided to employ a hybrid approach to increase accuracy. We augment the machine-learning-based framework spaCy through a rule-based approach that utilizes a dictionary to match the names of prominent Austrian persons, locations, and organizations. We first apply the rule-based scheme (i.e., the part in the NLP pipeline described in this section) to the corpus to fix some issues with tokenization, and afterward, spaCy is applied (see section 3.2.5).

We decided to use Wikidata [VK14] as a knowledge base for the dictionary because we found all candidates of the 2016 Austrian presidential election and a set of randomly chosen Austrian politicians in its repository. We found that most Wikidata entities have useful properties for our analysis. We used synonyms, which were available as properties, to implement a simple entity linking scheme (see section 3.2.5). Other available properties, such as gender, party affiliation, and location, were used in the statistical analysis (see section 5). In [GSG17], Wikidata is described as being especially well suited for the news analysis context due to its “up-to-date repository of persons and organizations”, compared to other knowledge bases like, e.g., DBpedia or Yago. They state that this benefit outweighs the risks involved in vandalism [HPSE15] because of the curation by users. Most entities were obtained from Wikidata through SPARQL queries specifically made to download all available Austrian entities. Furthermore, we used spaCy’s named entity recognition scheme to generate a list of the 500 most occurring named entities in our corpus and also retrieved them from Wikidata for the dictionary.

We use the retrieved names and synonyms from Wikidata to recognize named entities. Our scheme prioritizes matches based on full names, i.e., consisting of a first and last name. If no full name rule applies to a set of tokens and a full name was mentioned somewhere in the processed article, mentions of only last names are also matched as named entities. Matches of last names of Austrian politicians are also recognized as named entities when another Austrian politician or party is mentioned in the processed article. To further increase recall of Austrian politicians, we added a rule that matches names that are possessive nouns, such as "Hofers". Our scheme also considers inflicted forms in its named entity vocabulary.

In case of ambiguity due to overlapping or alike names, our matcher chooses the named entity with the longest name (greedy). If ambiguity still persists, first Austrian-named entities and then the ones with the most accumulated page views on Wikipedia in the month of publication of the corresponding article are chosen. This approach is inspired by [YTT15] where "temporal popularity knowledge," i.e., page views on Wikipedia in a corresponding time frame, was employed to improve entity recognition. To ensure reasonable accuracy of this rule-based approach, it is required for matched names in an article to have the first letter of first and last names in uppercase. In case names or synonyms we retrieve from Wikidata are in all caps, it is required for the matched tokens to also be in all caps.

We found that particular names of persons consisting of more than two tokens, like the winning candidate "Alexander Van der Bellen", are an issue to the spaCy named entity recognition scheme. To improve this, we merge all but the first token into one word, with the first letter in upper case and the rest in lower case. These transformed names improve recognition quality with spaCy in our example sentences. The performance of named entity recognition and linking is central to our case study because one of our main interests in this research on election reporting is how different news sources write about various candidates and parties. We provide an evaluation of our named entity scheme on a small corpus labeled by the ACDH in section 4.

After an article was processed, the token positions in the article and Wikidata-URLs of all found entities are stored in a list, which is later combined with the results of the named entity recognition of spaCy (see section 3.2.5). The tokenized texts of the articles are used to create spaCy "document" objects which are passed to a method for parallel NLP annotation. The title, content, and keywords are all encapsulated in separate spaCy "document" objects. All tags but paragraph tags are kept in a list to be restored after all NLP processing with spaCy was conducted.

3.2.3 Part-of-speech tagging

Part-of-speech tagging assigns syntactic markers, e.g., tags for nouns or adjectives, to words based on their grammatical use in a sentence [JM09]. This annotation is usually rule-based, machine-learning-based, or a combination of both. The probabilistic models work well in practice because they often consider context through modeling statistical relations and, therefore, can better cope with the variability of language. We provide an example in table 3.3 that illustrates how a sentence can be marked with syntactic tags.

Example:

Input:	Hannah	Arendt	spaziert	nach	England.
Output:	Hannah	Arendt	spaziert	nach	England.
	PROPN	PROPN	VERB	ADP	PROPN

Table 3.3: In the text the proper nouns "Hannah Arendt" and "England", the verb "spaziert" and the adposition "nach" are identified. The tags are based on the Google Universal POS tag set [PDM12].

This third step in our NLP pipeline is performed by the spaCy framework on a batch of tokenized texts in parallel. We chose the framework because it offers both speed and performance comparable to the state-of-the-art. [HJ15] It employs a probabilistic model[7], namely a greedy averaged perceptron, to predict Part-of-Speech (POS) which was trained on the TIGER corpus [AAB+03, BDE+04]. The corpus consists of annotated news articles from Germany, which means it likely performs also reasonably well on the German news articles of the AMC. SpaCy support both the Stuttgart-Tübinger [Sch95] and Google Universal [PDM12] POS tag sets. Since in our network analysis (see section 6), we are primarily interested in nouns and adjectives, we used the POS tags for filtering.

3.2.4 Sentence segmentation

This step in our NLP pipeline is concerned with detecting sentence boundaries. This task often depends on identifying punctuations indicating the end of a sentence or analyzing the syntax tree of a sentence. In section 3.2.7, we explain how syntax trees can be constructed through dependency parsing. Sentence segmentation also depends on tokenization (see section 3.2.1) as an approach based only on identifying punctuations will fail due to language ambiguities. For instance, the sentence: "Ich möchte Dr. Hauser besuchen.", would be wrongly interpreted as two sentences with an approach that simply looks for punctuation signs. A more sophisticated segmentation system would recognize the first punctuation as an indicator for an abbreviation. Decimal numbers are also sometimes written using a dot, such as "1.5", which poses a similar problem. There are a variety of other difficult expressions, such as dates (e.g., "1.1.2001"), URLs (e.g., "http://dbpedia.org"), email addresses (e.g., "gabriel.grill@tuwien.ac.at"), horizontal ellipsis ("..."). Consequently, improving sentence segmentation performance requires incorporating many special cases, which may also be domain-specific. We applied spaCy's sentence segmentation scheme in this work. It uses a syntax tree to determine sentence boundaries[8].

[7] https://explosion.ai/blog/german-model
[8] https://spacy.io/api/dependencyparser

3.2.5 Named entity recognition and linking

In this step of the NLP pipeline, spaCy is used for named entity recognition and Wikidata dictionary for entity linking afterward. We used spaCy version 1.10 with model de_core_new_md because it performed best in our evaluation (see section 4). The spaCy results are combined with the named entity list generated from our rule-based approach(see section 3.2.2). In case of conflicting classifications, we privilege the results from the rule-based approach. We mainly employ spaCy's model to detect named entities not so prominent on Wikipedia. In our evaluation (see section 4) of spaCy and the rule-based approach, we found that, as expected, the latter offers high precision while the machine-learning approach offerers much better recall. Still, we found the improvements through the hybrid approach relatively minor in the evaluation. We postulate that this result may be because our test corpus does not contain many articles featuring persons with names of more than three tokens and was not specifically crafted to cover election coverage. The more difficult names were responsible for most misclassifications in our other much smaller test set of sentences on Austrian politicians.

Adopting Wikidata in our rule-based named entity recognition enabled us to include a simple entity linking scheme into our NLP pipeline. The task of entity linking or co-reference resolution involves attributing all mentions of a named entity to the same entity. This problem is still actively researched. As expected, we only found a few frameworks for German texts, as the vast amount of NLP research is conducted on English texts. We tested the CorZU [Tug16] framework of the University of Zurich, the only German co-reference resolution framework we found during our initial review. It requires the Zurich dependency parser, which was very slow in our test and, therefore, infeasible for our pipeline. Instead, we tried to use the spaCy dependency parser, but we found the tag sets from spaCy, and the Zurich parser were too different for a conclusive translation. So CorZU was not usable for us. Recent efforts using deep neural networks led to significant performance improvements in German co-reference resolution [SHB21], but these results were published after we had already made all the tooling decisions for the NLP pipeline. Thus, instead, we implemented an entity linking scheme based on relating mentions of the name and synonyms in un- and inflicted forms of each named entity from Wikidata. The entities are identified by their ID on Wikidata. We also implemented a heuristic approach to linking mentions of the last name of a person to the corresponding named entity. In section 3.2.2, we describe the process of detecting different word forms of named entities in detail. The named entities detected by spaCy are also linked based on their lemmatized forms (see section 3.2.6), but since no semantic information about them but their entity type is available, more sophisticated linking based on, e.g., synonyms is not possible. We do not consider mentions of pronouns in the linking process.

3.2.6 Lemmatization

Lemmatization is the task of reducing inflectional word forms, like “ging” or “geht”, to their corresponding base form, in this example, “gehen”. [Eis19] Every word has an uninflected form called a lemma. Thereby lemmata establish a relation between inflected forms of a word to a shared base form. The context of a word is crucial to determine the correct lemmata of a word, as highlighted in table 3.4.

Example 1:

Input:	Elefanten	gingen	über	die	Straße.
Output:	Elefant	gehen	über	die	Straße.

Example 2:

Input:	Es	geht	mir	gut.
Output:	Es	gehen	mich	gut.

Table 3.4: The examples illustrates issues when a lemmatizer does not consider the context of a word. Example 1 correctly determines the base form of "geht" as "gehen", but in Example 2, "geht" actually is related to the lemmata "fühlen" (feeling).

We mapped nouns and adjectives extracted from news articles to nodes representing entities mentioned in the text. Such entities are best represented by their base form; therefore, lemmatization is critical to combat the sparsity of alike word forms in the texts and to yield simplified text networks. We apply a framework called IWNLP [LC15], which uses Wiktionary as a knowledge base. It achieves performance comparable to other state-of-the-art solutions. If words are unknown to IWNLP we employ spaCy's built-in German lemmatization dictionary or assume the word is already in its uninflected form.

3.2.7 Envisioned additional information extraction

We tried to incorporate several external semantic knowledge bases to extract more information from the unstructured news articles and thereby construct more expressive and concrete networks. We envisioned identifying subject-object-verb (SVO) triplets to extract more meaningful relations between entities. The goal was to classify the verbs according to valence to create directed networks highlighting how the relation between entities is portrayed in news. This approach of constructing such tonality networks was inspired by previous work on automated narrative network analysis [SDFFC15, SVC15]. However, state-of-the-art tools did not yield the desired accuracy, and results were too sparse. We only identified a few sentences that could be used to extract such relations between entities in the timeframe we investigated.

The extraction of SVO triplets can be done through dependency parsing. It aims to create a syntax tree out of a sentence with directed edges representing grammatical relations [Eis19]. These usually consist of a head and a dependent, e.g., for our case, we were interested in verbs are heads with a subject and object as dependents (child nodes). The choice of which types of grammatical relations are supported by the syntax tree depends on the standards adopted. We provide an example in table 3.5, which illustrates how a sentence is turned into a syntax tree. Only a few frameworks are available that support dependency parsing on German texts. The only parser we found that seemed reliable and still actively in development, besides the one of spaCy, was ParZu [SSVW09, SVS13]. We tried both with a selection of test sentences and found the results unsatisfactory after manually looking through the results.

Example:

Input:	Irmgard	Griss	fährt	nach	Zürich.
Output:	Irmgard	Griss	fährt	nach	Zürich.
		sb		mo	nk

Table 3.5: In the text, the verb "fährt" is the root node, and its dependent to the left is "Irmgard Griss" (subject) and to the right "nach" (modifier). The dependent of "nach""Zürich" (noun kernel element). The tags are based on the TIGER annotation scheme [AAB+03].

We tested several dictionary-based sentiment and emotion classification tools to determine their usefulness for extracting relations between entities. The sentiment classification tested was based on SentiWS v2 [RQH10] combined with the "German Political Sentiment Dictionary" [HJ17], which was specifically crafted to capture political sentiment in Austrian newspaper texts. After some initial testing, we identified too many sentences where we considered the classification incorrect. One big issue was that the dictionaries contain values that indicate the strength of either positive or negative sentiment, and depending on we set the thresholds, either sparsity became an issue or too many sentences were not correctly classified. For example, the sentence: Alexander Van der Bellen is a better candidate than Norbert Hofer, would be identified as a positive relation between Hofer and Van der Bellen, although the contents of the sentence have a completely different meaning. Finally, we tried out the German-Emotion-dictionary [KSR16], which promises to aid in recognizing seven basic emotions [Ekm99], namely anger, disgust, enjoyment, surprise, fear, sadness, and contempt. We found that the words in the dictionary were only sparsely present in relevant sentences, which made this approach unreliable. We still used sentiment and emotion classification to compare tonality across newspapers and time based on descriptive statistics of mentions (see section 5.3.3), but not for further network analysis.

3.3 Network extraction and analysis

The discussed NLP pipeline is crucial to generate expressive networks. The extracted named entities and adjectives represent nodes, and the identified sentences are used to construct co-occurrence relations between words. The figure 3.3 illustrates the resulting data model stored in the Neo4j graph database. The articles are represented by the Document class in the database, which has additional fields: publication date of the article (Date), the newspaper outlet it was featured in (Docsrc), the internal newspaper department responsible for it (Ressort), and the regions the article was available (Region). A Document instance can be related to multiple Sentence instances. The Sentence class also has a field that stores the full sentence in plain text (String). Sentence instances can be related to multiple Entity instances and vice versa. The Entity class represents lemmatized words and named entities. It has two fields: the original uninflected word or named entity (Label) and an identifier URL on Wikidata (Wikidata ID), which is only available for named entities. There are multiple types of relations between the Sentence and Entity classes. These relation types model the part-of-speech tag of a word or the named entity category identified through the NLP processing in a sentence. We used the Neo4j query language, Cypher, to retrieve co-occurrence networks from specific timeframes and with certain node types from this database. Our analysis focuses on co-occurrence networks that contain nouns, named entities, and adjectives as nodes and networks that only contain named entities representing persons and organizations (see chapter 6).

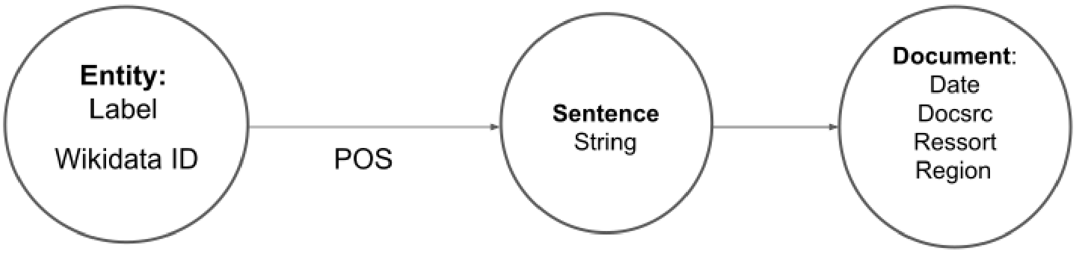


Figure 3.3: The text network model in the graph database Neo4j.

Since we wanted to make the networks expressive, data cleaning and preprocessing were necessary. For example, the named entity linking and lemmatization, part of the NLP pipeline, were meant to reduce the complexity of networks by aggregating more mentions meaningfully into fewer nodes. We also crafted stopword lists to exclude entities and words like weekdays, abbreviations, or newspaper-specific keywords. We also removed some politicians with last names corresponding to some very commonly used nouns to reduce the number of false positives. Words that contain less than two characters or non-alphanumeric characters except “-” and “.” are also excluded.

We also identified duplicate articles to exclude them during network extraction and avoid redundancy. An article is classified as a duplicate if it has an overlap of at least 95% with another article in the same issue. In total, 34.179 articles were excluded from the database through this process. Since our analysis is focused on the coverage of the 2016 Austrian presidential election, we also developed a simple article classification scheme to identify texts on domestic politics. We classified an article as such if it contained specific keywords like names of Austrian politicians or political organizations. This chapter has illustrated how we constructed expressive co-occurrence networks, addressing our first research question. The following sections explain the methods used to analyze the constructed networks and addresses research question two.

3.3.1 Community detection

This thesis is also concerned with identifying topics and themes in the corpus through a network-based approach. Thus, we employed the KeyGraph method [SR13], which involves using a community detection algorithm to identify topics. We use the Louvaine algorithm [BGLL08] for community detection. We present in section 6.1.2 an analysis of visualized communities detected in co-occurrence networks comprised of nouns and adjectives. The networks represent reporting during the election, and we discuss what the identified communities reveal about differences in reporting across outlets. We also present a modularity score for each network, which measures how well communities partition a network. Put differently, the greater the modularity, the more relations within communities and the fewer between them. The Louvaine algorithm is comprised of two phases. First, each individual node is classified as part of its own community. Then in the first phase, nodes are removed from communities and added to neighboring ones until modularity is locally maximized. In the second phase, the resulting communities are aggregated into nodes, and the first phase is applied again. This is repeated until it is no longer possible to optimize for modularity. The nodes are then disaggregated and the full communities are revealed.

3.3.2 Ranking named entities

This thesis is also concerned with identifying entities portrayed as relevant in election coverage. We present the results of this analysis and discuss the difficulty of modeling relevance in section 6.2. We employ the Textrank [MT04] algorithm to identify relevant entities in the co-occurrence networks. It involves calculating centrality measures based on the PageRank algorithm [PBMW99] for each node. The top-ranked nodes represent the keywords. We apply the algorithm to networks comprised of only named entities, either persons or organizations, to identify the most relevant entities. Usually, extracted keywords that frequently co-occur are combined into phrases. We omitted this step in our implementation because we apply Textrank to networks comprised of named entities that are already aggregated representations of mentions due to entity linking (see section 3.2.5). The importance of a node in PageRank [PBMW99] is determined by how many other important nodes are more closely connected to it. Initially, all nodes are assigned a PageRank value. Then a rank is calculated for each node repeatably until convergence based on a threshold is reached. A threshold can be, for instance, based on the number of iterations of the algorithm computed or the sum of all changes in the iteration.

CHAPTER 4

Evaluation of named entity recognition

The performance of named entity detection and linking is central to our case study because it is required to extract nodes and build networks. We conducted preliminary evaluations of several frameworks and discarded those lacking speed and performance on a small test set. Below we provide examples of misclassifications of the Austrian presidential candidate in 2016 "Alexander Van der Bellen" by state-of-the-art frameworks we tried out.

In table 4.1 we illustrate by example errors in NER employing the Stanfords CoreNLP framework [MSB^{+}14]. It could only correctly tag "Alexander Van" as parts of a named entity of type person, but it missed the last part of the name: "der Bellen".

Example:

Input:	Ich	mag	Präsident	Alexander	Van	der	Bellen.
Output:	Ich	mag	Präsident	<u>Alexander</u>	<u>Van</u>	der	Bellen.
				PERSON			

Table 4.1: The text was annotated using CoreNLP 3.9.1.

The DBpedia Spotlight framework [DJHM13] as shown in table 4.2 is not able to detect "Alexander Van der Bellen" but the presidential candidate "Norbert Hofer" is identified correctly.

Example:

Input:	Ich	mag	Alexander	Van	der	Bellen	und	Norbert	Hofer.
Output:	Ich	mag	Alexander	Van	der	Bellen	und	Norbert	Hofer.
								PERSON	

Table 4.2: The text was annotated using DBpedia Spotlight 1.0.0.

Lastly we tried the spaCy[1] framework, which also failed to detect "Alexander Van der Bellen" correctly as shown in table 4.3.

Example:

Input:	Alexander	Van	der	Bellen	ist	ein	Präsident.
Output:	Alexander	Van	der	Bellen	ist	ein	Präsident.
	Person						

Table 4.3: The text was annotated using spaCy v1 - de_core_news_md-1.0.0.

We suspect the inadequate quality in detecting "Alexander Van der Bellen" may be because the train corpora employed in the mentioned frameworks are based on German newspaper text and not Austrian. Consequently, based on these findings, we decided to enhance the NER performance in the Austrian politics domain of the framework spaCy through rule-based matching based on a knowledge base of well-known named entities. We choose Wikidata as a knowledge base on the assumption that important actors have a Wikipedia page. In order to strengthen our confidence in this assumption, we crafted a set of Austrian politicians, which we tried to find in different knowledge bases. Wikidata included them all with additional meta information, such as synonyms. The hybrid approach enabled the detection and linking of mentions of synonyms and stemmed forms, and last names in many cases. In table 4.4, we illustrate how spaCy is not able to detect "Van der Bellen" as the last name of a person. We found last names and abbreviations to be very commonly used to refer to well-known politicians, which are required for our analysis.

Example:

Input:	Van	der	Bellen	ist	ein	Präsident.
Output:	Van	der	Bellen	ist	ein	Präsident.

Table 4.4: The text was annotated using Spacy v1 - de_core_news_md-1.0.0.

While commencing our work, spaCy v2 was released, which employs deep convolutional neural networks for its NER. We tried it out, and to our delight, it was able to detect

[1] https://spacy.io/

"Alexander Van der Bellen" in several example sentences as shown through example in table 4.5.

Example:

Input:	Ich	mag	Präsident	Alexander	Van	der	Bellen.
Output:	Ich	mag	Präsident	Alexander	Van	der	Bellen.
					PERSON		

Table 4.5: The text was annotated using Spacy v2 - de_core_news_sm-2.0.0.

Still, we found example sentences incorrectly tagged, as illustrated in table 4.6. The misclassifications of spaCy v2 seemed to us very unintuitive. Classifying a whole sentence as "MISC", a category for named entities that are probably not organizations, persons, or locations, seems like a grave error.

Example:

Input:	Ich	mag	Bundespräsident	Alexander	Van	der	Bellen.
Output:	Ich	mag	Bundespräsident	Alexander	Van	der	Bellen.
				MISC			

Table 4.6: The text was annotated using Spacy v2 - de_core_news_sm-2.0.0.

In order to evaluate our scheme on actual domain-specific data and compare the performance of spaCy v1 and v2, our project partner, the ACDH, kindly employed students to tag named entities in articles randomly chosen from the AMC. The evaluation set was made of 1.156 articles from 36 outlets. In table 4.7, we illustrate the composition in more detail.

Kleine Zeitung	Krone	Standard	Kurier	NVB	Salzburger N.
115	124	24	58	23	30
OÖ Nachrichten	OTS	Presse	NVT	APA	Tiroler Tagesz.
57	32	22	17	95	24
Wiener Zeitung	BVZ	Falter	Format	KW	TTKOMP
21	39	4	5	1	3
SBGW	TV Media	MWVOLL	Heute	News	BAUERNZT
23	8	21	13	5	4
OBERRUND	Woman	ATVVOLL	E-Media	Trend	Medianet
7	2	1	1	1	2
Vorarlberger N.	NÖN	SPORTZTG	Augustin	HOR	Österreich
32	208	1	1	1	45

Table 4.7: Newspaper sources and amount of articles used in the evaluation set.

The result of our evaluation is shown in table 4.8. We calculate the Precision 4.1, Recall 4.2 and F_1-score 4.3 based on the number of tokens (in-)correctly assigned to the name entity categories person, location, and organization.

Configuration	Named Entity	Precision	Recall	F1-score
Austrian linking	Person	**0.84**	0.24	0.38
	Organization	0.45	0.09	0.15
	Location	0.41	0.18	0.25
Complete linking	Person	0.77	0.35	0.48
	Organization	0.42	0.22	0.29
	Location	0.27	0.31	0.29
Shortened linking & privilege Austrians	Person	0.81	0.33	0.47
	Organization	0.48	0.19	0.28
	Location	0.36	0.28	0.32
Spacy v1	Person	0.74	**0.89**	**0.81**
	Organization	**0.50**	0.37	**0.42**
	Location	**0.63**	0.54	**0.58**
Spacy v1 Austrian linking	Person	0.73	**0.89**	0.80
	Organization	0.46	0.38	**0.42**
	Location	0.55	0.56	0.56
Spacy v1 Complete linking	Person	0.75	0.87	0.80
	Organization	0.35	0.41	0.38
	Location	0.38	**0.59**	0.46
Spacy v1 Shortened linking & privilege Austrians	Person	0.73	**0.89**	0.80
	Organization	0.45	0.39	**0.42**
	Location	0.47	0.58	0.52
Spacy v2	Person	0.54	0.83	0.66
	Organization	0.38	0.46	0.42
	Location	0.39	0.54	0.45
Spacy v2 Austrian linking	Person	0.55	0.86	0.67
	Organization	0.36	0.46	0.40
	Location	0.37	0.57	0.44
Spacy v2 Complete linking	Person	0.55	0.86	0.67
	Organization	0.34	**0.47**	0.39
	Location	0.28	0.58	0.38
Spacy v2 Shortened linking & privilege Austrians	Person	0.55	0.87	0.67
	Organization	0.35	0.46	0.39
	Location	0.33	0.58	0.42

Table 4.8: Performance of different named entity recognition schemes.

$$precision = \frac{true \quad positives}{false \quad positives} \tag{4.1}$$

$$recall = \frac{true \quad positives}{false \quad negaties} \tag{4.2}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{4.3}$$

We evaluated only running spaCy v1 and v2 on the test corpus, where to our surprise, we found spaCy v1 to perform better. We presume this to be the case because, in v1, a much bigger corpus was used to train the NER model. We also tested three sets of entities retrieved from Wikidata integrated with our rule-based scheme in combination with the spaCy versions and standalone. The three entity sets contain Austrian persons, organizations, or locations retrieved through a specifically tailored SPARQL query. The "Complete linking" set was crafted by looking up all named entities detected through spaCy v1 in the corpus on Wikidata. The third set, "Shortend linking & privileging Austrians", is made of only the top 500 most occurring persons, organizations, and locations detected by spaCy and an additional step removing names and synonyms that are the same as names of Austrian entities. The set only contains entities where the full name makes up at least 25% of occurrences in our time frame in the AMC corpus, with very few exceptions. Through this approach, we got rid of synonyms that, by coincidence, also strongly resemble other words in the corpus and thereby would significantly increase the false positive rate. For this set, we also manually evaluated the quality of entity linking by checking sentences featuring one of the 200 most occurring entities in the corpus.

Overall the spaCy v1 standalone performed best, but we believe this only to be the case because most articles were not on politics (as highlighted in section 5.2). Consequently, the articles did not include many complicated names we targeted with our hybrid approach. Furthermore, the only spaCy v1-based scheme does not allow any form of entity linking, and we do have meta-information on named entities for additional analysis, like based on gender or party association of politicians. Of the three different Wikidata-based entity sets, the "Shortend linking & privileging Austrians" set performed best standalone compared on the basis of F_1-scores. As expected, the rule-based approach offers high precision but low recall. Also expected was that the "Complete linking" set has the best recall as the set of entities is the biggest while sacrificing precision. Generally, the performance for detecting persons with about 80% is good and comparable to the state-of-the-art. In contrast, the detection performance for organizations and locations could be better. We presume this to be the case firstly because spaCy is trained on German newspaper texts, which may entail that persons are easier to detect because of similarity in naming. However, this does not hold for locations or organizations where names may be more specific to Austria. Secondly, persons' names are often only comprised of a first and last name, which may be a more easily recognizable pattern. Last, but not least, we suspect increased ambiguity in our domain. The corpus consists, for instance, of various articles on sports, e.g., “The match Wien vs. Linz will take place in may”, where it is not trivial to determine if the named entities refer to the sports teams or locations. Sports articles are often not comprised of well-formed sentences but game results, e.g., “* Wien - Linz (3:0) *”. Similar problems exist, for instance, for articles on TV and cinema programs or events.

Ultimately we decided to employ spaCy v1 with the “Shortend linking & privileging Austrians” set, used for rule-based matching, due to comparable performance to the best-performing configuration. Additionally, it can detect important politicians well and provides meta-information on entities through Wikidata.

CHAPTER 5

Austrian Media Corpus

The Austrian Media Corpus is maintained by the ACDH and was first released in 2013 [1]. Since then, it has been constantly updated to encompass articles from over 50 newspapers and magazines, released starting in 1990 [DMPR14] up to the end of 2017. When this study was conducted, it was made of 10,549,292,505 tokens, 38,268,374 unique words, 696,694,153 sentences, and 42,001,922 articles, according to ACDH corpus statistics. In this chapter, we first introduce and motivate the case study. It spans a defined time frame, the 2016 Austrian presidential election, with a tangible but still significant amount of articles and a varied set of newspapers. We then provide various descriptive statistics to give some insights into the temporality of reporting, the distribution of coverage on the Austrian political spectrum, gender biases, and the writing quality across the papers.

5.1 Case study: Austrian presidential election in 2016

Our case focuses on the Austrian presidential election in 2016, which is an interesting object of study for several reasons. First, it received international attention as it was perceived as an election that could indicate a trend toward the right political spectrum in Western democracies. Inside the EU, it was feared that Austria would follow the UK, which voted to leave the union shortly before and elect a candidate critical of the EU. Second, it was understood by spectators as an indicator of the impact on public opinion of the so-called "refugee crisis" of 2015 in Europe. It was presumed that it had stirred up racist and nationalist sentiments in Austrians, which would make far-right opinions on migration and asylum-seeking, at the time most prominent in the Austrian Freedom Party, much more appealing to many. Ultimately, these assessments turned out to be correct as the final run-off was between a candidate promising to welcome refugees and the other insisting on the importance of closed borders. The two candidates were

[1] https://id.acdh.oeaw.ac.at/amc

not from the historically major Austrian parties, usually winning the presidency. This electoral duel was viewed as representative of a trend for other upcoming elections in the EU. Some argued that it could even have a signaling effect and influence outcomes in neighboring countries. Inside Austria, it was considered an election to determine the country's future direction in terms of EU and migration politics. Third, the election is also particularly interesting because the run-offs were repeated once and then again postponed. Both of these events are also visible in our statistical analysis. Since this election was held over several months, it opens up possibilities to compare many different points in time to understand how such events unfold in news reporting.

The timeframe we consider in our analysis starts on the 1^{st} of September 2015 and extends to the 30^{th} of December 2016. The first round of voting with six candidates was conducted on the 24^{th} of April 2016. We included in our analysis articles published several months before the first election date to detect the start of increased reporting on the candidates. On the first ballot, the voter turnout was at around 68,5 %. The electoral candidates received the following proportions of votes:[2]

1. **Norbert Hofer**: 35,1 %
2. **Alexander Van der Bellen**: 21,3 %
3. Irmgard Griss: 18,9 %
4. Rudolf Hundstorfer: 11,3 %
5. Andreas Khol: 11,1 %
6. Richard Lugner: 2,3 %

The candidates Norbert Hofer and Alexander Van der Bellen received the most votes, so they ran against each other in the supposedly final round of voting. Both candidates were not associated with any of Austria's two big ruling parties at the time, which never occurred beforehand in Austrian history. The second ballot was conducted on the 22^{nd} of May with Alexander Van der Bellen acquiring a majority with 50,35 % [3]. It was soon afterward repealed and a repetition of the vote was set to take place on the 2^{nd} of October. However, this date was postponed due to issues with the envelope glue used for postal voting. The final and definitive date was the 4^{th} of December 2016. The turnout was at around 68,5 % and the election result was:[4]

1. **Alexander Van der Bellen**: 53,79 %
2. Norbert Hofer: 46,21 %

[2] http://wahl16.bmi.gv.at/1604-0.html

[3] https://www.bmi.gv.at/412/Bundespraesidentenwahlen/Bundespraesidentenwahl_2016/start.aspx#pk_06

[4] http://wahl16.bmi.gv.at/

The new president of Austria in the end, therefore, was Alexander Van der Bellen, the candidate who ran a pro-European and pro-refugee campaign. To capture also some of the post-election reporting, we decided for our subcorpus also to encompass the whole of December and not to stop a day after the election date abruptly. In our analysis, we consider six daily newspapers, namely: “Heute,” “Krone” (“Kronen Zeitung”), “Kurier,” “Österreich,” “(die) Presse” and “(der) Standard .”We decided on this amount to make the analysis more tangible. Furthermore, the choices are widely read and cater to different audiences. Therefore we believe it can be considered a representative sample of the Austrian print media landscape.

5.2 Description of the corpus

The corpus crafted for our case study encompasses about seven hundred thousand articles, 3.3 million unique lemmas and named entities, and 16 million sentences. 34.179 duplicate articles were not included in the subcorpus. The total disk storage size of the annotated corpus is about 14 GB, the CSV files used for a batch import into the graph database amount to 4 GB, and the graph database to 10 GB. These measures illustrate the volume and variety of the data, which had to be considered in selecting tools and methods.

The articles are available to us in well-formed XML and usually consist of a main content body, a title, and some article properties. Due to various shortcomings, we decided to leave most of these properties out of our main analysis after some examination. We found that some newspaper articles have a few select topic words as specified properties. However, because of their sparseness, we decided not to consider them in our analysis. We also excluded in the main analysis the property on the number of tokens in an article. Instead, we chose to quantify word occurrences identified through our NLP processing pipeline and observed that our word counts frequently deviated from ones of the given property. This result is not surprising as tokenization is not an unambiguous task, as explained in section 3.2.1. The property for regions of issue publication has the same value across all articles and is thereby not meaningful. This result was unexpected as, for instance, the newspapers “Krone” and “Kurier” feature regional content.

The 'media type' property takes on the value “print” in all articles and, because of that, was discarded by us as well. The property specifying authors of articles only has a set value in only 84.796 articles and often contains abbreviations, which are only understandable with the appropriate domain-specific information of news publishers. For some articles, full sentences describing people involved in an article are provided. Meaningful analysis would require a separate information extraction routine to retrieve the relevant author names from such sentences. Consequently, we decided to also omit this property in our analysis because of its sparse presence and the hard to parse content. The following properties are considered in our data analysis: article id, publishing date, name of the publisher, and internal department (or “ressort” in German). We consider the department property a proxy variable for the composition of broader topical areas in the newspapers. It takes on values such as “politics”, “economy” or “sports” as the departmental structure is often based on topical areas. We decided to refer to this property in the text as "topic category" because this phrase better captures the meaning. The remainder of this section comprises several descriptive graphs to give readers of this thesis a better feeling of the data in this complex and varied corpus. In section 5.3, several more in-depth statistical measures and graphs are presented to allow for a more detailed understanding of the data.

While studying the amounts of articles published in the different issues closely over time, we noticed that for the 26th Dec. 2015, 16th May. 2016 and 26th Dec. 2016, no articles were present in the corpus. These dates are also Austrian holidays, which we think explains their absence. In figure 5.1, we visualize how many words per week on average occur in an issue published in the analyzed timespan. We choose to present only word instead of article counts in figure 5.1 because the length of articles varies heavily and is, therefore, less expressive as a measure to describe how much content was produced within a timeframe. The frequency polygon shows an apparent decrease in published content during December 2015 and 2016. We believe this decline is due to the Christmas holidays being widely celebrated in Austria. There were also declines in July and August of 2016, which indicates less activity of certain newspapers in the summer. Only some variations are visible throughout the other months, but none stand out firmly to us.

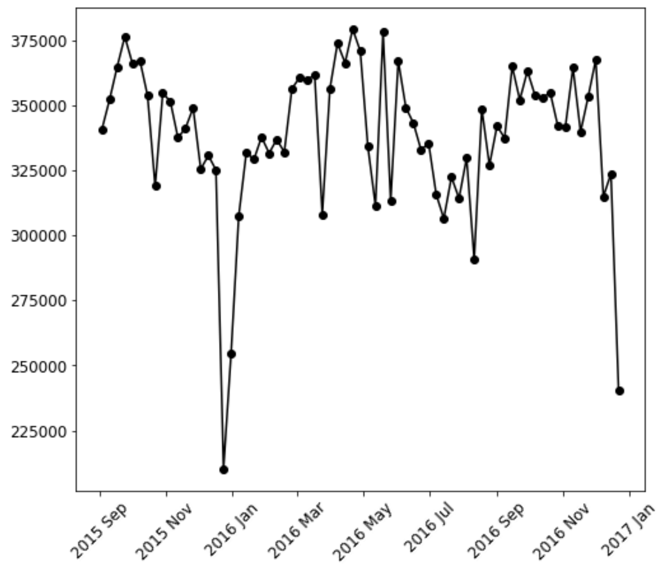


Figure 5.1: Words per issue on average per week of all newspapers.

In figure 5.2, we highlight how much content is published on weekdays on average. Interestingly we see an increase in content from Monday until Saturday and then again a drop on Sunday. The low amount of content on Sunday is due to several newspapers not publishing this weekday, which is illustrated in figure 5.2 in section 5.3.

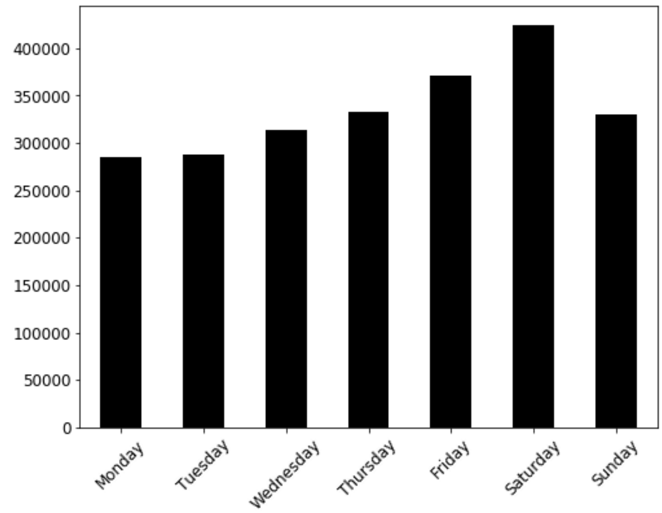
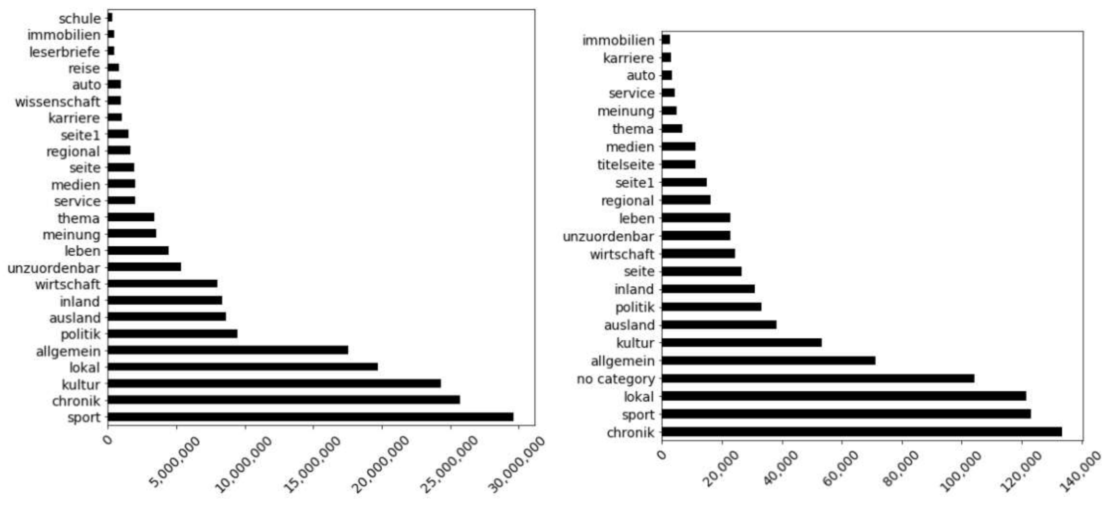


Figure 5.2: Average number of words per weekday.

The amount of words published per topic category is visualized in figure 5.3a to give some overview of how much certain topical areas are represented in the corpus. For example, this figure shows that the corpus contains more content on sports than politics. Still, the assignment of these categories is often ambiguous and usually only designated for internal use. Therefore, any interpretation needs to be done carefully and may require additional communication with the news outlets for clarification no how the categories were assigned. Only categories with at least 2.000 articles were included. Notably, many articles are associated with more than one category.



(a) Number of words per category.

(b) Number of articles per category.

Figure 5.3: Content published as part of a topic category.

The corpus contains 104.251 articles without a category and 2.974 solely classified as "uncategorized". In total, 34 categories are part of the corpus, and often multiple are assigned to one article. There are about 130 sets comprised of combinations of topic categories. In figure 5.3b, we visualize all categories associated with at least 2.000 articles. The categories "lokal" (local), "sport"(sports), and "chronik" (chronicle) have by far the most associated articles as they are featured heavily in "Krone" and "Kurier" (see also figure 5.7). Both papers publish the most content (see figure 5.5). There are relatively few articles on "kultur"(culture), but as highlighted through graph 5.3a, these contain much content compared to other topical areas.

The categories related to Austrian politics are "politik" (eng.: politics) and "inland" (eng.: domestic). They encompass 56.772 articles, whereas our classification, based on the detection of named entities representing Austrian politicians (see section 5.3.1), yields, in comparison, about 78.594 articles. We calculated the intersection of both article sets based on the different classification schemes and found an overlap of only 20.880 articles. The category set "inland politik" (eng.: domestic politics) is assigned to approximately 7.603 articles and shows the most similarities to our classification with an overlap of about 68%. Second are articles solely associated with the politics category with an overlap of 56% (5.100 articles). In contrast, articles associated with politics and other categories but not "inland" (eng.: domestic) overlap 34 % (5.569 articles). The ones tagged as solely domestic only have an overlap of about 15% (2.469 articles). These results made it clear to us that topic categories are not very useful for classifying articles for our election analysis.

The sheer amount and variety of content and properties highlighted in this section illustrates the complexity of conducting news analysis on the AMC. In figure 5.4, we illustrate the number of words in all newspapers associated with groups of topic categories we crafted manually with semantic similarity of topics in mind. This figure was created to illustrate how topic categories. The grouping with the corresponding consolidated categories is listed in section 8.2.

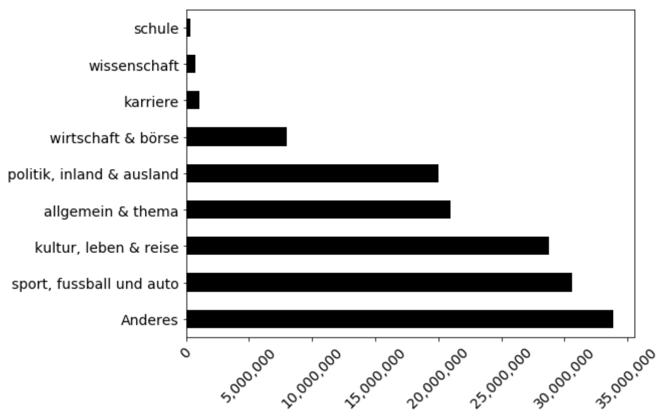


Figure 5.4: Number of words in articles according to topic categories.

5.3 Comparison of news outlets

In this section, we highlight differences based on counting mentions of named entities and other text properties in reporting between the news outlets we have chosen for our analysis. Our motivation is that making differences in reporting patterns visible matters can aid public discourse on media outlets. The set of newspapers we choose for our analysis are widely read in Austria and can be classified into featuring either predominantly sensational or high-quality news content. According to [GB17], "Standard" and "Presse" can be classified as high-quality papers while "Krone" and "Heute" as tabloids. The authors note that "Kurier" may be viewed by many as not being a tabloid newspaper, but they argue it still exhibits many properties of this category. Therefore, they classify it as a tabloid paper in their study. We use this categorization in our interpretation of results. The newspaper "Österreich" is assessed by us as a tabloid medium due to the types of articles it predominantly features and its distribution strategy as a freely available newspaper financed by ads. Our confidence in this categorization is further strengthened through the results of our statistical analysis of writing quality (see section 5.3.4).

In the first row of table 5.1, we list the amounts of articles published per news outlet in the corpus to highlight differences in quantity. "Krone" features by far the most articles, amounting to about 315 thousand. Ranked right after that are "Kurier" and "Österreich", which respectively published about 120.000 articles. Then lastly, "Standard", "Heute" and "Presse" with between 50 and 30 thousand articles.

	Heute	Krone	Kurier	Österreich	Presse	Standard
articles	48322.00	328630.00	127470.00	120974.00	62131.00	51576.00
mean	77.81	167.28	348.14	141.77	372.83	384.56
std	78.00	217.51	725.90	158.94	347.55	320.15
min	1.00	1.00	2.00	1.00	8.00	1.00
25%	37.00	54.00	81.00	70.00	94.00	165.00
50%	53.00	105.00	197.00	113.00	263.00	301.00
75%	103.00	216.00	375.00	181.00	594.00	537.00
max	6957.00	5529.00	19652.00	15739.00	4657.00	4364.00

Table 5.1: The first row lists the number of articles for each newspaper. The other rows contain statistical measures on the number of tokens in articles.

The ordering changes significantly when the number of words (tokens) instead of articles in news outlets is considered. Table 5.1 shows in row 5 the average number of words in the articles. "Heute" has the shortest articles with the least words. Also "Krone" and "Österreich" feature many articles with little content. Notably, according to our classification, all of these belong to the tabloid category. The exception is "Kurier", which publishes quite long articles on average, but as noted in [GB17], its classification as a tabloid may not be as evident and definitive as for the others. The median, listed in row five, positions “Kurier” between the newspapers that more clearly belong to the tabloid categories and those classified as quality news sources. “Standard” and “Presse” have the most prolonged articles on average, and both are classified as quality news sources. The high standard deviation across all measurements illustrates a high variability in length.

In figure 5.5, we illustrate the number of words (content) published weekly in each news outlet. “Krone” and “Kurier” feature the most content in our corpus, possibly due to both publishing multiple local issues. In contrast, “Heute” features particularly little content, which we expected may be due to its short length issues while featuring many commercial ads not included in our corpus. The amount of weekly content varies significantly across all newspapers, but the “Heute” remains relatively constant. Due to time constraints and the focus of this study on election coverage, we did not investigate the reasons for these variations in more depth.

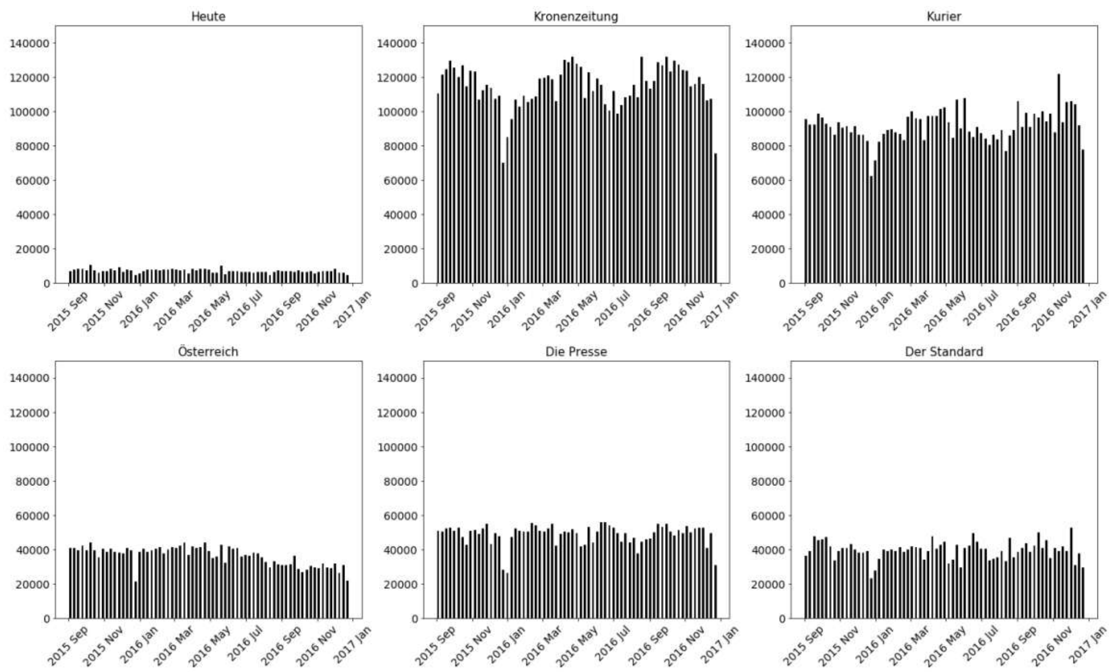


Figure 5.5: Words per issue on average per week in six newspapers.

Examining the average amount of content published per weekday in each newspaper (see figure 5.6), we observe that “Heute” only publishes issues from Monday to Friday and “Standard” until Saturday. All other papers are released throughout the whole week. Interestingly, only “Heute” seems to publish the same amount of content each weekday strictly. All others publish less during the beginning of the week and constantly more until the end of the week is reached. Mondays are an exception in “Krone”, “Österreich” and “Presse”, where more content is published compared to Tuesdays, Wednesdays, and Sundays in “Presse” and “Kurier”.

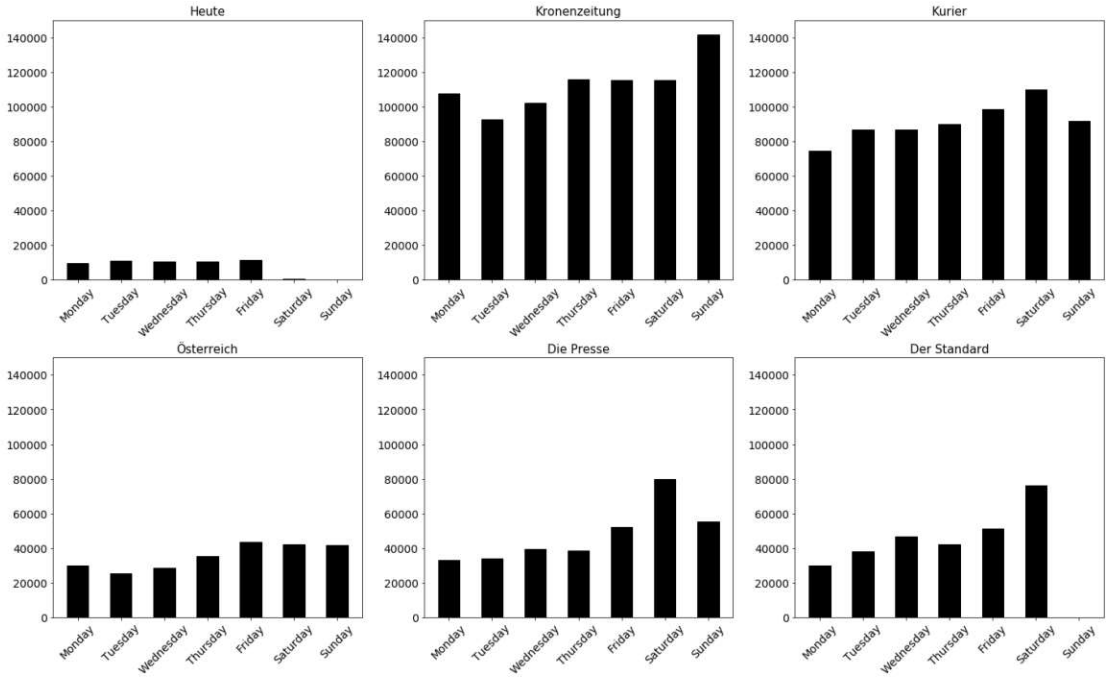


Figure 5.6: The average number of words per weekday in six newspapers.

The figure 5.7 shows the frequency of groupings of topic categories (see also section 5.2) across the newspapers in our set. We can observe that reporting on economics (ger.: wirtschaft und börse) appears to be most prominent in the quality papers "Standard" and "Presse". In contrast, "sports and cars" (ger.: sport, fussball und auto) are more dominant in the tabloid papers. Culture (ger.: kultur, leben & reise) is prominently represented in "Kurier" and quite strong in "Standard", "Presse". All of which are high-quality papers. "Standard" is the only paper with many articles explicitly labeled as reports on science (ger.: wissenschaft).

The general and other categories ("Anderes" and "allgemein & thema") comprise a significant amount of content in all papers. We believe that the topics of these articles may also be associated with more meaningful topic categories such as sports or science. Noticeably, "Krone" has very few articles (about 13.000) marked with a topic category on politics comparatively to the other papers. Through our political article classification scheme, we could identify about 24.000 articles on Austrian politics in the "Krone". We think this further shows that an analysis based on topic categories may not yield completely reliable insights, but it could still be useful as a heuristic. Consequently, we are convinced that our classification approach is necessary for more meaningful analysis. In section 5.3.1, we describe the content we identified through our approach based on detecting named entities of Austrian politicians.

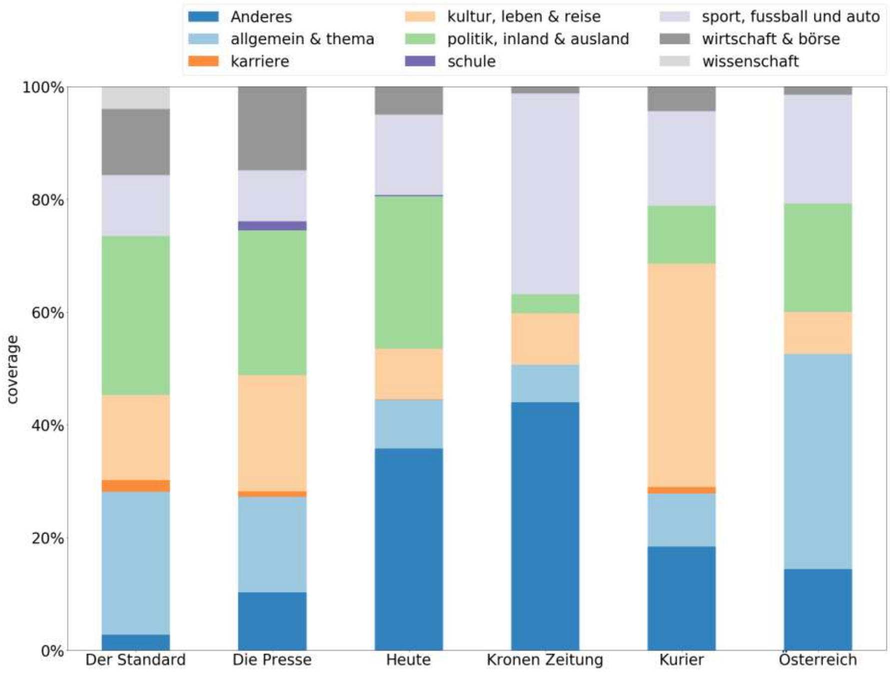


Figure 5.7: Amount of words associated with a topic category.

5.3.1 Coverage of Austrian politics

As we have elaborated in section 3.2.2, our analysis focuses on articles related to domestic politics identified through our detection scheme, that is, articles mentioning an Austrian politician or party. This filtering procedure is central to our project as the corpus comprises mostly articles not concerned with elections or domestic politics. Many of those articles are also not comprised solely of well-formed German sentences, like, for instance, the cinema program or game result announcements of sports matches. These would require special processing routines in our NLP pipeline since our machine learning tools were trained on corpora consisting of labeled, well-formed German sentences. Our analysis is partly based on frequencies of mentions of named entities or other phrases. Hence, problems with a subset of articles not related to politics arise due to the expected high frequency of some named entities that are, for our purposes, irrelevant to the election. These would need to be made visible as noise in our statistics and require separate, specifically crafted routines. Thus, we created through our detection scheme a sub-corpus, totaling 77.861 articles related to domestic politics, to avoid these issues.

In table 5.2, we illustrate how many articles and words the sub-corpus comprises compared to the total number per newspaper. The first row presents the number of articles related to domestic politics per newspaper. The second lists the ratio of these amounts compared to the total number of articles in a newspaper. The third row illustrates this proportion again but relates the number of words instead of articles. The "Krone" and "Heute" publish the least content on domestic politics compared to the other papers. According to our analysis, only about 11 % of their content is related to domestic politics. In contrast, the two high-quality newspapers “Standard” and “Presse” feature comparatively a lot of content related to domestic politics, whereas the former shows a slightly stronger focus. The “Österreich” surprisingly appears to publish more on domestic politics than the other two major tabloid papers. The “Kurier,” as a hybrid between high-quality and tabloid paper, reports more on politics than “Krone” and “Heute” but less than “Österreich”. The rows after line three in table 5.2 list statistical measures based on word counts in the articles. These highlight an apparent clustering between high-quality and tabloid papers. The former publishes, on average, much longer articles than the latter. All the elaborated statistics have highlighted differences between the newspapers, particularly the divide between high-quality and tabloid newspapers, which was also addressed in the previous section 5.3, becomes further visible.

	Heute	Krone	Kurier	Österreich	Presse	Standard
articles	4001.00	24594.00	14824.00	15830.00	9210.00	9402.00
ratio article	8.28%	7.48%	11.63%	13.09%	14.82%	18.23%
ratio words	10.88%	10.95%	12.73%	16.88%	18.93%	20.43%
mean words	102.21	244.72	381.09	182.91	476.20	430.88
std	97.82	326.98	324.15	196.11	339.73	305.02
min	12.00	11.00	11.00	2.00	12.00	16.00
25%	43.00	86.00	191.00	102.00	202.00	216.00
50%	77.00	168.00	319.00	153.00	434.50	376.00
75%	132.00	278.00	474.00	222.00	675.00	572.00
max	2182.00	5529.00	3820.00	15739.00	3208.00	3449.00

Table 5.2: First row: Number of articles on domestic politics for each newspaper. Second row: Share of articles on domestic politics. Third row: Share of word on domestic politics. Other rows: Statistical measures on word counts in articles on domestic politics.

The figure 5.8 shows how often the six Austrian presidential candidates were mentioned in absolute numbers across the newspapers over time. Irmgard Griss declared her candidacy in December, but she received increased coverage already in October, as our graph highlights. Her official announcement was about a month earlier compared to most other candidates. We found increased reporting activity starts one to two months before the official announcements for most candidates. For instance, Alexander Van der Bellen and Rudolf Hundstorfer got more media attention already in November 2015. In contrast, the decisions on who will run for president were less seamless for the FPÖ and ÖVP. This may explain why reporting on their candidates only increased during the official announcements in January. In the case of the ÖVP, a variety of well-known politicians declined until Andreas Kohl was officially designated. For the FPÖ, Ursula Stenzel was considered the best candidate by party leader Strache. However, social media campaigns against her likely ultimately led to Norbert Hofer becoming the FPÖ candidate. Finally, Richard Lugner announced his run in February. The tabloid newspapers, especially the

"Österreich", wrote regularly and extensively about him before this announcement and after the first ballot, which he lost. This is not surprising as he has used his high-society status for decades to get attention through media stunts. The high-quality newspapers only reported on him from January, when speculations around his run arose, until April when he lost in the first ballot. In the tabloid papers, interested in buzz and gossip, Lugner received attention outside of the race.

From March until the first ballot in April, reporting on all candidates was frequent. Alexander Van der Bellen remained on top in the frequency ranking across all newspapers. In May, reporting dropped rapidly for all candidates that did not advance to the second ballot. Then over the following months, Alexander Van der Bellen and Norbert Hofer remained present across all media outlets. A small peak in reporting activity is visible in September because the repetition of the second ballot was initially planned for the 2nd of October. A significant increase in mentions can again be observed in November and December when the final voting occurred. The high frequency of reporting on Van der Bellen in December is due to his victory and consequent coverage. The figure 5.8 shows how quantifications of candidates' mentions can be used to trace the discussed important events during the Austrian presidential election. We, ultimately, find that reporting on candidates most often starts before the official announcements.

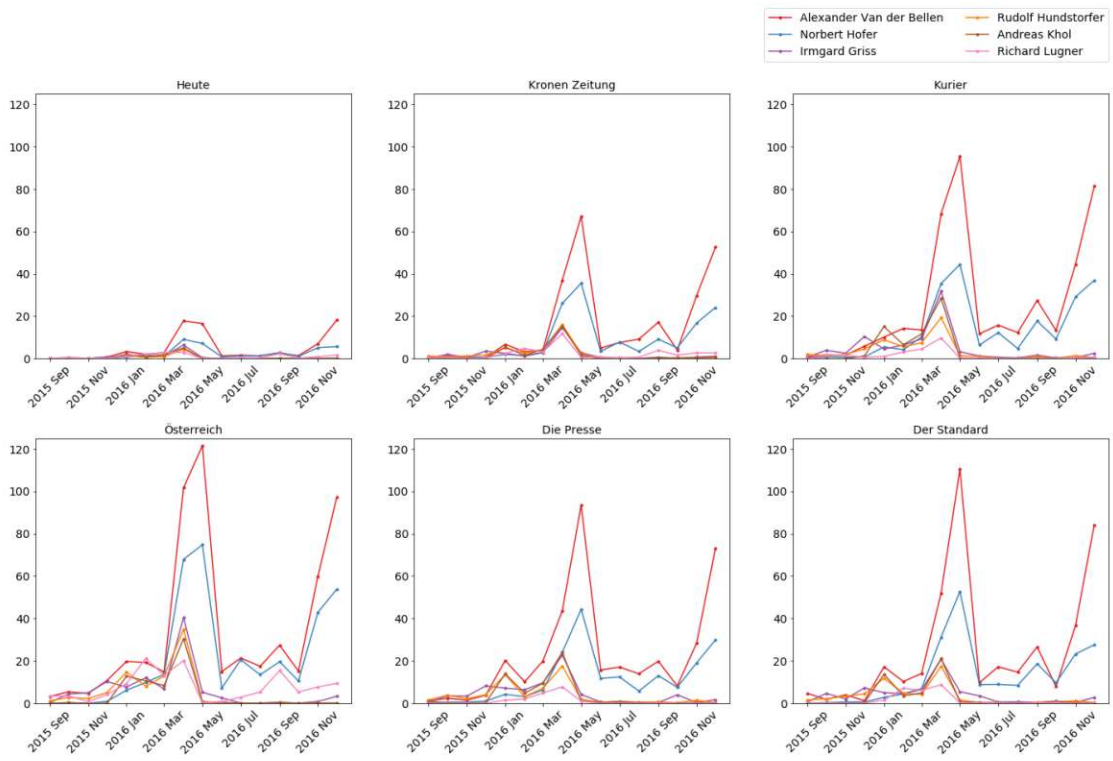


Figure 5.8: Mentions of presidential candidates per month in different newspapers.

In order to make the focus on different candidates by newspapers more visible, we decided to show in figure 5.9 how much each candidate is mentioned proportionally to the others. We only show mentions starting in January, as there was only little reporting on the candidates before that, as illustrated in figure 5.8. The graph again highlights how continuous reporting on Richard Lugner predicts whether a paper is part of the tabloid category. The most significant difference in reporting on the candidates can be observed in “Kurier”, which seems to feature significantly more articles on Alexander Van der Bellen throughout the elections. Andreas Kohl, the ÖVP candidate, received the most mentions during his run proportionally in “Presse” and “Kurier”, which are considered more conservative with some historical ties to the ÖVP. Irmgrad Griss was visible in all the high-quality publications and the “Kurier”. Alexander Van der Bellen received the most attention across all newspapers, while Norbert Hofer managed to be more present only a few times.

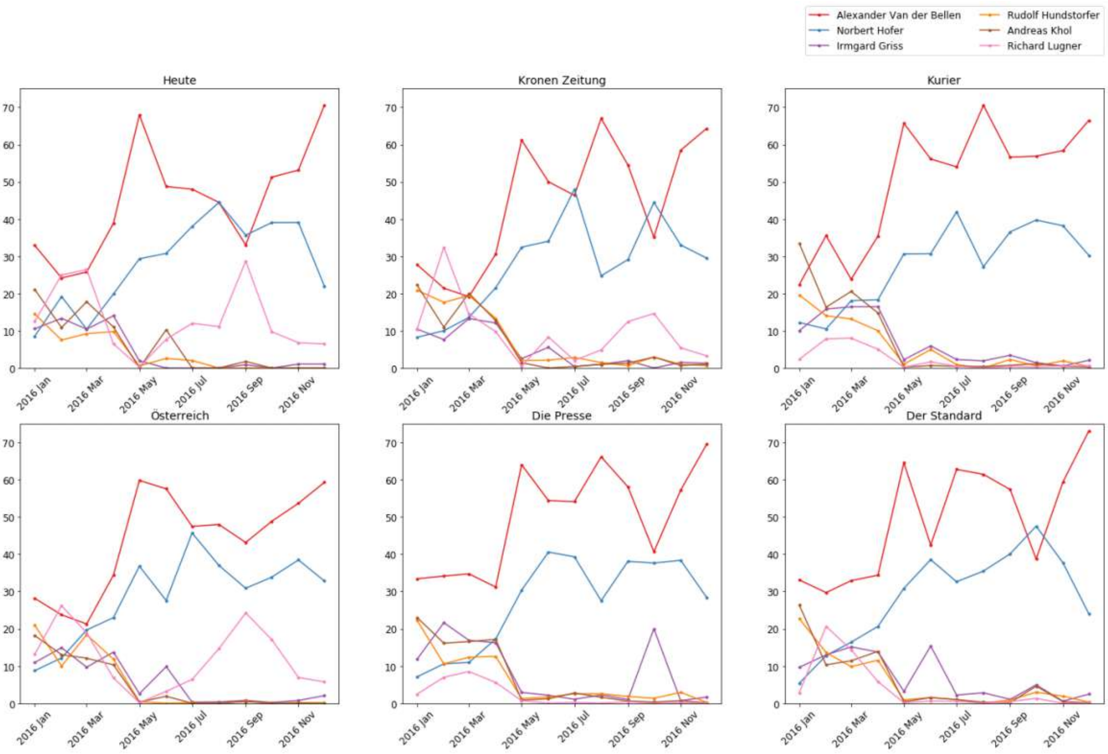


Figure 5.9: Mentions of presidential candidates proportional to another per month.

Finally, we visualize in figure 5.10 which parties received the most mentions. We calculate party mentions by summing up the number of detected named entities of parties and their members. The entities and party affiliations were retrieved from the Wikidata platform (see Appendix section 8.1 for a list of names of all downloaded members with party affiliation). We excluded the presidential candidates in this procedure as their situation was investigated in the previous paragraphs. We also decided not to include the president at that time, Heinz Fischer, since, in his role, he is expected to be impartial. The number of members retrieved from Wikidata for each party is as follows: SPÖ (1906 members), ÖVP (1678 members), FPÖ (526 members), Grüne (169 members), and NEOS (16 members). These numbers are not completely surprising as the SPÖ and ÖVP are long-established parties in Austria and strongly present in all states. The NEOS, with only 16 members on Wikidata, is, in stark contrast, a new party with little prominent representation in the states.

Figure 5.10 shows that across all newspapers, the governing parties, SPÖ and ÖVP, were mentioned the most, whereas the SPÖ, as the party of the Federal Chancellor in this term, was mentioned the most. This ranking separation between SPÖ on top and ÖVP below is most visible in the "Krone". The spike in reporting on the SPÖ in May is due to a transition of power from the former chancellor Werner Faymann to the new one, Christian Kern. The graph further shows that the high-quality newspapers, including "Kurier", report in similar frequencies on the Grüne and the FPÖ, whereas the tabloids favor the FPÖ. The NEOS is mentioned across all papers with low frequency compared to the other parties. Most surprising to us was how neatly and continuously the ranking of party mentions aligned with the last election's results. We think this may be due to similar editorial policies in the newspapers that may give more space to more popular parties or that governing parties produce more news.

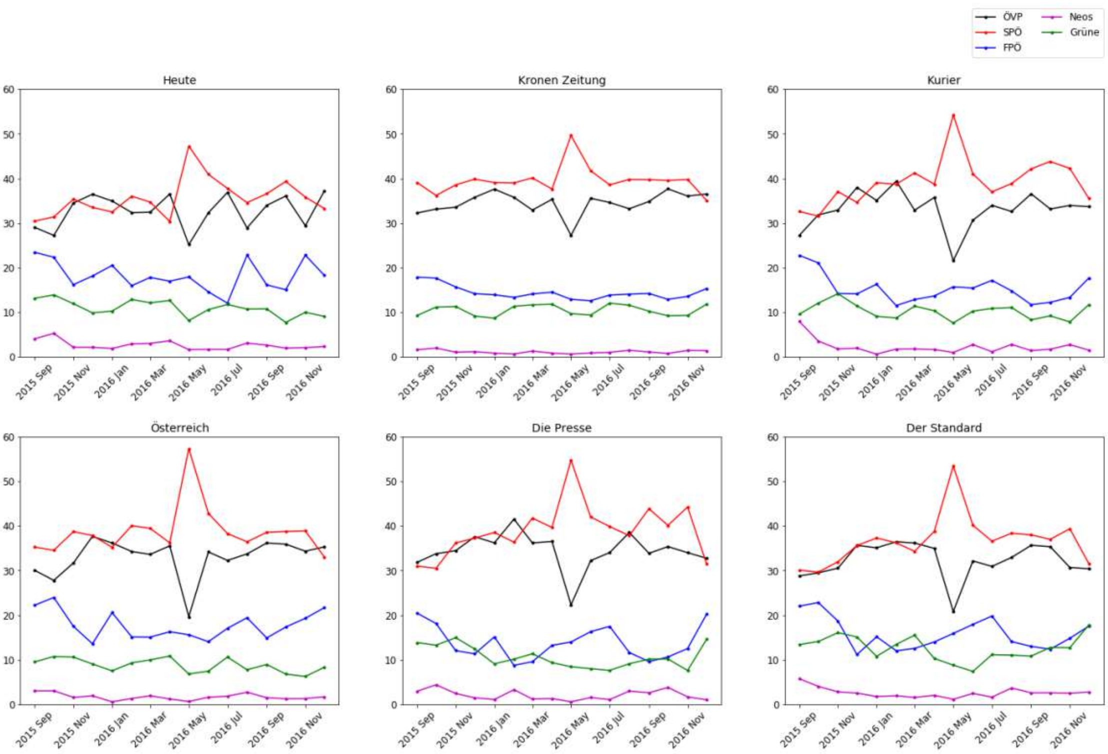


Figure 5.10: The number of mentions of parties proportional to another per month.

5.3.2 Gender bias

In this section, we explore gender bias in the newspapers. To conduct such an analysis of inequalities, we retrieved the recorded genders (only female and male were available) of all Austrians on Wikidata. In total, we found 27.650 male and 5.851 female entities. The underrepresentation and exclusion of women on Wikipedia is a well-studied fact [Tri23] and is also reflected in our data. Consequently, the results we present here must be read carefully as they also contain Wikipedia's gender bias. Figure 5.10 shows the gender divide in mentions across newspapers and time in absolute numbers. It highlights very clearly that a substantial disparity is present in all newspapers.

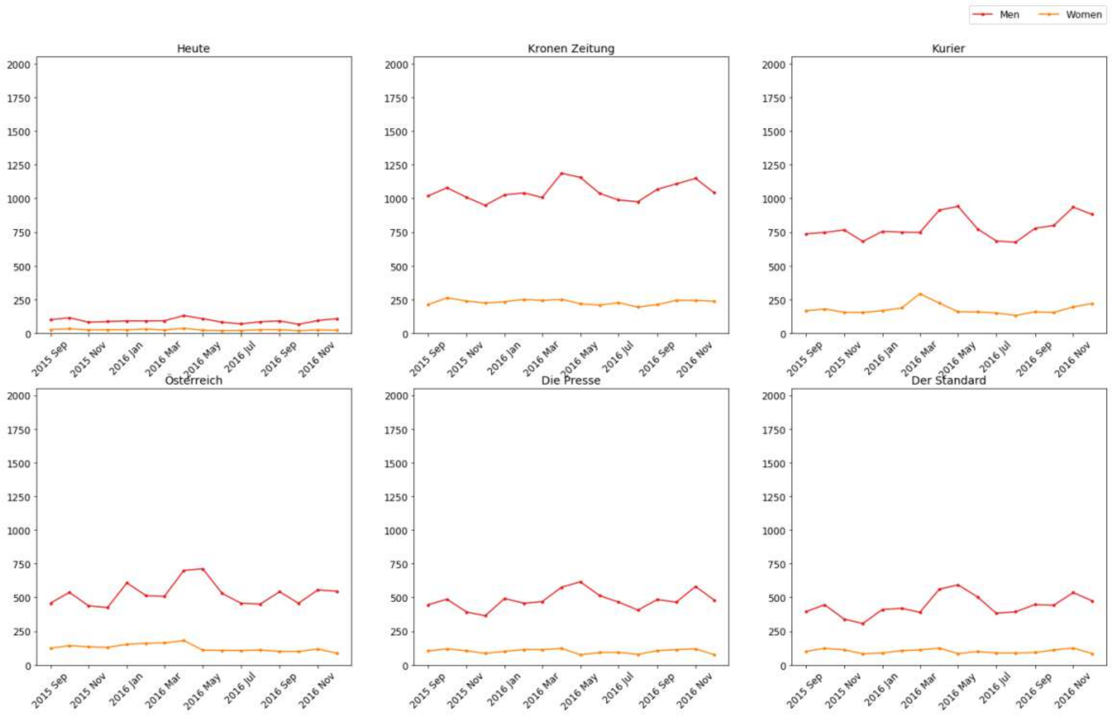


Figure 5.11: Total number of mentions of women and men per month.

The figure 5.12 shows the share of mentioned Austrian women (named entities) across different newspapers. The gray line indicates how many women were named at least once compared to men. The red line illustrates the cumulative share of all mentions of women. Finally, the orange line illustrates the proportion of women compared to men in the Wikidata dataset as a baseline. The lines thereby ultimately show that fewer women than men are named at least once in the newspapers. Furthermore, the named women are also less frequently referenced in the texts. This result is statistically significant according to a t-test (see table 5.3) as the t-statistics are positive and the p-values much smaller than 0.05.

	t-statistic	p-value
Heute	10.991556	9.948021e-17
Krone	10.594716	4.886063e-16
Kurier	10.724289	2.900627e-16
Österreich	9.280468	1.054644e-13
Presse	5.318933	1.257771e-06
Standard	5.471256	6.947208e-07

Table 5.3: Statistical significant: Women (named at least once) are even less frequently mentioned.

While studying figure 5.12, we also notice that across the different newspapers, gender bias varies. The “Presse” and the “Krone” name the fewest women on average with about 19% and 20% respectively. In “Heute” in contrast, about 25.28% of mentioned names are women. When considering the frequency of mentions, the newspapers exhibit similar bias, with “Heute” naming women the most (about 21.32%).

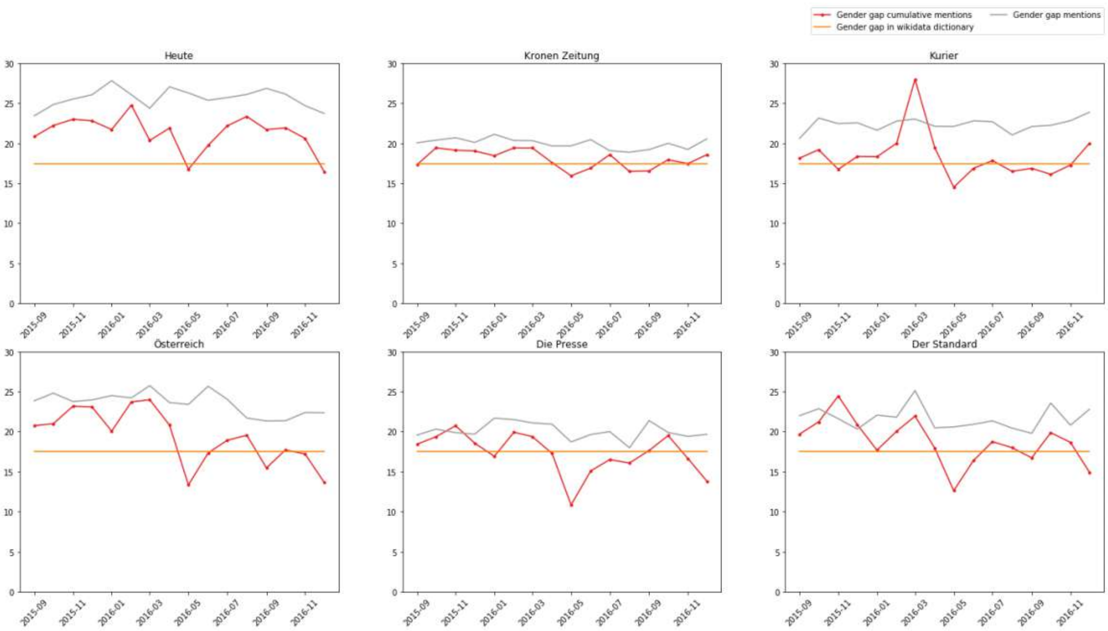


Figure 5.12: Mentions of Austrian women compared to men across time per newspaper.

In figure 5.13, we illustrate gender bias based on mentions of Austrian politicians. The set of women named is similar across all newspapers, which may indicate that all papers report on some shared subset of events when politicians are involved. In contrast, the frequency of mentions varies more strongly, but there is also a high overlap in certain months. The ratio of women to men in our Wikidata entity set is more biased against women than the subset based on female Austrian politicians named at least once. This difference may be because female politicians are less present on Wikipedia than in the news, and our subset consists of all Austrian politicians alive - in previous decades, even fewer women than now were holding political offices. The figure again shows that fewer women than men are named in the newspapers. Furthermore, those named are also less frequently invoked in the texts than men. This result is statistically significant according to a t-test (see table 5.4).

	t-statistic	p-value
Heute	5.738607	2.419201e-07
Krone	6.717817	4.555367e-09
Kurier	8.848522	6.338181e-13
Österreich	10.553269	5.775040e-16
Presse	4.760943	1.047682e-05
Standard	6.100934	5.660001e-08

Table 5.4: Statistical significant: Female Austrian politicians (named at least once) are even less frequently mentioned.

The graph also shows that the cumulative mentions and the naming of female Austrian politicians are consistently apart in all the included months.

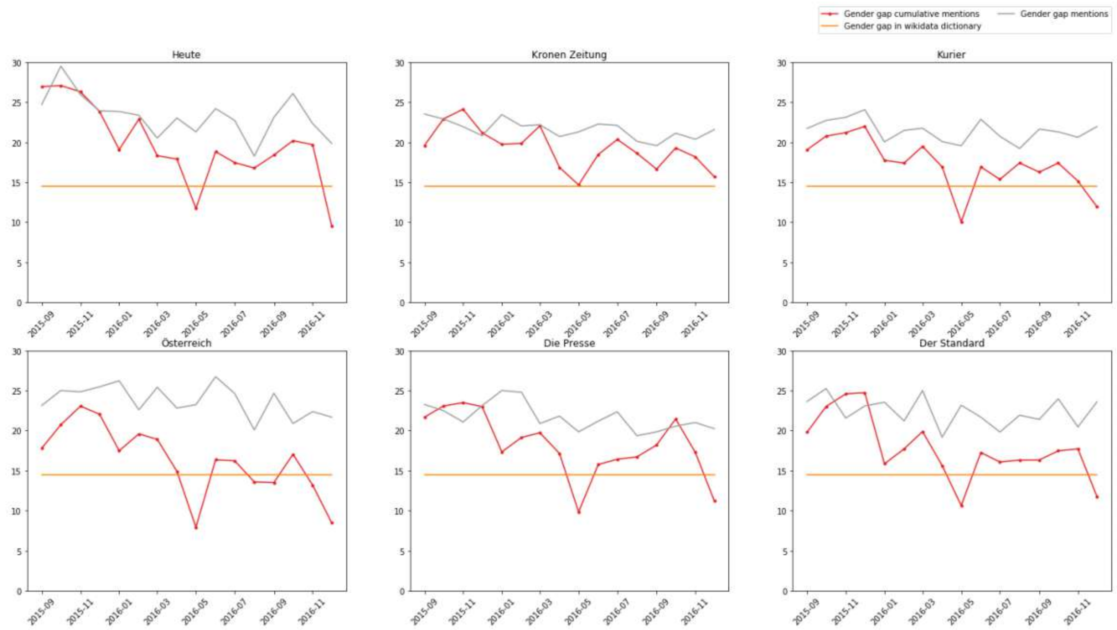


Figure 5.13: Mentions of female Austrian politicians to male Austrian politicians across time per newspaper.

5.3.3 Emotion and Sentiment Analysis

We also tried to employ sentiment and emotion dictionaries to characterize the different newspapers based on how frequently certain words are mentioned. These aggregations are made over many varied texts, and therefore many nuances may be lost, but we still hoped that strong signals in the data become visible. In figure 5.14, we show proportional positive and negative sentiment across newspapers based on matching words and summing up given weights from several sentiment dictionaries. The graph shows we found more negative than positive sentiment in all newspapers. The proportions are quite similar over time and in different newspapers. This may show that consistent, mostly neutral language is used in reporting, but we think it more likely questions the validity of this dictionary-based approach to characterize sentiment. Future work could explore machine learning-based approaches trained through corpus samples since they may provide more conclusive results.

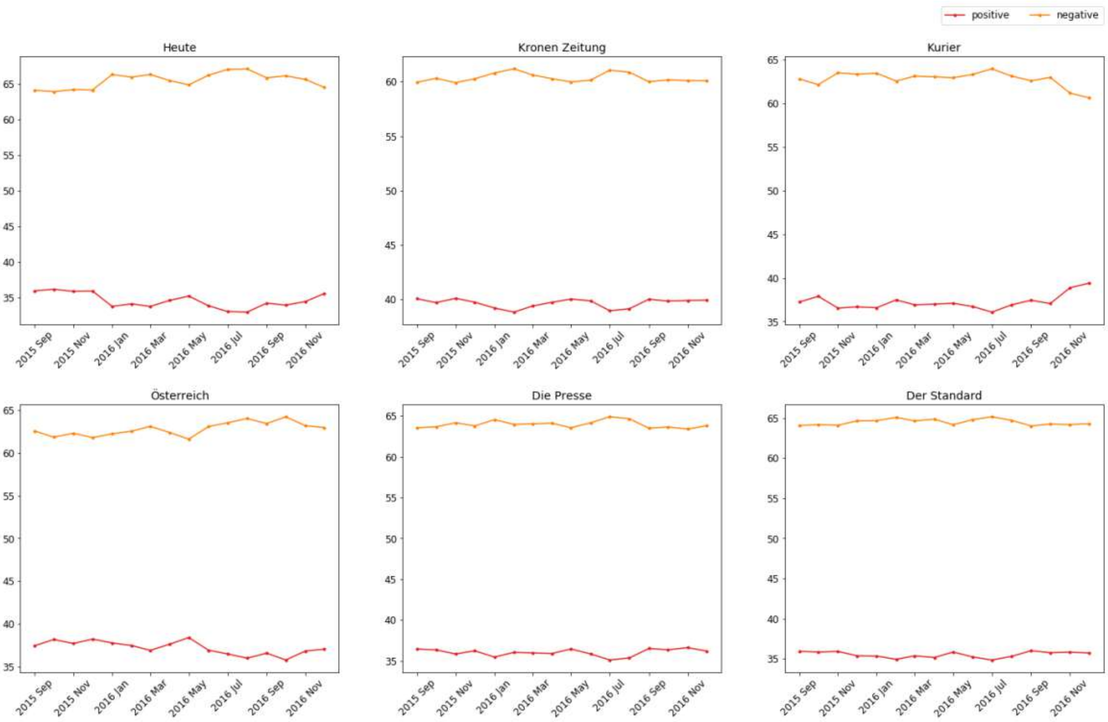


Figure 5.14: Sentiment across time per newspaper.

In contrast, figure 5.15 highlights proportional counts of mentions of emotional words, providing a better basis for differentiating the newspapers. Disgust ("Ekel") and anger ("Wut") are mentioned the least across all papers. The "Krone", "Österreich" and "Kurier" features many articles with words representing happiness ("Freude") compared to the other papers. These are also the papers with the greatest number of articles. Fear ("Furcht") is common in the tabloid papers "Österreich" and "Heute". Interestingly, the quality newspapers "Standard" and "Presse" express a similar number of emotional words over time.

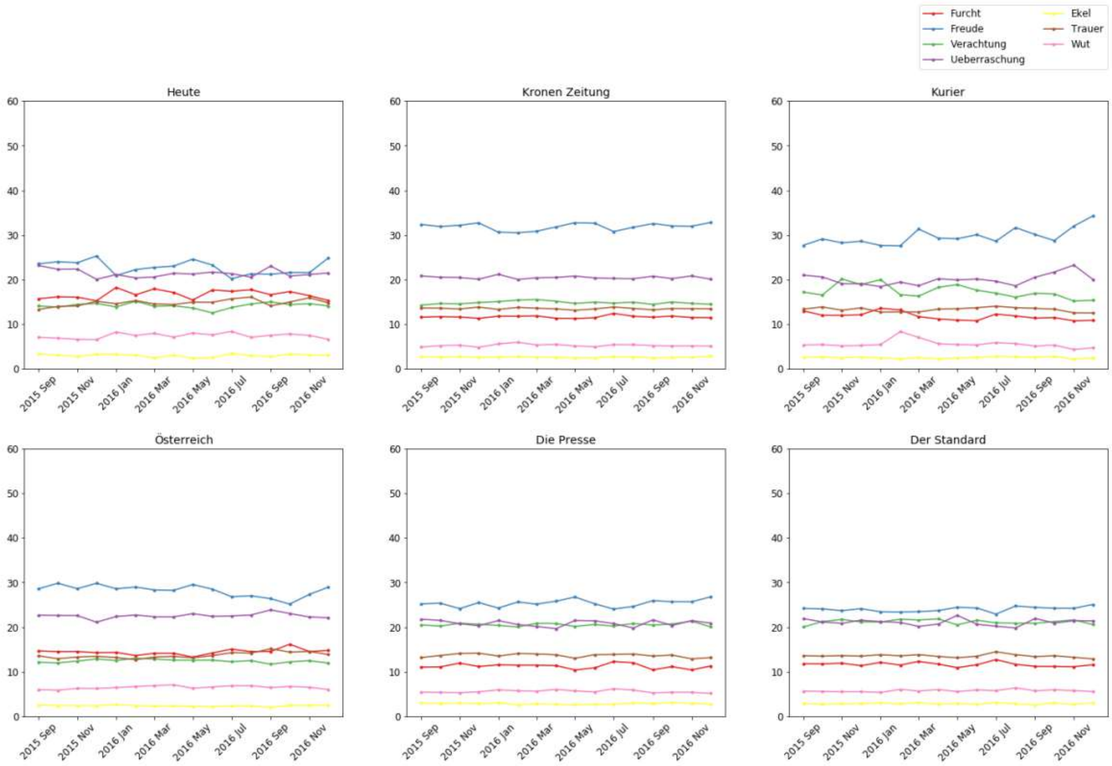


Figure 5.15: Emotion across time per newspaper.

5.3.4 Quantitative analysis of writing quality

In this section, we further illustrate the differences between tabloid and high-quality newspapers by presenting two measures for readability. The rankings based on simplicity/complexity of language may provide further insights into which audiences the newspaper may intend to target. The "Wiener Sachtextformel" [5] is calculated as follows:

$$\begin{aligned} score = (19.35 * \frac{\#polysyllable\_words}{\#words}) + (0.1672 * \frac{\#words}{\#sents}) + \\ (12.97 * \frac{\#long\_words}{\#words}) - (3.27 * \frac{\#monosyllable\_words}{\#words}) - 0.875 \end{aligned} \tag{5.1}$$

[5] `https://chartbeat-labs.github.io/textacy/api_reference/misc.html#textacy.text_stats.wiener_sachtextformel`

It is important to note that the higher the score, the more difficult a text should be. The other measure we list is the "Flesh reading ease" for the German language [6] and it is derived as follows:

$$score = 180.0 - \frac{\#words}{\#sents} - (58.5 * \frac{\#syllables}{\#words}) \tag{5.2}$$

The scores range from 0 to 100, and in contrast to the "Wiener Sachtextformel", a lower score indicates a less difficult text. In our opinion, the "Wiener Sachtextformel" is the better measure of the two, as it was specifically developed for the German language. As table 5.5 illustrates, the scores appear to match the high-quality and tabloid paper classification we presented earlier. The coloring of cells in the column "Wiener Sachtextformel" represent clusters based on scores interpreted as similar. These classifications of language use ultimately indicate that the classical tabloid papers have shorter sentences and use words with fewer syllables. It also reveals that there may be a gap in Austria's media landscape for a high-quality print outlet that uses more easily readable and accessible language.

Newspaper	**Wiener Sachtextformel**	**Flesh reading ease**
Heute	5.23	126.36
Krone	5.34	127.04
Österreich	5.37	125.08
Kurier	5.82	122.58
Presse	6.43	117.82
Standard	6.48	117.71

Table 5.5: Writing quality according to two widely used indices.

We also calculated readability scores for the different topic categories and gained interesting insights. We found that sports (German: "Sport") across all newspapers has mostly low readability scores (i.e. contains less 'difficult' text). Culture (German: "Kultur") and the title pages (German: "Titelseite") also have low scores in newspapers except for the high-quality ones, "Standard" and "Presse". We found mostly high readability scores for the topics economy (German: "Wirtschaft"), chronicle (German: "Chronik"), politics (German: "Politik"), and opinion (German: "Meinung").

[6] `https://chartbeat-labs.github.io/textacy/api_reference/misc.html#textacy.text_stats.flesch_reading_ease`

CHAPTER 6

Network analysis of the Austrian presidential election

This chapter presents our results on using co-occurrence networks for news analysis. Such a network-based approach is intended to center relations between entities and thereby provide a different perspective into the AMC beyond analysis based on counting the occurrence of mentions. First, we highlight how filtering and parameter tweaking are needed to make visualized networks more interpretable. Then, we apply a cluster detection algorithm on the resulting networks to show themes and topics that can be identified in the collection of articles. Lastly, we explore using a node centrality measure to highlight entities presented as highly relevant in election coverage.

6.1 Comparing networks

We present in this section the analysis and interpretation of constructed co-occurrence networks that represent timeframes with a lot of election reporting. We also do not consider the election day and the two following days to avoid reporting focused on election results. The included timeframes encompass: 1st of January to 23rd of April, 28th of April to 21st of May, and 1st of October to 3rd of December. The nodes in the networks are extracted nouns and adjectives, which were found to be particularly well-suitable for text summarization [MT04]. The relations are co-occurrences in a sentence. We decided on this configuration after trying to interpret differently constructed networks, such as ones with verbs or where relations are formed after co-occurrence in an article.

The present visualizations of text co-occurrence networks are meant to provide a different perspective on the corpus, focusing on relations, which contrasts the descriptive statistics presented in the previous chapter 5. They provided a more atomized analysis of entities based on mentions. The construction and analysis of appropriate networks require parameter tweaking, filtering, and interpretive work to deal with the messiness and complexity of the underlying unstructured texts. It is an explorative and iterative task that requires researchers to try out until useful results are generated. The concept of usefulness is qualitative and interpretative, which means researchers need to be closely involved in the analysis. This also poses a danger of confirmation bias, but on the other hand, this flexibility also enables researchers to tune results based on contextual knowledge. We use the available parameters to focus on more prominent trends and signals in the data to address this challenge. For instance, networks that contain too many nodes and edges can become difficult to read, but with appropriate filtering, the focus can be put on the most relevant insights. We illustrate in this section how different visualizations can be used to make sense of differences in the reporting of the six news outlets.

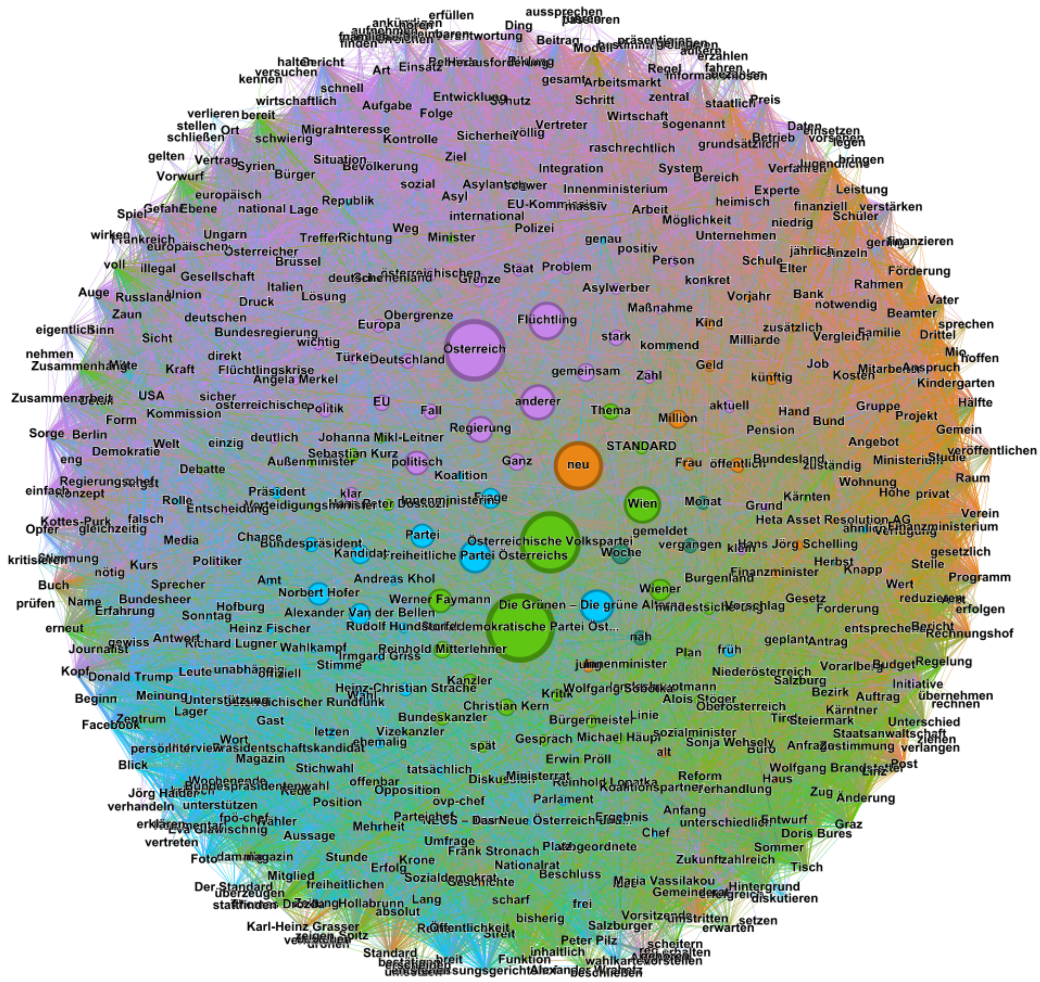
6.1.1 Configuring networks

Parameters and filters need to be tweaked to make the aggregated text networks visually appealing and more easily interpretable. This process is about putting the focus on specific aspects of the data; in our case, these are strong trends. We illustrate below how a co-occurrence network based on "Standard" articles was iteratively refined with Gephi.

In all visualized networks, Pagerank determines the size of the nodes to illustrate their centrality in the networks, and the color of the nodes highlights clusters according to the Louvain community detection algorithm. The Force Atlas 2 layouting algorithm [JVHB14] was applied to better accentuate clusters spatially.

A few descriptive statistics accompany the network visualizations to quantify some of the structural properties of the network. The "#Nodes" and "#Edges" columns record the number of nodes and edges in the networks. The "Avg. Degree" column details the average degree of nodes based on weighted edges, and "Avg. Path" describes the average path length between nodes. The "Modularity" column provides a measure of the robustness of the detected community (see section 3.3.1), and "Components" lists the number of connected components. The descriptive statistics and the process to generate the network visualizations are for our comparative analysis in section 6.1.2 and section 6.1.3.

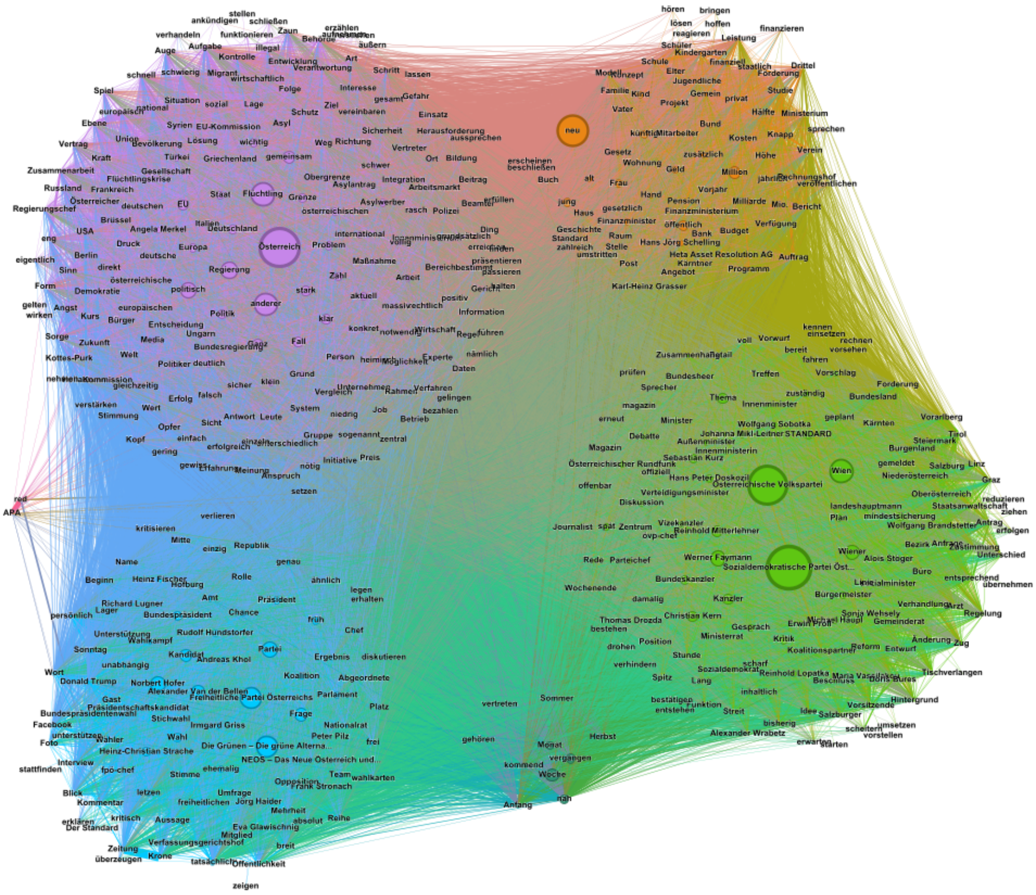
The figure 6.1 shows the network without any filtering. The resulting network is so extensive that it is difficult to interpret, and clusters, visible through node coloring, meant to illustrate themes, seem to overlap.



#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
515	48470	509.814	1.666	0.131	1

Table 6.1: A filtered co-occurrence network based on Standard articles during the election.

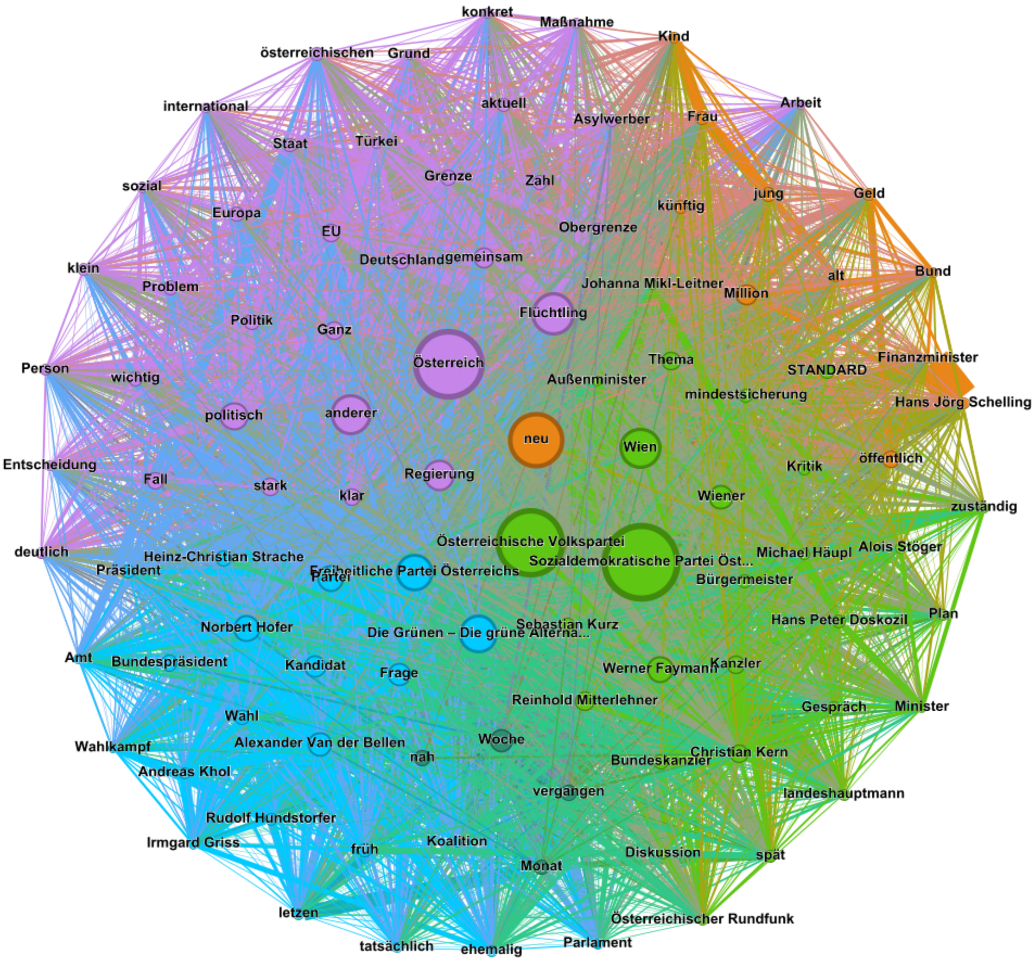
The figure 6.2 shows the same network again, but each identified cluster was moved into a different corner of the visualization. This arrangement makes recognizing the different clusters as topics or themes easier.



#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
515	48470	509.814	1.666	0.131	1

Table 6.2: A filtered co-occurrence network based on Standard articles during the election.

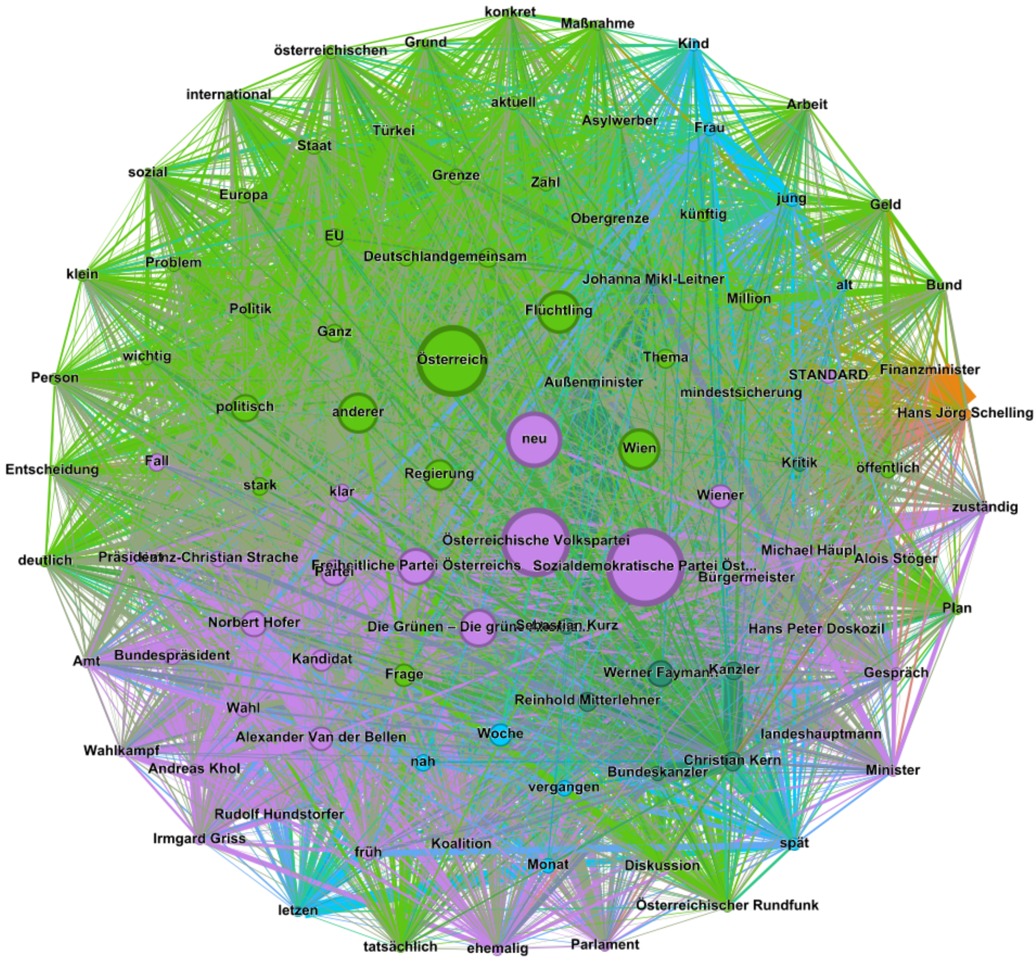
The figure 6.3 shows the network after all nodes except the 100 most prominent ones, according to Pagerank, were filtered out. This makes the network and the communities more easily readable, but the many connections between nodes still make it seem chaotic.



#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
100	4472	648.66	1.097	0.131	1

Table 6.3: A filtered co-occurrence network based on Standard articles during the election.

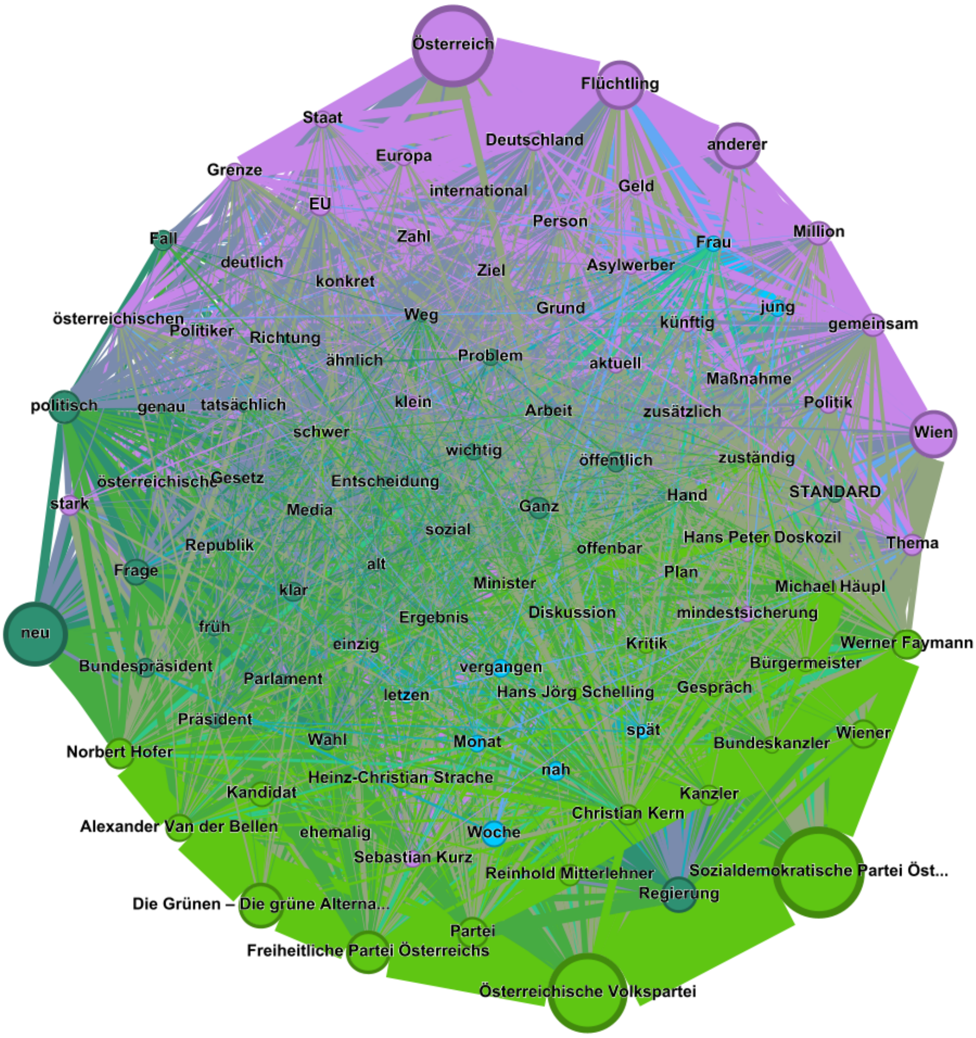
The figure 6.4 shows the network after community detection was applied again. Different clusters were identified, highlighting how the algorithm is not deterministic and results are influenced by randomized parameters. This sensitivity to parameterization points to a fluidity of identified themes that need to be considered when interpreting the results.



#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
100	4472	648.66	1.097	0.113	1

Table 6.4: A filtered co-occurrence network based on Standard articles during the election.

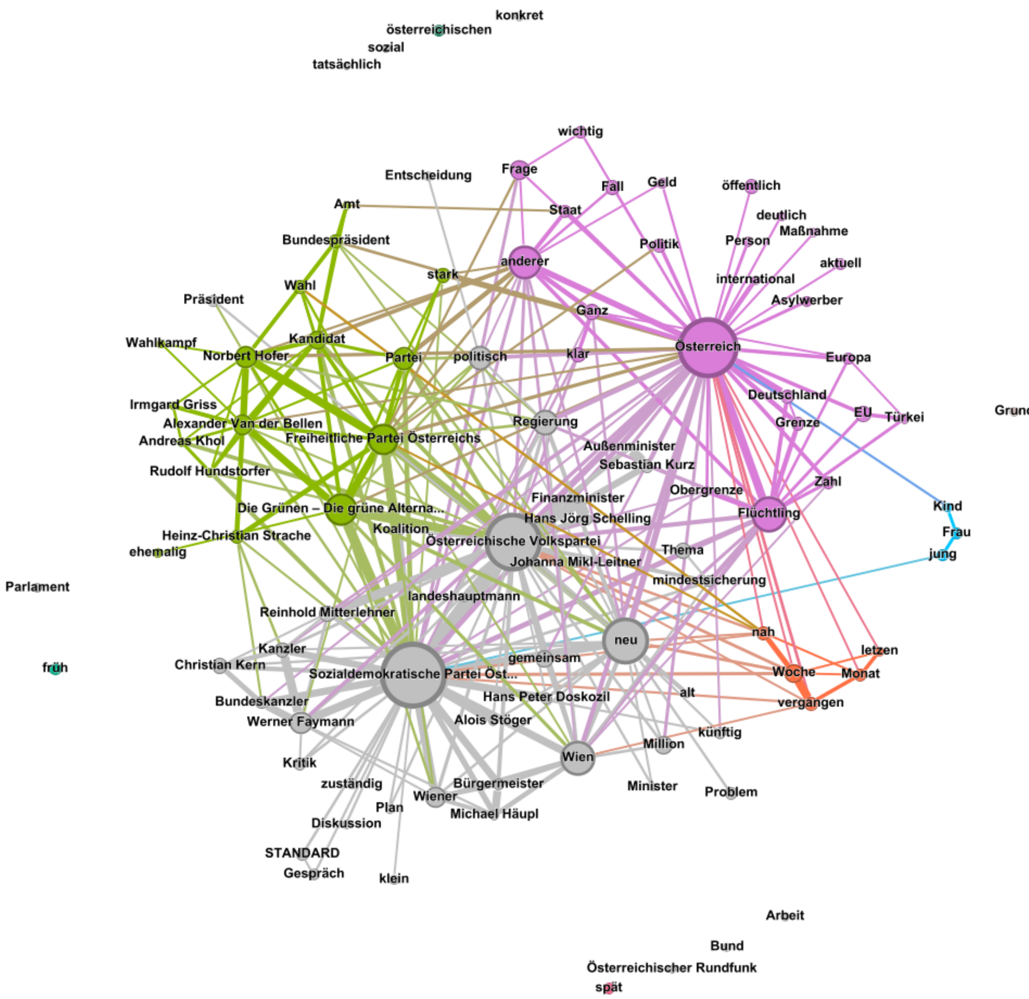
The figure 6.5 shows the network after the parameters for the layouting algorithm were tweaked. The different arrangement of nodes highlights how layouting also has lots of flexibility through parameterization. There is a danger that depending on how nodes and relations are displayed, results may be interpreted differently, which could enable misinterpretation and misuse.



#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
100	4472	648.66	1.097	0.097	1

Table 6.5: A filtered co-occurrence network based on Standard articles during the election.

The figure 6.6 shows a visualization where less prominent edges were removed. This filtering significantly increases the modularity of identified clusters and makes the network more easily interpretable. After the filtering, only less than 300 relations are visualized. It decreases the average unweighted degree of nodes to about five or six edges per node.



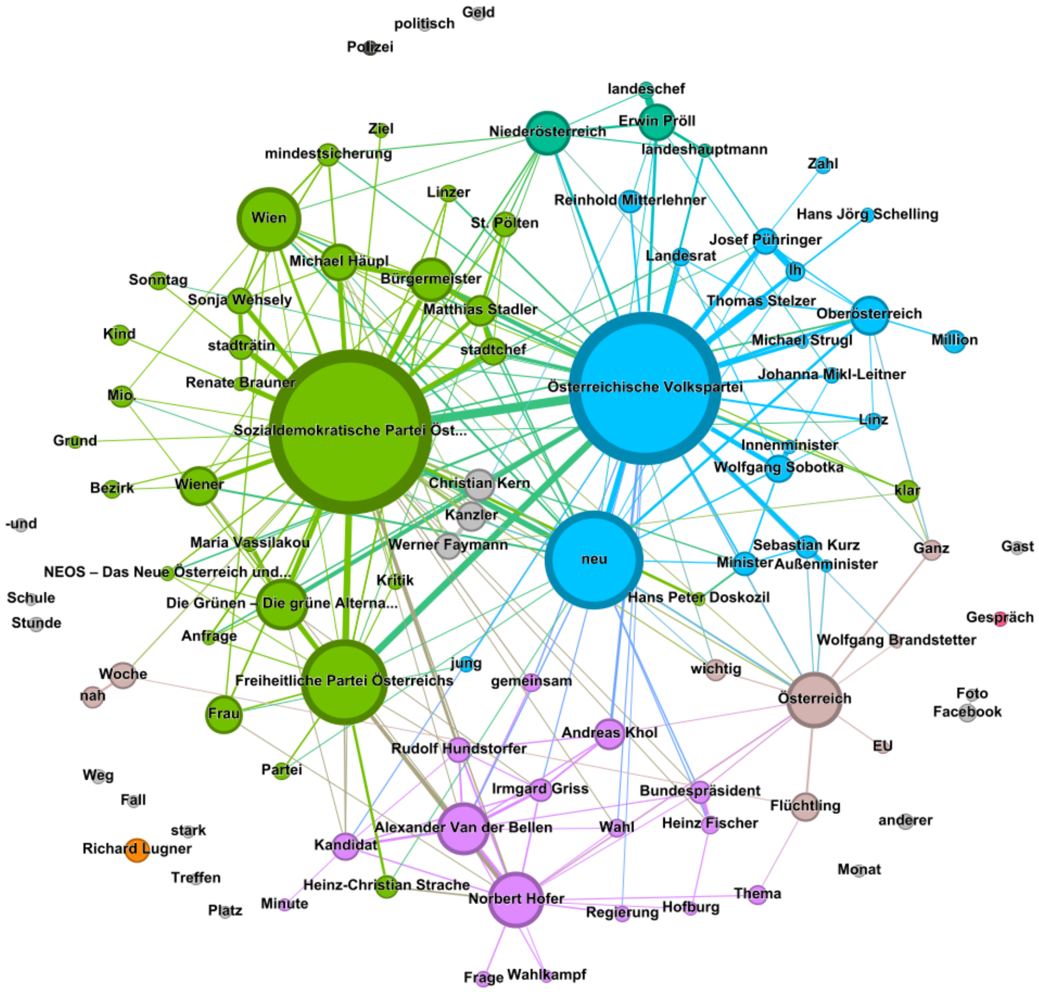
#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
100	287	213.38	2.267	0.344	12

Table 6.6: A filtered co-occurrence network based on Standard articles during the election.

6.1.2 Comparing politics-coverage networks

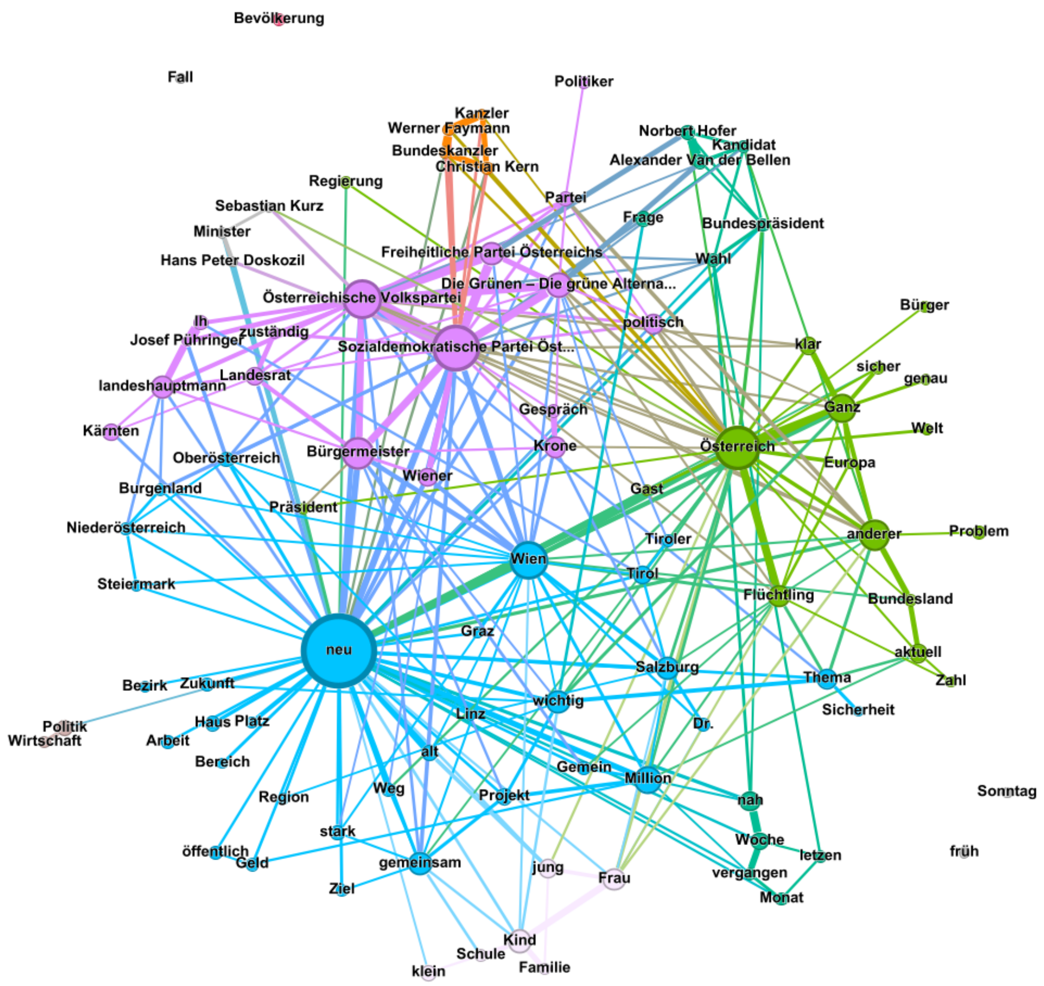
We analyze network visualizations for all newspapers created as described in section 6.1.1. Some of the most central nodes in all networks include the Austrian parties, states, and various politicians. The word new (“neu”) is also central in all networks and often connected to parties, candidates, locations, and topics in domestic politics. We theorize this may be due to issues in our data cleaning and newspapers more generally reporting mainly on “new” events. All of the networks contain some words that could be classified as stopwords, highlighting possibilities for improvement of our NLP pipeline in future work. Furthermore, some of the strong relations are between job titles of politicians such as “Erwin Pröll”, “Niederösterreich” and “landeschef/hauptmann” for a prominent governor in figure 6.7. Future work, could seek to make the networks even more expressive by combining such strongly related nodes. In most networks, separate clusters containing the two final presidential candidates and keywords indicative of an election were identified. Some also contain parties (most often the Grüne and FPÖ) and other presidential candidates. The presidential candidates were most prominent in Heute and Österreich, in other papers reporting on domestic politics appears to be even more dominated by the governing parties SPÖ and ÖVP. The parties are often strongly connected to their members and thereby also part of the same clusters. All networks have one or two clusters for the two major parties, the SPÖ and ÖVP. They also all have clusters for important domestic political topics during the election, such as social security, education, policing, national security, and refugees. However, there are different words associated with the clusters. For example, the networks representing “Österreich”, “Standard”, “Kurier”, and “Presse” also mention the word threshold in relation to refugee.

We unpacked identified clusters as themes but want to stress that this usually involves ignoring some nodes and connections as the clusters are messy and don’t neatly represent topics. Relatedly, there is a danger of conformation bias as the interpretative flexibility of clusters affords different ways of reading and understanding them. Both the social sciences and humanities have developed ways to deal with this kind of ambiguity and uncertainty and don’t see them necessarily as problems but instead embrace how different interpretations are possible. This requires adopting an epistemological stance that questions objectivity and instead ascribes to different values like plurality of perspectives, which are still uncommon ways to interpret results in computer science. Still, we think when engaging with text as cultural products, it is important to adopt such critical stances. The parameterization, which allows researchers to construct and visualize networks differently, further increases flexibility, possibly resulting in errors and misuses. Furthermore, the clustering algorithms are not statistically consistent and depend on randomization, which means that certain results may not be easily reproducible and clusters should be seen as analytical guidance and not always definitive. There is no standard, established way to visualize and construct text networks, so researchers should embrace that these data science methods can only focus on certain partial aspects of the analyzed texts and not provide completely certain results. It is important that limits are discussed and the partiality of perspectives and approaches reflected.



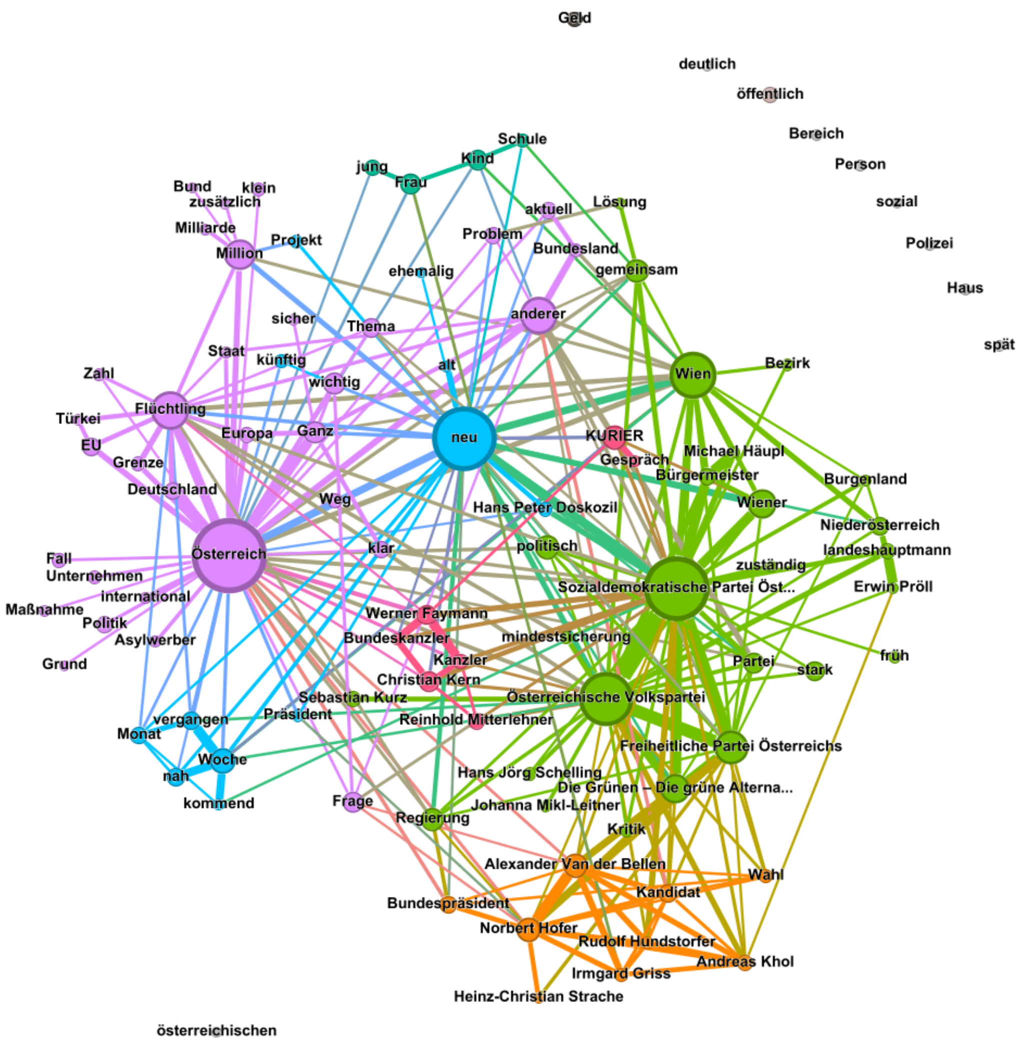
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	99	244	41.616	2.293	0.37	18
Full	360	6980	60.978	1.972	0.21	1

Table 6.7: A filtered co-occurance network based on Heute articles during the election.



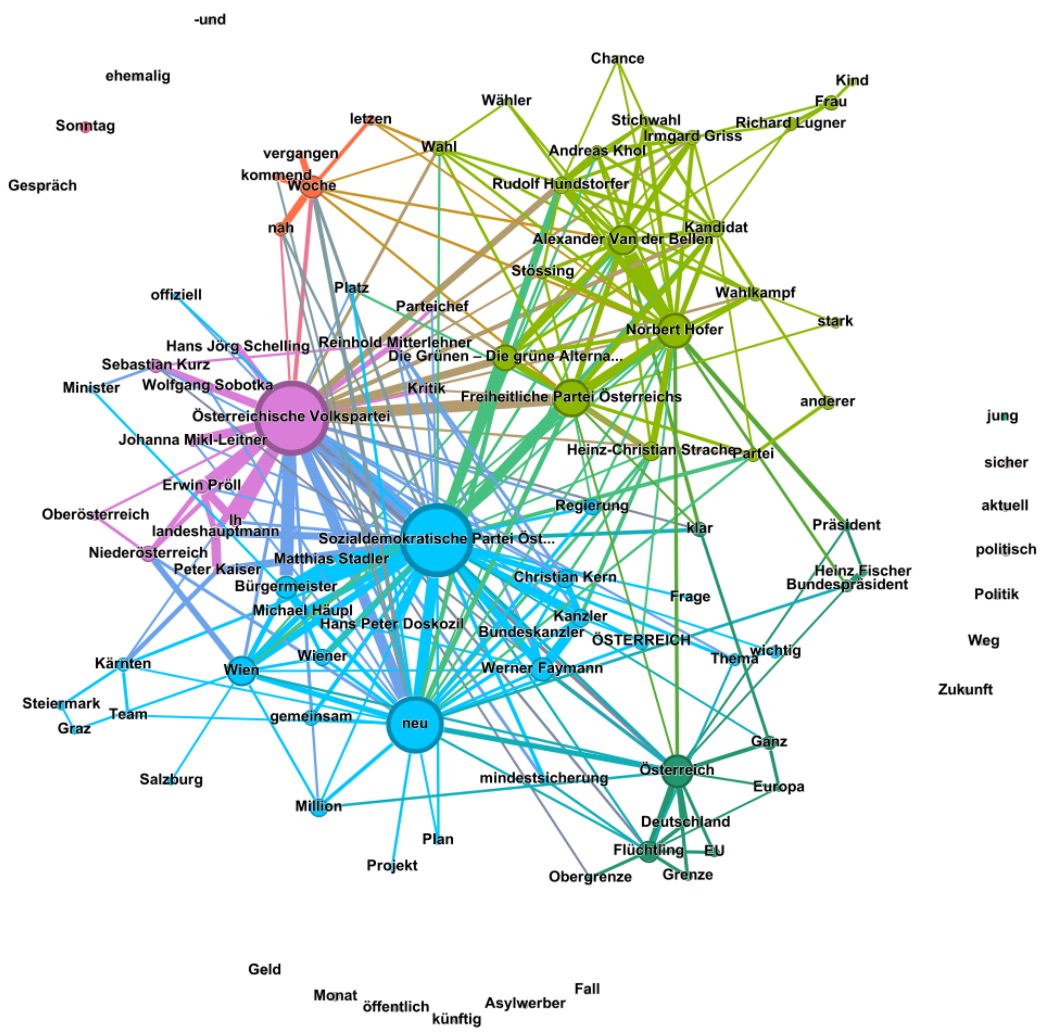
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	96	288	179.792	2.308	0.37	5
Full	516	45913	490.171	1.703	0.131	3

Table 6.8: A filtered co-occurrence network based on Krone articles during the election.



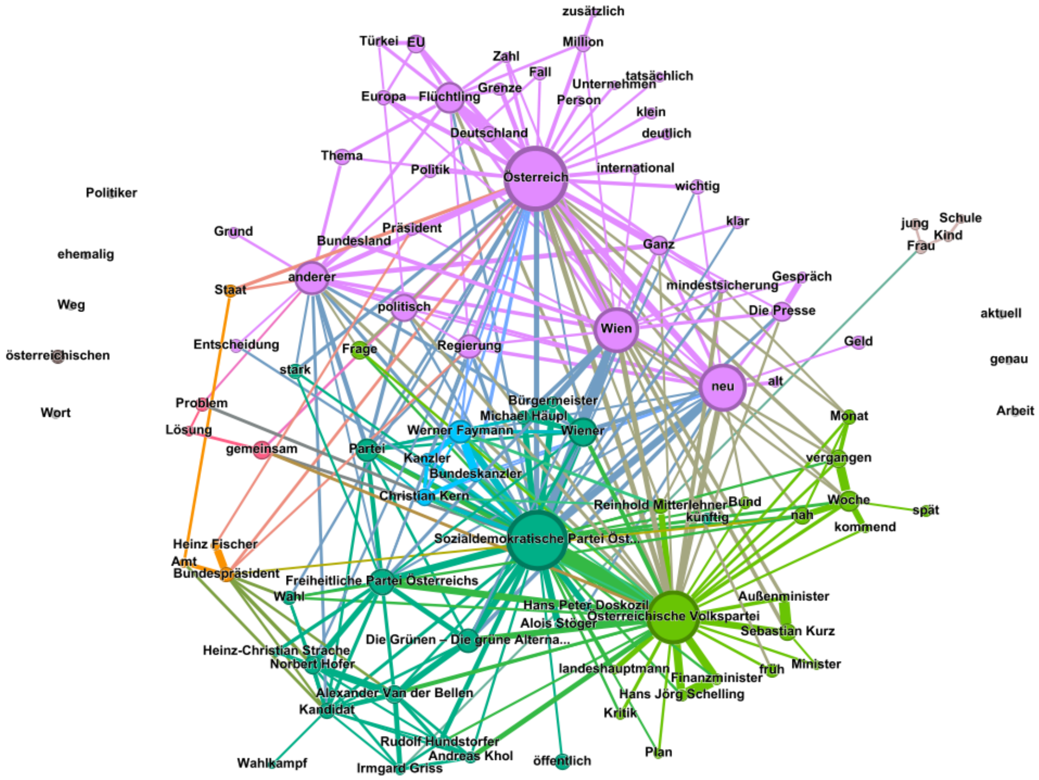
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	100	285	214.04	2.319	0.38	11
Full	503	46097	527.614	1.681	0.134	1

Table 6.9: A filtered co-occurrence network based on Kurier articles during the election.



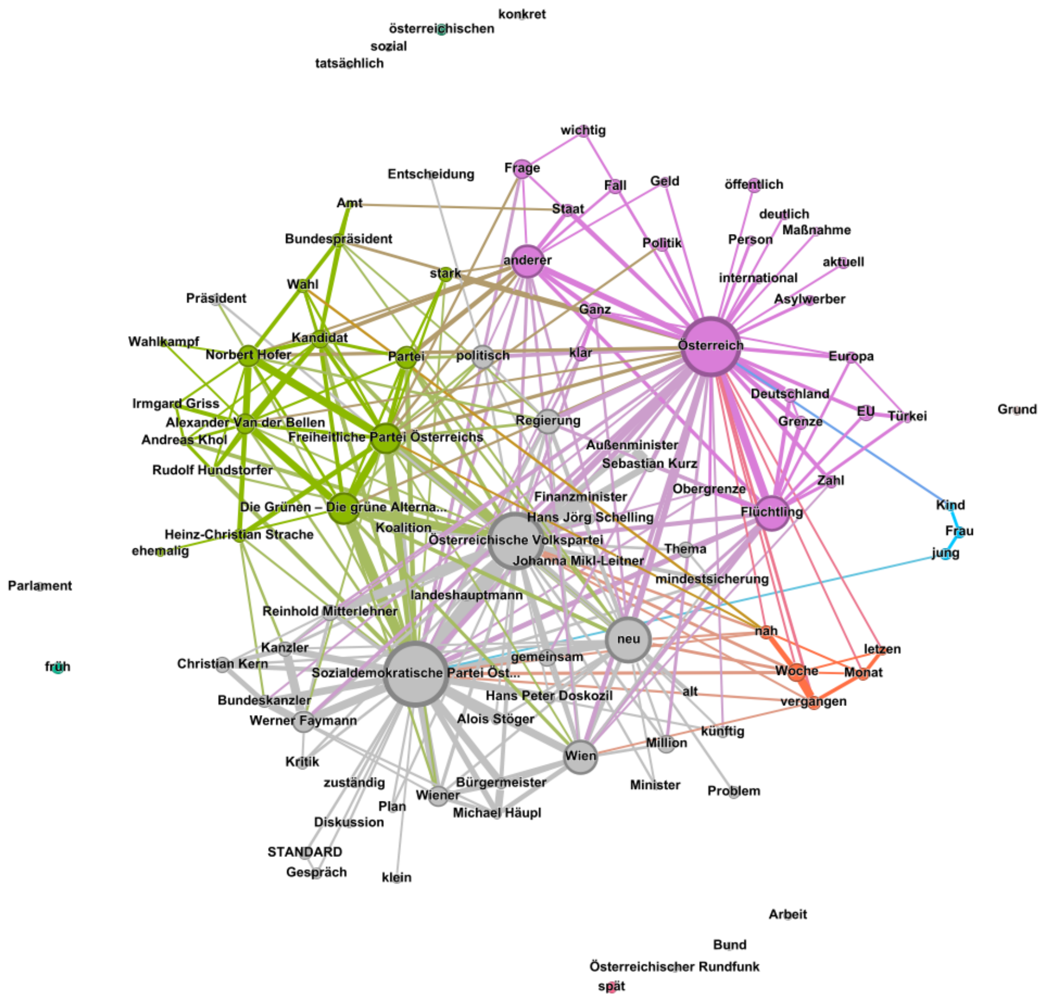
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	99	272	198.929	2.338	0.366	18
Full	472	30826	353.305	1.778	0.163	1

Table 6.10: A filtered co-occurrence network based on Österreich articles during the election.



Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	99	291	219.01	2.527	0.33	9
Full	551	53307	522.051	1.694	0.125	1

Table 6.11: A filtered co-occurrence network based on Presse articles during the election.



Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	100	287	213.38	2.267	0.344	12
Full	515	48470	509.814	1.666	0.131	1

Table 6.12: A filtered co-occurrence network based on Standard articles during the election.

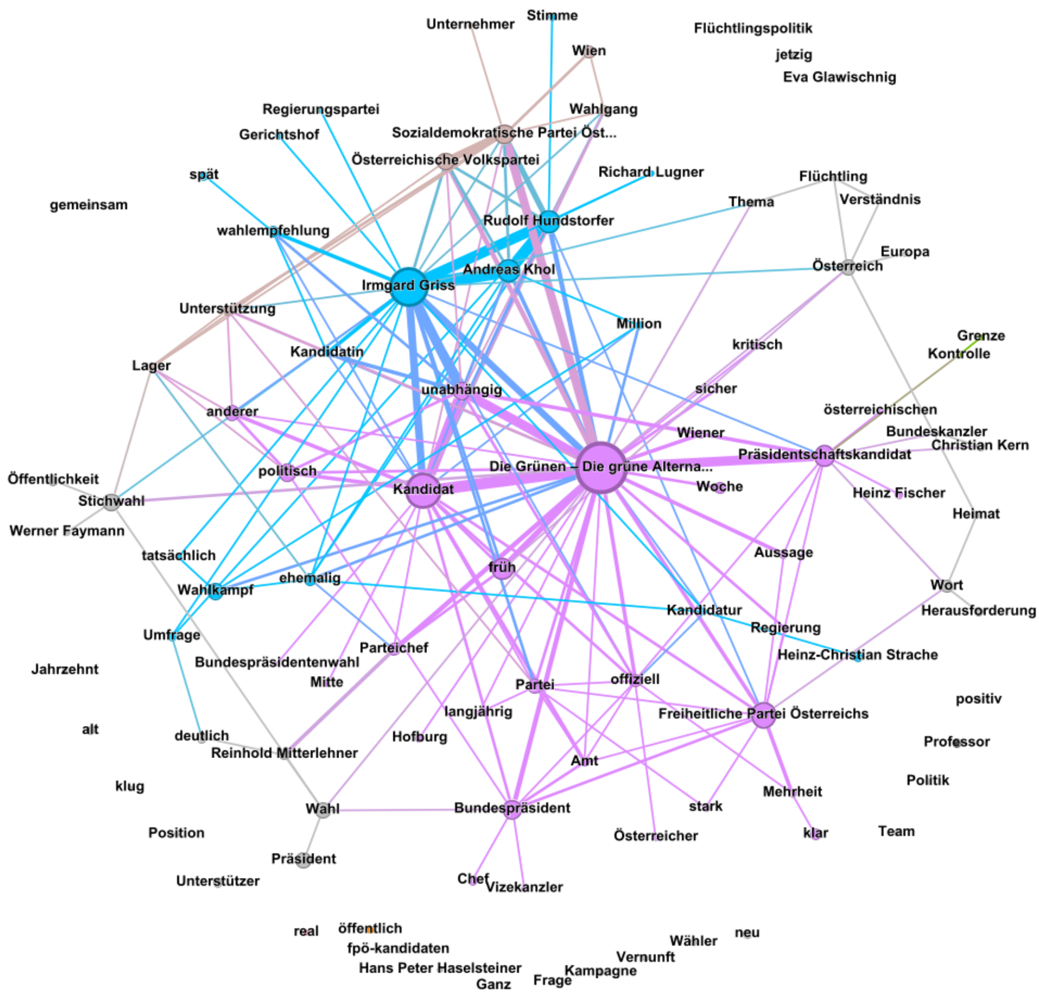
6.1.3 Comparing egocentric networks

We also constructed ego networks [Fre82] for each of the two final presidential candidates. They only present nodes directly related to the candidates and exclude their opponents. These networks are intended to provide a more in-depth comparison between political figures by illustrating their neighborhoods in the previously discussed politics-coverage networks (see section 6.1.2). We decided to use only two newspapers to highlight the efficacy of these visualizations.

The networks on Alexander Van der Bellen (see figure 6.13 for “Presse” and figure 6.14 for “Österreich”) present several clusters with some overlap. There is one cluster on the other presidential candidates, which is quite similar in both networks. The “Presse” has a cluster encompassing Van der Bellen’s Grüne and the FPÖ of Hofer. There is also a small cluster for the ÖVP and SPÖ. In contrast, the “Österreich” has a cluster encompassing the Grüne, SPÖ, and ÖVP, and a separate one for the FPÖ. The FPÖ cluster contains many nodes related to the party, which stands in contrast to “Presse”. There are also topics and international figures associated with this FPÖ cluster, like a refugee node and one on Donald Trump’s win. In “Presse,” the refugee node is part of a small separate cluster.

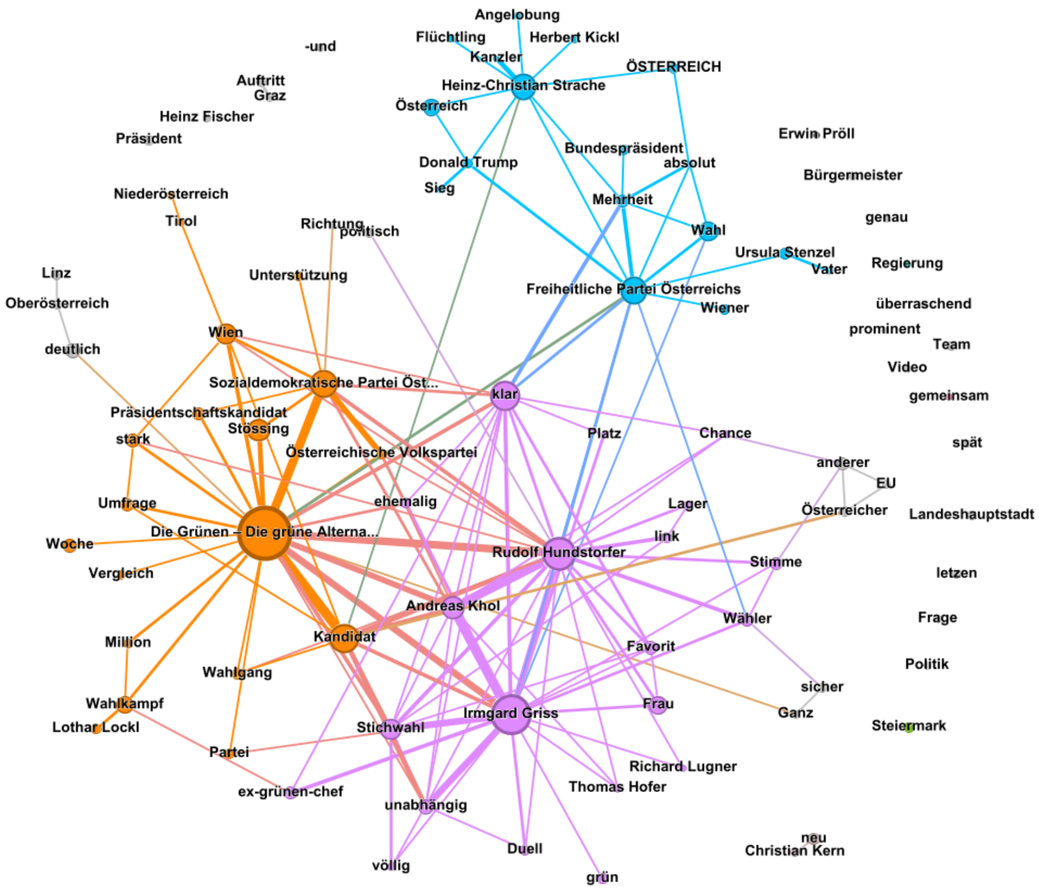
The networks on Norbert Hofer (see figure 6.15 for “Presse” and figure 6.16 for “Österreich”) also illustrate several clusters with some overlap. They both have a big cluster on the FPÖ, while “Presse” has also a separate smaller cluster on the party leader Strache. There are different clusters on the other presidential candidates and parties, similar to Van der Bellen. Interestingly, the Grüne represents only a very small node in the Hofer networks, while the FPÖ was very present in the neighborhood of Van der Bellen, especially in “Österreich.” Both visualizations contain also, more prominently than in the Van der Bellen networks, a refugee and Donald Trump node.

An examination of the mentioned words also yields interesting differences. We find that Alexander Van der Bellen networks are associated with the word old while Norbert Hofer with young and nice in the “Presse.” Ultimately, we find in reporting on Hofer and Van der Bellen, the FPÖ party is prominently discussed while the Grüne only in Van der Bellen’s networks. “Österreich” emphasizes the FPÖ party and its members more than “Presse.”



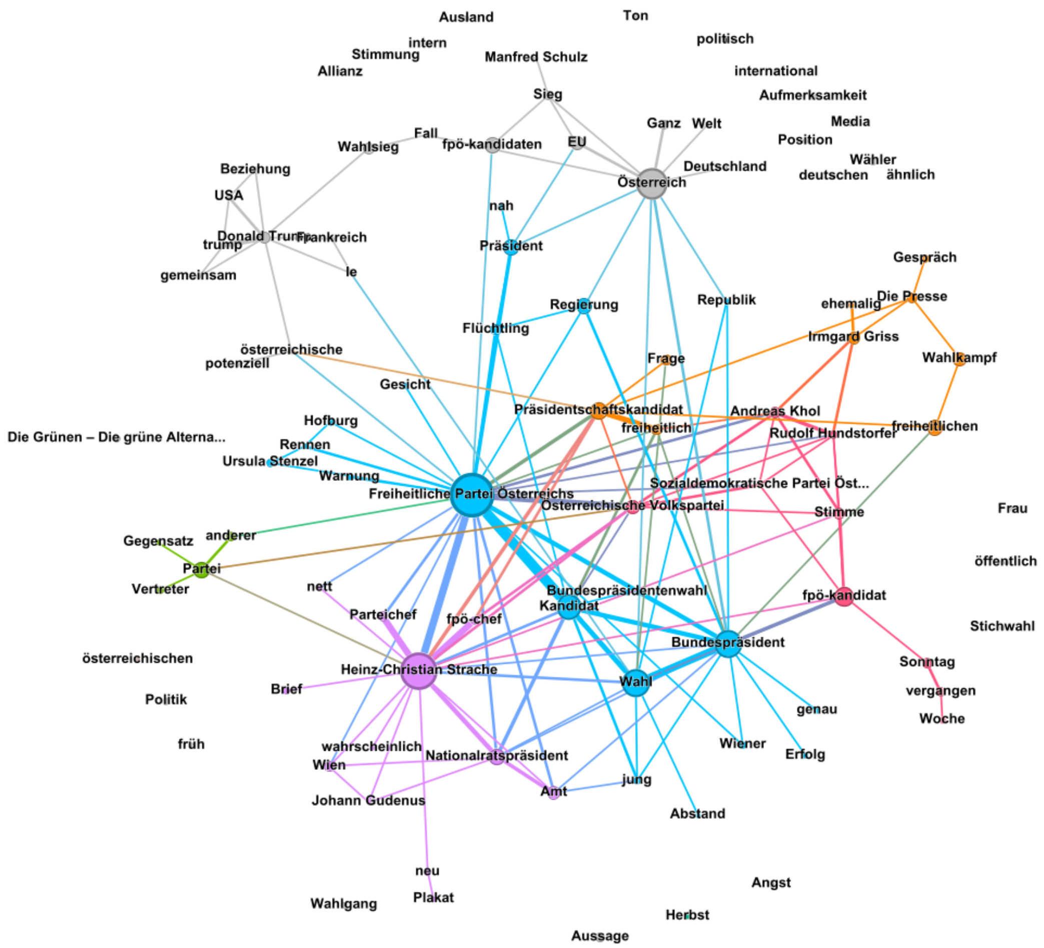
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	97	176	10.948	2.714	0.286	24
Full	775	4104	14.963	1.986	0.172	1

Table 6.13: A filtered ego co-occurrence network of Alexander Van der Bellen based on Presse articles during the election.



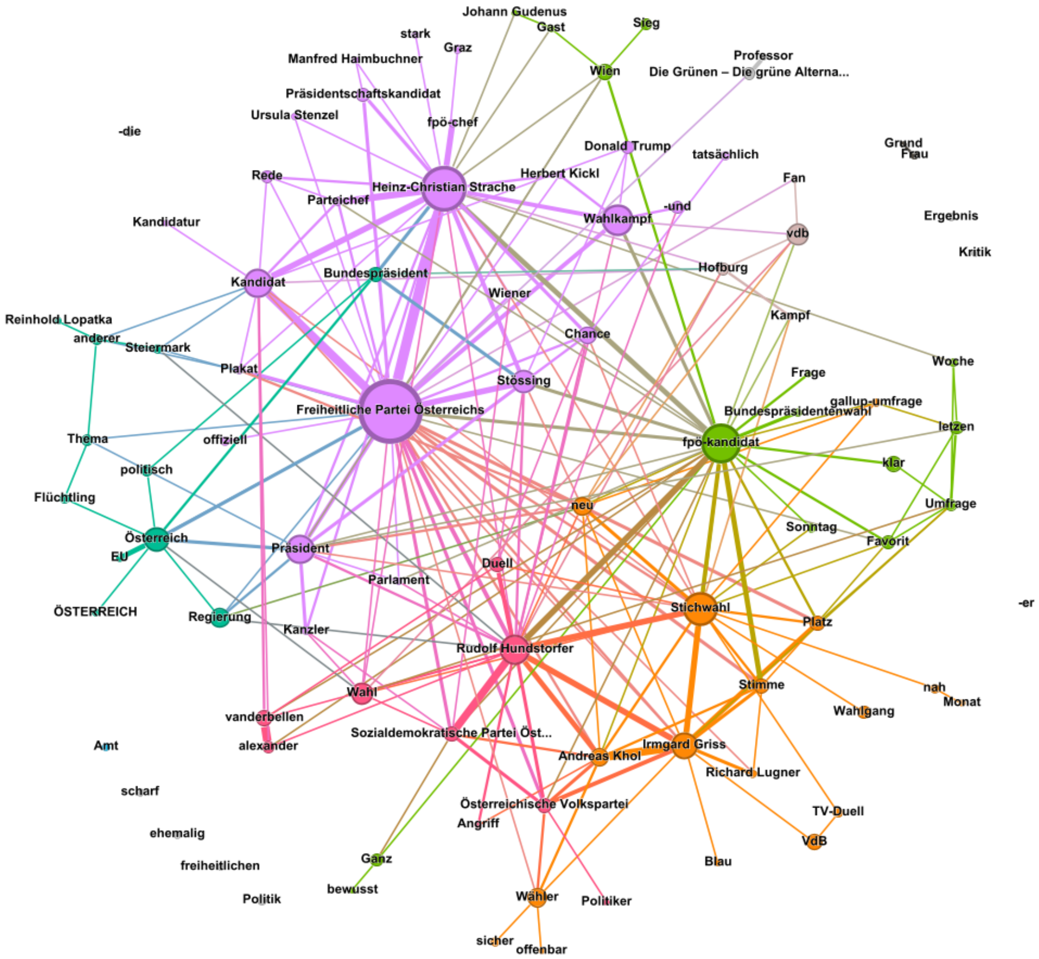
Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	89	152	9.933	2.872	0.375	21
Full	531	2693	16.618	1.981	0.121	1

Table 6.14: A filtered ego co-occurrence network of Alexander Van der Bellen based on Österreich articles during the election.



Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	97	147	8.082	3.152	0.441	25
Full	851	4474	14.482	1.988	0.185	1

Table 6.15: A filtered ego co-occurrence network of Norbert Hofer based on Presse articles during the election.



Type	#Nodes	#Edges	Avg. Degree	Avg. Path	Modularity	Components
Filtered	94	232	14.489	2.598	0.323	11
Full	488	2853	21.148	1.976	0.078	1

Table 6.16: A filtered ego co-occurrence ego network of Norbert Hofer based on Österreich articles during the election.

6.2 Ranking named entities

This section explores a measure of relevance of named entities, such as politicians, in texts. The articles used to construct networks are classified as focusing on politics and encompass the timespan of the Austrian presidential election from 1^{st} of September 2015 to the 30^{th} of December 2016. These are the same articles examined through the descriptive statistics in section 5.3.1, which we also reference multiple times in the interpretation of the results. The nodes of the networks consist only of named entities classified as either a person, organization, or location. The extracted relations between them are co-occurrences in a sentence. The networks are thereby meant to provide insights into how political networks and their influence are portrayed in media. We presume that co-occurrence is indicative of some important represented relation between actors. We expect that through more fine-grained semantic analysis, more expressive and precise networks could be constructed for such analysis. As discussed in section 3.2.7, the models and data sources available to us when the NLP pipeline was finalized did not allow for more sophisticated semantic relation extraction.

Our approach to identifying relevant entities in the extracted article networks is inspired by Textrank [MT04]. As discussed in section 3.3.2, this involves applying the Pagerank algorithm to score and extract important, central keywords from texts. Textrank is an established approach in NLP research for text summarization. It is an unsupervised method and does not require specific training data, making it language-agnostic and applicable to our German-language data. As noted by media scholar Tarleton Gillespie [Gil14], relevance is a "fluid and loaded judgement" for which "no independent metric" exists. Thus, whether a conception of relevance is accepted often depends on the audience. We thereby do not claim or strive to derive a supposed objective measure for relevance but instead hope this approach can provide another limited perspective to debates on portrayal of relevance of political actors in media reporting. In computer science, taking an epistemological stance that questions objectivity is not the norm [DK20], but recent research efforts indicate a shift in sentiment toward acknowledging interpretative pluralism. For example, disagreement in data labeling is increasingly not seen simply as an error but indicative of different perspectives on a data point that need to be considered in modelling [MPY22].

We use Textrank as a measure of relevance during the election because we understand it as approximating the embeddedness of political actors in networks. In contrast to degree centrality, it is not calculated only considering relations to neighbors but, instead, how connected a node is to other well-connected nodes. In the original Textrank paper [MT04], they describe it as nodes recommending other nodes. In the following sections, we provide various rankings and line graphs to illustrate how Textrank can be used to study relevance in news content. We interpret the resulting descriptive statistics to illustrate their analytical utility to readers. We find significant discrepancies when comparing relevance based on Textrank with measures based on the number of mentions of an entity. Future research could explore the significance of differences between these measures in more detail to better explain how they emerge. We did not evaluate whether different people would agree with the presented relevance assessments. Thus, such an investigation could be an interesting follow-up study. Since news reporting is a highly contested domain (see section 1) with various ascriptions of bias being raised against different outlets, various perspectives on measuring relevance would likely emerge from such a study.

6.2.1 Consistency in ranking

We tried to identify a set of named entities within the networks with consistently high relevance, according to Textrank. This inquiry was inspired by prior work [DCT12] on identifying key actors in reporting on Sudan. We constructed co-occurrence networks for every week in the studied timeframe. Then, we extracted a set of the named entities that are consistently part of the top 100 most relevant nodes according to Textrank in all weeks. Next, we rank the named entities according to the sum of all Textrank scores multiplied by the number of nodes in the respective weekly networks.

The table 6.17 illustrates the results of this ranking process. It shows that the two major Austrian parties, ÖVP and SPÖ, are consistently relevant over the whole timespan. This is an intuitive result due to their present and historical influence on both state and national levels. The EU is consistently relevant in the high quality newspapers "Presse" and "Standard," possibly illustrating that these papers are more concerned with international politics. Interestingly, no single politician was consistently relevant within our timeframe. Instead, major parties, countries, international organizations, and big Austrian cities such as Wien and Salzburg were relevant every week. The FPÖ, which was the party of Norbert Hofer, was consistently relevant in the tabloid papers "Österreich" and "Kurier," but also in the high quality paper "Standard." The Grüne, the former party of Van der Bellen, was consistently relevant in all papers besides the tabloid paper "Heute."

	Heute	Krone	Kurier	Österreich	Presse	Standard
1	SPÖ	SPÖ	SPÖ	SPÖ	SPÖ	SPÖ
2	ÖVP	ÖVP	ÖVP	ÖVP	ÖVP	ÖVP
3		Wien	Wien	Wien	Wien	Wien
4		Grüne	Grüne	Grüne	Grüne	Grüne
5		Salzburg	FPÖ	FPÖ	EU	FPÖ
6			Deutschland			EU
7						Deutschland

Table 6.17: Named entities consistently ranked in top 100 over all weeks.

We also constructed a ranking of consistent relevance that features presidential candidates. The resulting table 6.18 shows which named entities were consistently top 100 ranked in 70 % of all weeks. The table only lists the top 15 results since they include all ranked presidential candidates. The displayed results show the same entities as in the table 6.17.

Only the newspapers "Kurier" and "Österreich" contain the two final presidential election contenders. This indicates that they put the most significant consistent focus on the presidential election. Norbert Hofer was, in both instances, calculated as being more prominent.

	Heute	Krone	Kurier	Österreich	Presse	Standard
1	SPÖ	SPÖ	SPÖ	SPÖ	SPÖ	SPÖ
2	ÖVP	ÖVP	ÖVP	ÖVP	ÖVP	ÖVP
3	Wien	Wien	Wien	Wien	Wien	Wien
4	Grüne	Grüne	Grüne	Grüne	Grüne	Grüne
5	FPÖ	FPÖ	FPÖ	FPÖ	FPÖ	FPÖ
6	Oberösterreich	EU	EU	EU	EU	EU
7	NÖ	Deutschland	D. Trump	D. Trump	D. Trump	D. Trump
8	E. Pröll	A. Merkel	Deutschland	Deutschland	Deutschland	Deutschland
9		Europa	N. Hofer	N. Hofer	USA	USA
10		S. Kurz	USA	A. Van der Bellen	Türkei	Türkei
11		Salzburg	Türkei	A. Merkel	A. Merkel	A. Merkel
12		Oberösterreich	A. Van der Bellen	W. Faymann	Syrien	Syrien
13		NÖ	C. Kern	S. Kurz	Europa	Europa
14		Kärnten	A. Merkel	R. Mitterlehner	S. Kurz	S. Kurz
15		Burgenland	Europa	H.C. Strache	R. Mitterlehner	R. Mitterlehner

Table 6.18: Top 15 named entities consistently ranked in top 100 in 70 % of weeks.

6.2.2 Ranking Austrian political actors

In this section, we rank presidential candidates according to relevance in co-occurrence networks constructed from all articles published in the studied timeframe. Table 6.19 shows the candidates' ranking and corresponding Textrank score.

	Heute	%	Krone	%	Kurier	%
1	Norbert Hofer	0.73	Norbert Hofer	0.22	Norbert Hofer	0.39
2	A. Van der Bellen	0.72	A. Van der Bellen	0.18	A. Van der Bellen	0.37
3	Rudolf Hundstorfer	0.24	Rudolf Hundstorfer	0.07	Rudolf Hundstorfer	0.13
4	Andreas Khol	0.22	Andreas Khol	0.05	Andreas Khol	0.13
5	Irmgard Griss	0.15	Irmgard Griss	0.03	Irmgard Griss	0.1
6	Richard Lugner	0.05	Richard Lugner	0.02	Richard Lugner	0.02
	Österreich	**%**	**Presse**	**%**	**Standard**	**%**
1	Norbert Hofer	1.08	Norbert Hofer	0.39	Norbert Hofer	0.49
2	A. Van der Bellen	0.79	A. Van der Bellen	0.38	A. Van der Bellen	0.44
3	Rudolf Hundstorfer	0.29	Rudolf Hundstorfer	0.15	Rudolf Hundstorfer	0.13
4	Andreas Khol	0.2	Andreas Khol	0.13	Andreas Khol	0.11
5	Irmgard Griss	0.19	Irmgard Griss	0.1	Irmgard Griss	0.11
6	Richard Lugner	0.07	Richard Lugner	0.02	Richard Lugner	0.03

Table 6.19: Ranking of presidential candidates over the whole timeframe.

The presidential candidate Norbert Hofer is the most prominent in tableTable 6.19, but most ranking scores are close to Alexander Van der Bellen, with the difference ranging from 0.01 to 0.05 %. The only exception and the most significant discrepancy is the “Österreich” newspaper, where Hofer’s score is ahead by 0.29 %. This result may be explained by the ego network of Hofer shown in section 6.1.3. It indicates that “Österreich” frequently mentioned the candidate in relation to the FPÖ party and its members. This strong network around the FPÖ is likely the reason for this high Textrank score. In contrast, in terms of total mentions within the timespan, Alexander Van der Bellen is ahead. Richard Lugner was overall not very prominent, but most relevant in the tabloid papers like “Heute” and “Österreich.” The candidates of the SPÖ and ÖVP have similar scores, while the SPÖ candidate is slightly ahead. Irmgrad Griss is consistently above Lugner and below the other candidates.

The graph 6.1 visualizes a ranking of presidential candidates based on Textrank where each data point was determined based on a co-occurrence network representing a month of reporting. Norbert Hofer is slightly ahead of Alexander Van der Bellen most of the time in all newspapers. This aligns with the results from table 6.19. In contrast, in the analysis based on mentions discussed in section 5.3.1, Alexander Van der Bellen is ahead. This discrepancy indicates that Van der Bellen may be mentioned more often, but Hofer is more connected with prominent entities. The graph also indicates that in “Presse” and “Kurier,” Andreas Kohl was featured more centrally in April 2016 than in other papers.

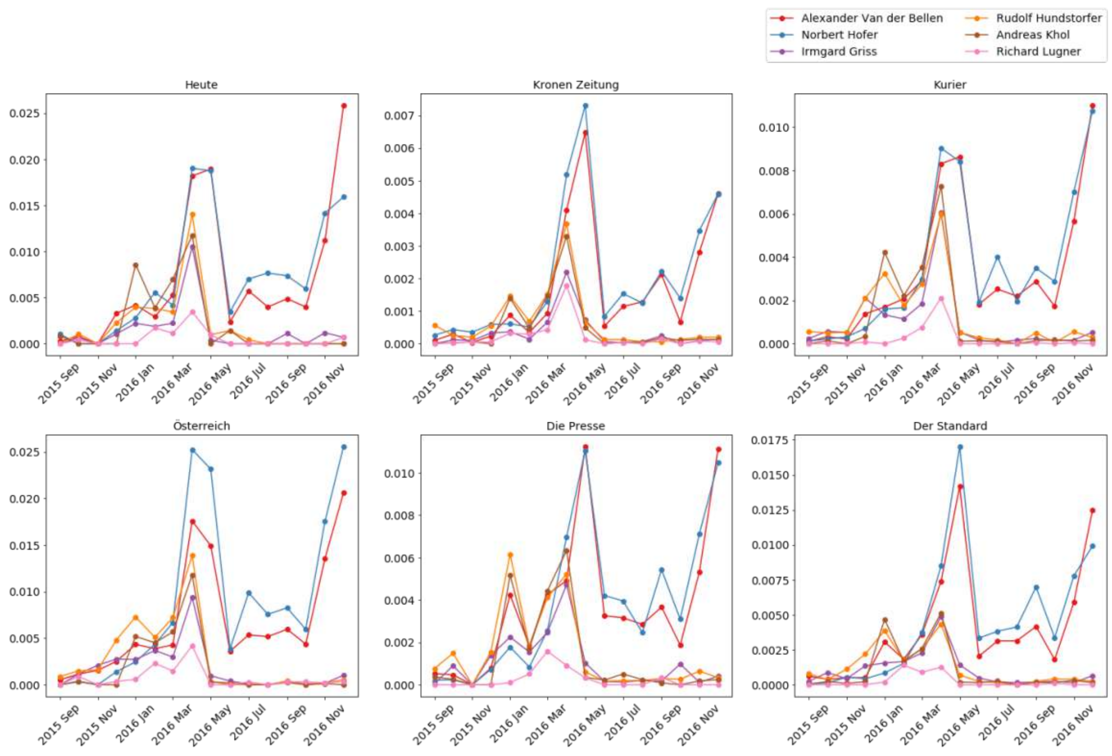


Figure 6.1: Ranking presidential candidates across time with Textrank.

We use Textrank to compare monthly relevance of political parties over time. The total score for each party is based on the sum of the Textrank scores of all identified members and the party entity. The presidential candidates were not counted as members in this analysis. The graph 6.2 visualizes the results. We identify several differences to the results presented in figure 5.10 based on counting mentions of parties and their members. For example, in the Textrank-based graph, the ÖVP is ahead of the SPÖ in more months. The starkest example is “Krone” in the mentions-based analysis, where the SPÖ is only behind the ÖVP in December. In the Textrank-based one, this is reversed, and the SPÖ is only ahead in May. In terms of mentions, the Grüne is ahead of the FPÖ in some months in “Presse” and “Standard.” In the graphs based on Textrank, the FPÖ is consistently ahead. The Textrank-based results for “Krone” and “Kurier” also significantly increase the lead of the FPÖ. These results indicate that the FPÖ and ÖVP, i.e., the parties and their members, are more prominent and often mentioned with other prominent entities. We interpret this also as them possibly being portrayed as more embedded in relevant Austrian political networks.

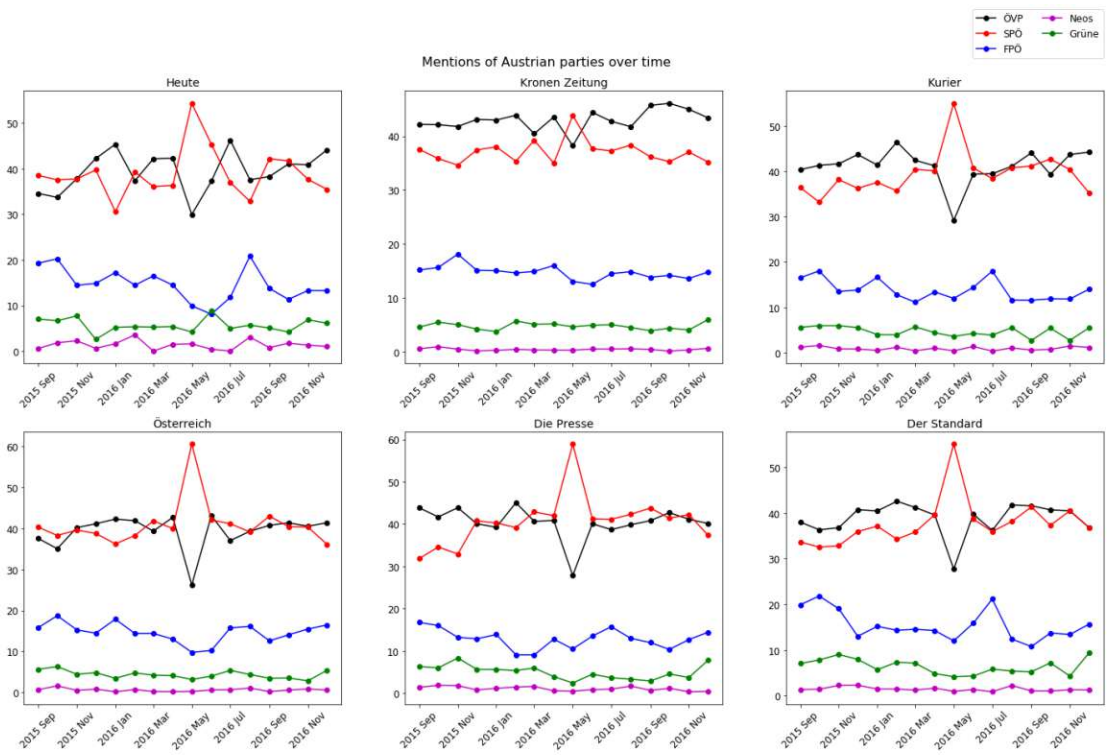


Figure 6.2: Ranking political parties across time with Textrank.

CHAPTER 7

Conclusion and future work

In this thesis, we presented a co-occurrence network-based approach to news analysis and illustrated its viability through an empirical examination of coverage of the 2016 Austrian presidential election. We reviewed the literature on news analytics and text network analysis. Then, we discussed our methodological approach and provided an evaluation of our named entity detection and linking scheme. Various descriptive statistics were presented to give insights into publishing patterns of the newspapers during the election, such as how much space in issues is given to different topics, what sentiments and emotions are predominant in what papers, and the writing quality of the papers according to metrics. We also quantified how much the presidential candidates and the political parties were covered and identified gender bias in reporting disadvantaging female Austrian politicians. Finally, we presented our network analysis, highlighting differences between the newspapers, such as what themes were covered during the election, how presidential candidates were portrayed, and explored which candidates and other named entities were most relevant in the articles. Several key ideas of the thesis were presented at the 2017 Vienna Young Scientists Symposium [GJH17]. The overall aim of this work was to address the following overarching research questions:

(1) To what extent can we make sense of the quality of extracted named entities and relationships used to construct co-occurrence networks?

(2) Which methods are appropriate to detect relevant entities and topics in the extracted networks?

(3) What can a co-occurrence network-based approach reveal about reporting on the 2016 Austrian presidential election?

(4) What are the benefits and risks of a co-occurrence network-based approach to news and text analytics?

In the remainder of this section, we provide partial answers to these questions and discuss possibilities for future work.

The first research question was concerned with the quality of the node and relationship extraction. In the creation of this thesis, we spent a lot of time on cleaning and preparing data to improve the extraction of entities and relationships. Initial efforts without any preprocessing resulted in networks that were very difficult to interpret. We encountered various issues through testing, such as many synonyms making it harder to see relations between concepts and entities, redundant articles inflating certain relations, different ways meta-information about the articles were included, and a lot of articles with information we deemed unhelpful for our analysis like listings of TV programs.

In section 3.2, we describe our NLP pipeline, which includes tokenization, part of speech tagging, named entity recognition and linking, sentence segmentation, and lemmatization. Since we focused on articles during the 2016 Austrian presidential election, central actors needed to be recognized to represent them as network nodes. We evaluated different named entity detection schemes on a test set of articles (see section 4). We found that combining an available open-source model with a dictionary-based approach crafted for the Austrian context worked best because it had comparable precision and high recall with entities necessary for the election analysis. We also had to create stopword lists to exclude certain keywords specific to the newspapers and corpus, such as abbreviations like "Wh." and "AUT." The available open-source tools for sentence segmentation work very well, so we did not conduct a separate evaluation of them. We tried to extract semantic relations through dependency parsing and sentiment analysis but found that the tools did not perform well enough in our tests. In section 3.2.7, we discuss several other NLP techniques and tools we tried to use to construct more expressive networks, like co-reference resolution, which would improve recall but discarded after testing.

Ultimately, we found that entity and relation extraction on the corpus required lots of "data work, " including understanding the data, cleaning it, and constructing data for testing approaches and training models. Trying to improve quality in network construction is an ever-moving target, as there is always room for improvement when dealing with great amounts of data, and many different conceptions of quality could be conceived. At some point, we had to stop refining our approach and focus on pragmatically improving the steps in the NLP pipeline according to the state-of-the-art determined during our initial review. We hope the provided evaluation and the explanations of issues we encountered add to debates on challenges around quality in named entity and relationship extraction.

The second and third research questions were concerned with what methods can be used to extract relevant named entities and topics and what their application reveals about election coverage. We unpacked the concept of relevance in section 6.2 and argued that it is contested and qualitative and that the acceptance of certain notions depends on the audience. Thus, our efforts to detect relevant entities and topics are not meant to settle this issue, which we argue is not possible, but instead, provide new perspectives on relevance in news articles through a network-based approach. We used Pagerank to identify relevant named entities and the Louvain community detection method to uncover themes and topics in the networks representing reporting on the presidential election. After our review, we decided to use these methods because they are established and tested, and lots of research builds upon them (see section 2.3).

We found that our approach reveals differences in reporting compared to the results based on counting mentions. The networks presented in section 6.1 illustrate topics we identified in networks representing different newspapers. Some identified differences include, for example, how some newspapers seemingly emphasized thresholds in relation to refugees more prominently. In our ego network analysis on coverage of the two final candidates in "Presse" and "Österreich", we found that the FPÖ is prominently mentioned in relation to both Van der Bellen and Hofer, whereas the Grüne only with Van der Bellen. The FPÖ was more prominently mentioned in "Österreich." We also noted differences between the outlets based on what entities were more relevant according to Pagerank, as described in section 6.2. For example, candidate Norbert Hofer was particularly highly ranked in "Österreich" compared to other outlets, which we think is due to the amount of coverage of the FPÖ and its members, who are frequently mention with Hofer. We found that all papers were concerned with similar themes in their election coverage. The differences become more visible when the relations of keywords and their strength are examined more closely.

We discussed problems with both the ranking and community detection. We discussed how the results of Pagerank are difficult to interpret. Furthermore, we highlighted how chosen parameters determine which clusters are identified and pointed to issues like confirmation bias, which could lead to overly generous interpretations. The interpretations of clusters usually require ignoring several nodes and relations to build a coherent argument for a theme. This further questions the reliability of the approach. So, the picture we present is mixed, and further research on evaluating the value of these networks is necessary for more conclusive answers. This may require rethinking current evaluation paradigms to also account for plurality of in ground truth construction. As we discussed, if we evaluate node ranking measures based on how well they reveal relevant nodes, then different humans are needed to construct different standards for relevance. This kind of rethinking of evaluation is challenging in the context of big data, where browsing through a lot of data to be better able to define relevance also takes a lot of time.

The fourth research question was concerned with the benefits and risks of our network-based approach to news analysis. The exploration of texts through networks puts the relations between words, entities, and nodes into focus. This differentiates it from methods concerned with simply counting occurrences and can provide different insights into data. Thus, network-based approaches remain a relevant research topic that challenges more atomized views of data. This is also illustrated by recent efforts to use graph-based representations of text to enhance cutting-edge deep learning techniques [PNPV23]. However, no obvious answer exists to how texts should be represented as networks. As we highlighted in our review in section 2, various ways to represent text as a network have been explored. Ultimately, different values and priorities shape how nodes and their relations are constructed. This includes what kinds of knowledge or (meta)data the networks are enriched with; in our case, we used information about entities on Wikidata and an open-source classification model to derive nodes. In section 6.1, we illustrated how influential different choices in parameterization were for constructing and visualizing networks and calculating measures. We warned that this flexibility might be misused and that conformation bias could distort results when the goal is to construct expressive and useful networks. Such qualitative markers of quality undermine the promises of mechanical objectivity [DG07], but good documentation could aid in still enabling some reproducibility. The AMC is not publicly available, limiting the reproducibility of this project.

We also argued that the flexibility in network construction can be understood as a feature that enables tweaking, exploration, and possibly making it necessary to engage with design interests reflexively. It stands partially in contrast to more automated techniques, which may also consider relations but often present users with seemingly certain results. This can hide uncertainty by foreclosing opportunities to construct or interrogate different results. The issues of interpretability and explainability of such complex models and their results also remain an open research topic. In our approach, important aspects are also dependent on black-boxed models in the NLP pipeline. Thus, ultimately, a compromise between seamful and seamless design has to be made when dealing with big, complex data. We understand the interactive flexibility and interpretability a network-based approach, like ours, can afford as a double-edged sword that can lead to misuse but also provide more agency to researchers. As noted by [FAL+13], "automation of many tasks in news content analysis will not replace the human judgment needed for fine-grained, qualitative forms of analysis." So, recovering agency in analysis is valuable and will remain important as news analysis is an interpretive task that requires humans to form questions and contextualize the results. We also think purely unsupervised approaches are limited in usefulness and must be combined with external knowledge to be more helpful. For example, we used trained models and the Wikidata knowledge base for the named entity extraction, which made it possible to construct more readable networks that could then be analyzed with unsupervised methods. Texts are cultural products embedded in a social context; therefore, externalized knowledge is needed for analysis. Humans can contribute their contextual insights into the domain at different points in the pipeline. For example, humans can provide knowledge through constructing training and testing data, but also when they tweak parameters to explore different interactive visualizations, and can give them meaning. Thus, we think this kind of interactivity paired with supervised approaches that draw on other knowledge are possibly fruitful future research directions. A recent study [ZSLZ22] illustrates, for instance, how the performance of Textrank can be improved by combining it with more complex Word2Vec models that improve the expressiveness of relations between nodes.

Future work could seek to refine and evaluate the web application for ease of use as a news analytics tool for digital humanities scholars and also put the usefulness of interactivity in network generation and exploration into focus. Generally, more qualitative, human-centered research that further explores different perspectives could improve network-based approaches by creating new, plural quality standards. These could center around clarifying and developing better notions of important goals, such as improving the expressiveness and interpretability of extracted networks, entities, and relations, identifying more relevant entities in articles, and extracting better themes. The named entity recognition and co-reference resolution could be improved by constructing appropriate and sizable training sets for the corpus and integrating them with larger, pre-trained models. The availability of this kind of data could also make the discarded semantic relation extraction techniques viable, thereby enabling new interesting forms of analysis and network construction. We think more “data work” that focuses on cleaning the corpus data and making more training and test data available will be required for further improvement. However, this will be difficult as “data work” in machine learning continues to be undervalued and hidden [SKH+21] although, as this work has again highlighted, remains central.

CHAPTER 8

Appendix

8.1 List of named entities of type politician

We present below the named entities we retrieved from Wikidata that represent Austrian politicians.

- **SPÖ**: Franz Schrangl, Konrad Reif, Karl Stix, Fritz Prechtl, Hans Reither, Maria Derflinger, Gertrude Schreiberhuber, Alois Müller, Martina Schröck, Johann Berghammer, Ferdinand Strasser, Susanne Kövari, Annemarie Reitsamer, Ernst Koref, Albrecht Gaiswinkler, Josef Schweiger, Heinrich Ulrich, Julius Linder, Kurt Wimmer, Sepp Rieder, Wolfram Enzfelder, Josef Stoll, Christine Haager, Johann Gmeiner, Thomas Punkenhofer, Engelbert Pernerstorfer, Franz Fruhstorfer, Isidor Wozniczak, Gabriele Titzer, Wolfgang Hager, Sylvia Rinner, Walter Hofer, Anna Rieder, Hans Mayrhofer, Maria Pokorny, Waltraud Schütz, Hans Herke, Martina Malyar, Edwin Schuster, Hermann Lipitsch, Rudolf Strunz, Erhard Koppler, Barbara Blaha, Fritz Marsch, Adalbert Duschek, Manfred Scheuch, Edmund Entacher, Werner Breithuber, Johann Kuba, Franz Pomper, Friedrich Hillegeist, Otto Weber, Gabriele Mörk, Franz Vesely, Karl Frais, Roswitha Bachner, Erwin Preiner, Martin Pappenheim, Herbert Moritz, Ernst Theumer, Walter Scholger, Josef Palme, Hans Jiricek, Kilian Brandstätter, Ernest Brezovszky, Gerhard Pongracz, Karl Gerfried Müller, Max Lotteraner, Robert Lucas, Herbert Bösch, Viktor Schwarzmann, Karl Hager (Politiker), Anita Fleckl, Leopold Spitzer, Valentin Blaschitz, Robert Häuser, Josef Grunner, Hubert Zankl, Josef Heigl, Karl Knapp, Ernst Neuhauser, Maria-Luise Mathiaschitz, Alois Reicht, Hermann Lechner, Josef Guttenbrunner, Hans Unger, Karl Steinocher, Hans Peter Doskozil, Gabriele Votava, Alfred Hintschig, Peter Koits, Adelheid Ebner, Roland Meisl, Bruno Wögerer, Hans Resel, Stefan Steinle, Josef Faustenhammer, Franz Olah, Heidelore Wörndl, Gabriele Schiessling, Paul Weiß, Pius Moosbrugger, Fritz Freyschlag, Heinrich Kandl, Detlef Gruber, Siegfried Herrmann, Wilhelm Eich, Michael Luptowits, Michael Maderthaner, Günther Prutsch, Karl Hohenberg, Eckhard Oberklammer, Alfred Gusenbauer, Maria Leibetseder, Josef Muchitsch, Werner Gruber, Muna Duzdar, Karl Pfeffer, Erna Musik, Manfred Ebner, Erich Schmidt, Robert Janschitz, Johann Kraml, Franz Tratter, Erich Kosler, Michaela Rösler, Hans Schober, Franz Dobusch, Josef Jaksch, Wolfgang Michlmayr, Anna Boschek, Theodor Binna, Franz Pülsl, Ludwig Winkler, Josef Peter, Walter Prior, Reinhold Einwallner, Hans Menzl, Alfons Bernkopf, Wolfgang Dolesch, Erich Prattes, Johann Steirer, Walter Steidl, Helmut Braun, Hans Sallmutter, Josef Schützenberger, Karl Reinhart, Julius Spielmann, Hans Sima, Andreas Korp, Hermann Fister, Franz Mager, Harald Troch, Agnes Prandler, Evelyn Messner, Karl Zlatarits, Leopold Weinhofer, Leopold Horacek, Fridolin Schröpfer, Karl Weigl, Gustav Ehart, Franz Kreuzer, Heinz Hufnagl, Robert Zehentner, Josef Reichl, Adalbert von Springer, Wolfgang Erlitz, Josef Posch, Hermann Findeis, Walter Mondl, Ewald Lindinger, Ernst Steinbach, Evelyn Regner, Andreas Zettel, Franz Schuster, Otto Klupp, Rudolf Schlager, Alois Hofer, Johanna Langanger, Friedrich Strobl, Anneliese Albrecht, Franz Kottek, Franz Horr, Karl Krisch, Josef Kaut, Kordula Schmidt, Michael Sika, Christa Krammer, Dietmar Wedenig, Ewald Ritter, Edmund Holzfeind, Albin Dötsch, Sonja Wehsely, Alexander Böhm (Politiker), Franz Reifmüller, Oskar Helmer, Anton Weber, Johann Glaser, Karl Pospischil, Franz Weber, Hans Krutzler, Ferdinand Brunner, Johann Svitanics, Rudolf Hundstorfer, Alfred Schreiner, Thomas Reindl, Anton Feistl, Josef Schlager, Paul Posch, Alois Stöger, Ernst Winkler, Hermann Fischer, Wolfgang Riedler, Karl Neuwirth, Hans Nießl, Niko Pelinka, Paul Kiss, Peter Köpf, Hans Czettel, Ludwig Tuller, Christian Füller, Ingeborg Bacher, Rudolf Kaminger, André Heller, Brigitte Tegischer, Ernst Schmid, Stefan Plaimauer, Wilhelmine Moik, Otto Pendl, Harald Krassnitzer, Safak Akcay, Rudolf Fertl, Karl Gruber, Norbert Neururer, Josef Heger, Wilhelm Scheibein, Klaus Soukup, Rudolf Scharinger, Tibor Karny, Hans Helm, Bettina Stadlbauer, Angelika Fußenegger, Peter Graf, Leopold Gratz, Manfred Gruber, Johann Müllner, Caspar Einem, Viktor Fadrus, Josef Fohringer, Erich Tschernitz, Doris Kampus, Herwig Seiser, Hannes Fazekas, Helmuth Stocker, Karl Volkert, Otto Binder, Marcus Schober, Franz Fux, Sonja Kato-Mailath-Pokorny, Adolf Reitmaier, Alois Weidenhillinger, Ludwig Deusch, Victor Theodor Slama, Manfred Moser, Gebhard Arbeiter, Herbert Egg, Rosl Moser, Peter Schachner-Blazizek, Max Neugebauer, Franz-Karl Effenberg, Sylvia Kögler, Gabriele Kolar, Günter Kiermaier, Hannelore Reiterer, Helmut Manzenreiter, Josef Gammer, Karl Waldbrunner, Hermann Tancsics, Franz Mikulits, Heidrun Silhavy, Hermine Mospointner, Anton Afritsch (Bundesrat), Ignaz Hinterleithner, Josef Pleyl, Hans Riemer, Alfredo Rosenmaier, Helmut Wöginger, Bruno Pittermann, Viktor Klima, Klaus Uwe Feichtinger, Alfons Schröcker, Josef Weichenberger, Harry Kopietz, Rudolf Nürnberger, Marie Lang, Florian Bergauer, Thomas Wagner, Josef Eidenberger, Michael Kretz, Andreas Rudas, Franz Bednar, Renate Winklbauer, Gustav Scherbaum, Hermann Bruno Wetschnig, Reinhard Resch, Günther Sallaberger, Andrea Kalchbrenner, Fritz Muliar, Karl Miksch, Othmar Schneglberger, Johann

Blümel, Erna Sailer, Helmuth Vogl, Rosa Mayreder, Paula Wallisch, Peter Keppelmüller, Elisabeth Petznek, Hans Wastl, Hermann Hermann, Franz Babanitz, Josef Fridl, Johannes Schwarz, Leopold Thaller, Daniela Holzinger-Vogtenhuber, Alois Roppert, Monika Kaufmann, Anton Koczur, Marie Hautmann, Günther Kräuter, Gabriele Sprickler-Falschlunger, Doris Bures, Irene Szep, Anton Wodica, Hans Brachmann, Erich Beck, Günther Albel, Tarik Mete, Quirin Kokrda, Edith Dobesberger, Felix Butschek, Josef Witternigg, Kurt Zach, Josef Deutsch, Karl Schmiedbauer, Arnulf Häfele, Rupert Zechtl, Brigitte Wohlmuth, Anton Weiguny, Koloman Markart, Franz Popp, Anton Afritsch, Josef Pfeifer, Richard Leutner, Jacques Hannak, Johann Hartmann, Paul Truppe, Gerald Mader, Franz Koci, Friedrich Verzetnitsch, Otto Libal, Arnold Schenner, Petra Müllner, Hans Karl Schaller, Koloman Wallisch, Josef Diem, Emil Nesler, Sepp Hopfer, Hermann Stecher, Theodor Kery, Gerda Weichsler-Hauer, Leopold Wolf, Adolf Duda, Stefan Kettner, Heinz Kraupner, Franz Walcher, Hans Kouba, Raimund Sassik, Laurenz Widholz, Walter Thaler, Franz Rauscher, Anton Schlinger, Felix Stika, Matthias Herrmann, Ingrid Schubert, Viktor Stein, Wolfgang Radlegger, Josef Wieser, Alfred Stingl, Gerhard Kubik, Sonja Ablinger, Ewald Wagner, Johann Pölzer senior, Ewald Gossy, Erwin Kaipel, Hermann Leithenmayr, Ferdinand Freund, Ruth Becher, Alfred Porges, Ferdinand Doblhammer, Mario Trinkl, Berthold Fuchs, Maria Jacobi, Wilhelm Steingötter, Dominik Löw, Norbert Horvatek, Erika Krenn, Alois Grath, Karin Scheele, Karl Grünner, Anton Azwanger, Karl Kislinger, Harald Ogris, Ines Obex-Mischitz, Josef Rohata, Anton Berger, Christian Meidlinger, Karl Sekanina, Norbert Tmej, Andreas Haitzer, Johann Sipötz, Hans-Joachim Ressel, Gottlieb Unger, Maria Köstler, Kurt Stepancik, Anna Demuth, Karl Konrath, Franz Vranitzky, Hans Ludwig, Peter Florianschütz, Ferdinand Chaloupek, Claudia Laschan, Franz Slawik, Johann Schmölz, Philip Kucher, Walter Guggenberger, Karl Schlögl, Vinzenz Übeleis, Alfred Migsch, Margaretha Obenaus, Josef Voithofer, Georg Hubmann, Josef Gruber, Erwin Spindelberger, Ludo Moritz Hartmann, Peter Pelinka, Johann Oberhammer, Heinz Nittel, Elisabeth Vitouch, Karl Stoiser, Marie Beutlmayr, Erwin Lanc, Josef Cap, Alfred Magaziner, Hans Barwitzius, Georg Oberhaidinger, Elfriede Strobel, Josef Jahrmann, Rudolf Plessl, Anton Linder, Christian Oxonitsch, Julia Röper-Kelmayr, Josef Taucher, Elfriede Karl, Kurt Heindl, Wolfgang Böhmer, Josef Pichler, Dieter Fuith, Günter Rehak, Wilhelm Schiegl, Ernst Ritter, Sonja Hammerschmid, Otto Bauer, Hubert Pfoch, Veronika Floigl, Johanna Dohnal, Andreas Babler, Franz Steininger, Josef Baliko, Franz Kielwein, Theodor Pauppill, Josef Kreiner, Rudolf Huber, Karl Honay, Pavel Bizjak, Erwin Scharf, Victor Band, Ignaz Till, Franz Glaserer, Franz Willinger, Leopold Stípčak, Franz Glockner, Albert Hummel, Hermann Heßl, Hanna Sturm, Klaudia Friedl, Franz Siegel, Franz Harringer, Günther Sidl, Anton Schrammel, Vinzenz Knor, Hermann Krenn, Franz Rosenberger, Hans-Karl Uhl, Werner Faymann, Gerhard Zechner, Olga Pircher, Ingobert Mayr, Wilhelm Holper, Michael Schickhofer, Ilse Barea-Kulcsar, Kurt Preiß, Franziska Appel, Anita Strebl, Manfred Wegscheider, Rudolf Exler, Anton Heinzl, Heinz Grabner, Karl Seitz, Gerhard Köfer, Kurt

Nekula, Josef Fuchs, Johann Höll, Hans Lala, Andreas Sucher, Hanna Hager, Hilde Krones, Markus Vogl, Christine Schirmer, Rudolf Pöder, Rudolf Ceeh, Helmut Weinberger, Karl Würbel, Edith Paischer, Rudolf Scholten, Ernst Winder, Karl Horejs, Franz Brand, Christoph Klauser, Franz Gruber, Hannes Derfler, Werner Brenner, Josef Lenzi, Harald Seidl, Franz Illig, Josef Brandauer, Rosalie Zull, Oskar Janicki, Josef Wimmer, Otto Gerhartl, Richard Wolf, Erich Pilsner, Ludwig Wutschel, Andreas Stampler, Eduard Euller, Harald Weiss, Rudolf Singer, Adelheid Popp, Heinz Lehner, Leopold Ruckteschl, Leopold Weber, Fritz Konir, Josef Holaubek, Werner Kummerer, Ferdinand Grandits, Maria Matzner, Roman Felleis, Hans Jungwirth, Wilhelm Sieß, Kurt Maier, Walter Bacher, Gertraud Jahn, Ella Zipser, Ferdinand Gstöttner, Peter Jankowitsch, Ulrike Sima, Laurenz Genner, Renate Csörgits, Rudolf Thalhammer, Peter Marizzi, Josef Ortner, Wolfgang Knes, Eduard Lindner, Silvia Rubik, Michael Pinter, Gerhard Schmid (SPÖ), Harald Repar, Helmut Dietachmayr, Hermine Kubanek, Michael Nagele, Josef Gabriel, Fritz Rudigier, Franz Hartl, Franz Voves, Siegfried Tromaier, Edmund Aigner, Hans Muchitsch, Alfred Kollmann, Karl Petinger, Johann Zwanzger, Maximilian Unterrainer, Wolfgang Moitzi, Waltraut Hladny, Leopold Bieder, Andreas Matthä, Franz Hillinger, Ferdinand Reisinger, Johann Mandl, Leopold Zechner, Josef Moser, Rudolf Kolb, Erich Fenninger, Inge Jäger, Hella Hanzlik, Karlheinz Hora, Arnold Auer, Gustav Pomaroli, Ernst Holzmann, Friederike Seidl, Friedrich Lehr, Kai Jan Krainer, Adele Obermayr, Karl Kysela, Gerhard Buchleitner, Viktor Kleiner, Stefan Wölfer, Hans Böck, Theodor Grill, Barbara Prammer, Hans Piesch, Martin Apeltauer, Susanne Bluma, Otto Liedl, Roman Rautner, Hans Hammerstorfer, Julia Herr, Gabriele Arenberger, Pauline Hautz, Petr Baxant, Erika Jirkovsky, Sepp Eberhard, Ernst Hoffenreich, Eduard März, Heinrich Keller (Politiker), Richard Stockinger, Franz Käfer, Hans Handl, Karl Flöttl, Franz Mrkvicka, Barbara Taufar, Renate Egger, Käthe Leichter, Irene Crepaz, Josef Seidl, Anton Scheiblin, Franz Schulz, Gaby Schaunig, Hans König, Christoph Matznetter, Leopold Wally, Helene Postranecky, Ernst Winter (Politiker, 1958), Hedwig Wechner, Rosa Jochmann, Vinzenz Müller, Otto Leichter, Norbert Scheed, Wilhelm Kainrath, Torsten Engelage, Vera Lischka, Heinrich Übleis, Alois Rechberger, Paul Blau, Rudolf Parnigoni, Wolfgang Schimböck, Hans Pusch, Franz Nekula, Maria Ducia, Margit Pfatschbacher, Elmar Steurer, Ferdinand Wultsch, Peter Juznic, Adelheid Praher, Anton Schäfer, Harald Lettner, Hubert Bauer, Fritz Sulzbacher, Karl Swoboda, Karl Schmidlechner, Georg Niedermühlbichler, Günther Kaltenbacher, Kurt Stürzenbecher, Hermann Schulz, Hannelore Buder, Jürgen Czernohorszky, Franz Ockermüller, Adolf Laser, Josef Reschen, Hannes Bauer, Johann Schweigkofler, Michael Unterguggenberger, Marina Hanke, Werner Posch, Hans Lenz, Johann Hechtl, Josef Czerwenka, Herbert Koller, Alois Kaltenbrunner, Hans Gumplmayer, Josef Mohnl, Franz Kirchgatterer, Hannes Weninger, Ernst Längauer, Christa Kranzl, Eduard Speck, Barbara Teiber, Hans Pawlik, Peter Spannring, Hans Frenzel, Emil Kuntner, Wilhelm Wache, Franz Schnabl, Helmut Seel, Karl Fürstenhofer, Monika Kemperle, Ernst Eugen Veselsky, Heinrich Allina, Hertha Firnberg, Karl Richter, Helmut Schamberger, Anja Hagenauer, Waltraud Rohrer, Otto Tschadek, Eduard Schönfeld, Maria Fischer, Georg Hahn, Christian Kern, Roswitha Bauer, Kurt Gartlehner, Melitta Trunk, Therese Schlesinger, Albin Schober, Peter Stauber, Karl Gföller, Michael Häupl, Cornelia Schweiner, Fritz Leitner, Harald Kunstätter, Konrad Nimetz, Franz Großmann, Johann Pettenauer, Josef Hammerschmid, Marianne Pollak, Johann Zauner, Paul Johannes Schlesinger, Franz Binder, Klaus Köchl, Arnold Eisler, Ewald Persch, Johann Holztrattner, Josef Rauchenberger, Otto Sagmeister, Berthold Roithner, Christian Faul, Christian Hursky, Claudia Schmied, Lona Murowatz, Sophie Bauer, Karin Kadenbach, Helene Auer, Elisabeth Dittrich, Wolfgang Petritsch, Jasmine Chansri, Johann Köteles, Franz Asboth, Ludwig August Bretschneider, Hans Muzik, Alois Pisnik, Gisela Peutlberger-Naderer, Johann Smitka, Siegfried Kokail, Heinz Illigen, Alfred Gisel, Hedwig Petrides, Sandra Frauenberger, Erich Sulzer, Bernd Stöhrmann, Friedrich Adler, Franz Staffa, Rupert Gmoser, Stefan Hofer, Josef Weidenholzer, Renate Gruber, Anton Rupp, Günter Willegger, Gerald Schmid, Manfred Lackner, Robert Sigl (Politiker), Alfred Ströer, Friedrich Austerlitz, Georg Stangl, Ferdinanda Flossmann, Werner Feurer, Ludwig Kostroun, Oliver Wieser, Tanja Wehsely, Reinhart Rohr, Karl Appel, Franz Krammer, Gerhard Reheis, Alois Zanaschka, Johann Neubauer, Anton Woschitz, Franz Skotton, Theodor Körner, Ernst Pfleger, Otto Felix Kanitz, Hedda Kainz, Josef Wiesmayr, Josef Ninaus, Georg Pehm, Josef Franzmair, Franz Gruener, Franziska Fast, Fritz Mayer, Doris Hahn, Gerhard Zatlokal, Kurt Mühlbacher, Erwin Niederwieser, Rupert Hartl, Rudolf Schober, Peter Schieder, Karl Leuthner, Jörg Leichtfried, Josef Modl, Hans Rosenberger, Helmut Bachmann, Hilde Eisl, Willi Mernyi, Manfred Wurm, Sonja Steßl, Paul Rosenberger, Ferdinand Lacina, Robert Uhlir, Alois Jirovetz, Kurt Flecker, Adolf Aigner, Josef Schlömicher-Thier, Fritz Stadler, Andrä Idl, Reinhard Machold, Volkmar Harwanegg, Rudolf Tonn, Inge Zankl, Josef Wondrak, Rudolf Kaske, Ludwig Oberzaucher, Julius Deutsch, Hans-Peter Schlagholz, Josef Prantl, Willi Grafeneder, Erich Kessler, Josef Veleta, Andreas Mailath-Pokorny, Michael Ehmann, Josef Broukal, Franz Pichler, Karl Traxler, Leo Lukas, Erich Rippl, Hilde Pleyer, Martina Ludwig-Faymann, Ursula Puchebner, Josef Enslein, Josef Kostelecky, Josef Horak, Josef Schmidl, Hartmut Prasch, Gertrude Fröhlich-Sandner, Rudolf Marchner, Rupert Dworak, Emmy Freundlich, Alfred Fister, Rudolf Häuser, Isabella Kossina, Alois Lang, Edmund Hamber, Herbert Haas, Ivan Wurglics, Elisabeth Pittermann, Robert Elmecker, Manfred Bauer, Georg Puhm, Rudolf Forsthuber, Anton Falle, Harry Rudolf Buchmayr, Peter Kaiser, Alois Hammer, Alexander Hareter, Hubert Kuzdas, Georg Sailer, Sonja Hartl, Jakob Viehauser, Klaus Firlei, Alois Buttinger, Leopold Wedel, Bruno Kreisky, Michael Ludwig, Elisabeth Hakel, Katharina Schinner, Hannelore Reischl, Konrad Fous, Josef Schweighofer, Walter Resch, Walter Geppert, Sieghard Hasler, Karl Drochter, Oscar Pollak, Rudolf Sigmund, Angela Lueger, Alois Gossi, Otto Mödlagl, Josef Staribacher, Ottilie Matysek, Walter Kröpfl, Josef Eksl, Hermann Krist, Johann Kaps, Johann Pölzer junior, Michael Dannereder, Renate Kaufmann, Franz Winterer, Karl Falschlunger, Karin Achatz, Eduard Baumgartner, Eduard Keusch, Ingrid Leodolter, Bettina

Vollath, Siegfried Schrittwieser, Robert Strobl (Politiker), Otto Schweda, Ilse Fetik, Georg Kriz, Rudolf Appel, Walter Posch, Hermine Kraler, Winfried Seidinger, Karl Troll, Willi Stiowicek, Raissa Adler, Alexander Kulman, Erwin Schramm, Benedikt Kautsky, Josef Hirscher, Barbara Novak, Leonhard Treichl, Michael Scherz, Erich Moser, August Neutzler, Johann Böhm, Rosa Weber, Simon Klug, Alois Bauer, Gabrielle Traxler, Alfred Aichinger, Eduard Gargitter, Rudolf Tirnthal, Kurt Steyrer, Christoph Peschek, Christine Hies, Hans Mayr, Gregor Stögner, Hermann Schnell, Martha Tausk, Hans Müllner, Albert Hofbauer, Katharina Graf, Doris Prohaska, Franz Probst, Rudolf Streicher, Leo Geiger, Brunhilde Plank, Franz Zeller, Kajetan Weiser, Adelheid Hirschbichler, Christian Dickinger, Alfred Horn, Josef Hafner, Matthias Konrad, Sybille Straubinger, Erich Suchanek, Vinzenz Dresl, Heinz Gradwohl, Karoline Graswander-Hainz, Christian Drobits, Franz Fiedler, Kurt Maczek, Josef Brunauer, Elmar Mayer, Josef Matejcek, Elisabeth Reich, Elisabeth Ficker, Andreas Scherwitzl, Josef Zoppoth, Christian Deutsch, Helmut Wolf, Peter Heger, Maximilian Brandeisz, Fritz Hofmann, Franz Domes, Christian Makor-Winkelbauer, Karl Mark, Franz Köck, Josef Edler, Maria Emhart, Hannes Androsch, Ilona Graenitz, Franz Klinger, Hermann Lackner, Alfred Haufek, Karl Dobnigg, Sepp Oberkirchner, Erhard Meier, Heinrich Kuba, Albert Sever, Josef Ackerl, Stefan Billes, Herbert Kautz, Julius Lukas, Albrecht Konecny, Nurten Yilmaz, Matthias Achs, Helene Potetz, Martin Schlaff, Karin Renner, Johann Kranitz, Josef Püchler, Maria Hlawka, Richard Bernaschek, René Pfister (Politiker), Maria Hagleitner, Josefine Oschmalz, Josef Kratky, Therese Šip, Cornelia Schmidjell, Josef Öller, Bernhard Vogel, Georg Thomschitz, Herbert Pansi, Franz Fritsch, Arno Kosmata, Beatrix Eypeltauer, Leopoldine Pohl, Max Ellmann, Franz Hellwagner, Barbara Gross, Josef Gönitzer, Josef Gerl, Franz Schleich, Siegmund Astner, Stefan Demuth, Klaus Zenz, Leopold Millwisch, Josef Schneeweiß, Alois Stöllinger, Eduard Kittl, Jolanda Offenbeck, Anton Brennsteiner, Bruno Marek, Willibald Stacherl, Hellmuth Schipani, Josef Loos, Richard Seidel, Johann Wohl, Otto Rösch, Gerhard Freund, Jürgen Schabhüttl, Franz Samwald, Roland Kaltenbrunner, Rudolf Schicker, Rosa Lohfeyer, Meinrad Hämmerle, Walter Schopf, Johann Pölz, Otto Probst, Siegfried Lindenmayr, Ferdinand Fageth, Adolf Müller, Heinrich Zwenk, Herbert Schmidtmeier, Matthias Kriegner, Leopold Berthold, Johann Brunauer, Dietmar Keck, Josef Hesoun, Brigitte Schwarz, Josef Pazelt, Karl Spielbüchler, Sabine Promberger, Matthias Eldersch, Maria Tusch, Katharina Kucharowits, Siegfried Nasko, Gisela Wurm, Gerhard Steier, Anton Kohl, Josef Weißkind, Edith Mühlgaszner, Hans Laimer, Martin Weber, Otto Kernstock, Hans Mayer, Ilse Oberländer, Josef Peck, Godwin Schuster, Hans Kaiser, Reinhold Entholzer, Kurt Wallner, Jonny Moser, Alois Pelzmann, Robert Bechinie, Stefan Wirlandner, Michael Frühwirth, Peter Milford-Hilferding, Bonaventura Berloschnik, Edith Sack, Elisabeth Hlavac, Franz Zelenka, Karl Ausch, Helmut Wiesenegg, Peter Rezar, Gisela Laferl, Jörg Neumayer, Franz Bezucha, Helmut Schagerl, Käthe Kainz, Walter Blachfellner, Johann Driemer, Cornelius Flir, Hans Jörg Schimanek, Otmar Brix, Sepp Steinhuber, Fritz Preiß, Rudolf Gelbard, Waltraud Karner-Kremser, Franz Hüttenberger, Franz Gartner, Franz Schön, Josef Podlipnig, Maria Bommer, Gerhard Mock, Josef Hannich, Johann Ginzinger, Julius Grünwald, Ferdinand Wedenig, Rudolf Posch, Sieglinde Trannacher, Ruth Körner, Anton Benya, Josef Tauber, Christine Krammer, Stephan Buger, Georg Kotek, Matthias Euler-Rolle, Ernst Maurer, Johann Hatzl, Franz Birnecker, Karl Blecha, Adele Stürzl, Gabriele Proft, Johann Grillenberger, Josef Mayer, Robert Hein, Emmerich Koller, Heinz Krendl, Alois Mayer, Waltraud Bachmaier-Geltewa, Susanne Kurz, Günther Reiter, Leopold Grünzweig, Josef Pongratz, Petra Bayr, Josef Hartmann, Christine Lapp, Verena Dunst, Alois Mentasti, Johann Pitzler, Konrad Antoni, Marianne Hagenhofer, Franz Moßhammer, Johanna Auer, Rudolf Gschweidl, Willi Gruber, Günter Kovacs, Cornelia Ecker, Josef Afritsch, Herbert Salcher, Johann Janecek, Leo Stern, Wolfgang Schantl, Hans Hobl, Florian Gröger, Stefan Prähauser, Erich Valentin, Matthias Stadler, Andrea Mautz, Markus Rumelhart, Eleonora Hostasch, Elisabeth Jenewein, Franz Ruhaltinger, Hans Ferlitsch, Johann Hell, Josef Bilkovsky, Maria Kraichel, Stella Klein-Löw, Christian Broda, Heinrich Scheibengraf, Heinz Vettermann, Hugo Breitner, Franz Fleischmann, Ludwig Leser, Brunhilde Fuchs, Ernst Pechlaner, Ludwig Hiermann, Ursula Pasterk, Felix Slavik, Walter Bösch, Heinrich Salfenauer, Fritz Klenner, Renate Bauer (SPÖ), Max Winter, Franz Zamponi, Günter Kraft, Ursula Lackner, Herta Winkler, Elfriede Krismanich, Julius Tandler, Karl Krammer, Josef Gassner, Friedrich Hoffmann, Helga Konrad, Ferdinand Skaret, Alexia Getzinger, Karin Schrödl, Franz Mellich, Leopoldine Glöckel, Nicole Cernic, Gerhard Spitzer, Alfred Schachner-Blazizek, Rudolf Fischer, Brigitte Bierbauer-Hartinger, Anna Körner, Franz Resch, Franz Riepl, Edgar Schranz, Karl Steidl, Josef Breitenfelder, Reinhold Diessner, Helene Tschitschko, Josef Brunauer junior, Gabriele Burgstaller, Norbert Bacher-Lagler, Beate Schasching, Friedrich Svihalek, Johann Koppensteiner, Erich Haider, Erika Seda, Adam Pichler, Elisabeth Thury, Josef Leichtfried, Karl Dampier, Helga Hieden-Sommer, Gertrude Spieß, Conrad Lötsch, Erika Scharer, Ludwig Klein, Hubert Voraberger, Nicole Berger-Krotsch, Albin Dostal, Gerald Klug, Viktor Schneider, Juliane Lugsteiner, Günther Novak, Bernadette Kerschler, Max Jessner, Vinzenz Muchitsch, Josef Wesenauer, Ferdinand Häuslmayer, Karl Pick, Günther Leichtfried, Walter Strutzenberger, Rosa Rück, Adolf Schütz, Helmut Hirt, Ferdinand Hanusch, Helga Ahrer, Karin Greiner, Birgit Gerstorfer, Friedrich Szenkuröck, Günther Keckeis, Lorenz Schöfbeck, Erwin Buchinger, Rosemarie Schönpass, Emil Schabl, Sepp Wendl, Rupert Fuchs, Alfred Weichselberger, Karl Maisel, Richard Gebert, Hans Ciresa, Anton Mayrhauser, Paul Prigl, Karl Renner, Matthias Gelbmann, Hannes Ebner, Georg Dornauer, Rudolf Müller, Alexander Eifler, Johann Giegerl, Fanny Blatny, Gertrude Wondrack, Johannes Gradenegger, Georg Maier, Johann Wohlmuth, Gottfried Albrecht, Johann Reif, Viktor Wuganigg, Leopold Prucha, Fritz Schorn, Anton David, Adalbert Sebastian, Dietmar Hoscher, Ernst Piller, Wilhelm Pröckl, Franz Löschnak, Hannes Swoboda, Walter Fabics, Friedl Fürnberg, Anton Lang, Josef Schmidt, Josef Breurather, Ernst Nedwed, Alexander Stangl, Rosemarie Polkorab, Karl Heinz, Herbert Tieber, Richard Wilhelm, Ernst Haselwanter, Herbert Würschl, Gebhard Grabher, Traude Votruba, Ludwig Bernaschek, Ludwig Parise, Anna Huber, Robert Hergovich, Franz Jonas, Hans Morawitz, Engelbert Rückl, Maria Kubitschek, Matthias Pinter, Renate Mann, Martin Rom, Josef Gamsjäger, Humbert Spitzer, Erich Gumplmaier, Max Lercher, Rudolf Moser, Wilhelm Ellenbogen, Willibald Krenn, Helene Pecker-Zörner, Josef Schilowetz, Ferdinand Icha, Rudolf Friemel, Heinrich Knotzer, Otto Skritek, Norbert Pichler, Franz Sebek, Omar Al-Rawi, Oswald Hillebrand, Franz Riedmann, Karl Düh, Gertraud Knoll, Ana Blatnik, Hans Schuster, Julie Rauscha, Werner Stadler, Josef Ostermayer, Klaus Luger, Irene Sponner, Michael Rosecker, Karl Schramek, Hermann Reichl, Reinhard Winterauer, Anna Czerny, Alois Striny, Josef Kalina, Hilde Seiler, Franz Strobl, Ernst Höger, Franz Rubant, Reinhard Todt, Silvia Stöger, Karl Hrdlicka, Emmerich Schwemlein, Emil Polke, Josef Handler, Josef Geßl, Elisabeth Märk, Johann Schorsch, Inge Posch-Gruska, Margit Schneider, Jakob Strauß, Arnold Grabner, Martin Tatzber, Heinrich Gross, Hilde Hawlicek, Günther Molzbichler, Wolfgang Katzian, Max Lakitsch, Wolfgang Ruttenstorfer, Franz Scheidl, Anton Weidinger, Anna Mayr, Karl Toman, Alois Alberer, Gerhard Abraham, Rudolf Wehrl, Josef Novak, Ulrike Königsberger-Ludwig, Roman Heinz, Leopold Vogl, Franz Ekkamp, Hans Hladnik, Mirjam Fischer, Maria Metzker, Christa Prets, Franz Koller junior, Josef Schoder, Karl Krikler, Gerhard Frasz, Emmerich Wenger, Karl Markut, Simon Abram, Franz Reiter, Winfried Bruckner, Michael Pay, Eugen Freund, Rudolf Niklas, Hans Schabes, Gerhard Razborcan, Hubert Schnofl, Anna Schlaffer, Gabriele Heinisch-Hosek, Tobias Piff, Rudolfine Muhr, Klaus Hutterer, Otto Möbes, Wolfgang Bulfon, Johannes Schnizer, Otto Schmidt, Günter Weninger, Rudolf Prinz, Josef Horn, Edmund Reismann, Wilhelm Liwanec, Hans Witzany, Hubert Koller, Karl Graf, Alfred Tiefnig, Anton Franz Hölzl, Annemarie Pawlik, Jakob Reumann, Leopold Winarsky, Josefine Welsch, Wolfgang Rainer, Alois Junker, Mario Mühlböck, Johann Kalliauer, Ernst Nevrivy, Katharina Steindl, Johann Affenzeller, Hans Schickelgruber, Thomas Pupp, Adolf Populorum, Valentina Adler, Eduard Weikhart, Leopold Eckhart, Kurt Eder, Gertrude Tribaumer, Viktor Pratl, Josef Kunst, Adolf Fauland, Franz Thanhofer, Anton Fux, Josef Fischer, Anton Regner, Elisabeth Grossmann, Johann Resch, Franz Zingler, Stefanie Psonder, Helmut Kapeller, Ernst Hallinger, Ingrid Salamon, Otto Glöckel, Robert Danneberg, Max Klein, Christine Muttonen, Heinz Fischer, Gerhard Rupp, Herbert Dinhof, Josef Breitenbaumer, Marianne Klicka, Franz Lanik, Michael Berthold, Grete Laska, Franz Gallent, Günter Haiden, Stefan Trenovatz, Christian Peterka, Ingrid Winkler, Katharina Pfeffer, Renate Anderl, Eduard Rieger, Karl Zillner, Michaela Grubesa, Harald Baumgartner, Maurice Androsch, Ewald Schnecker, Pius Schneeberger, Alois Wessely, Johanna Schicker, Fred Sinowatz, Helmut Cerwenka, Eleonore Hödl, Rudolf Edlinger, Adolf Schärf, Friedrich Karrer, Max Adler, Ferdinand Obenfeldner, Renate Seitner, Johann Windsteig, Andreas Kollross, Josef Heinzl, Eva-Maria Hatzl, Johannes Jarolim, Helga Hammerschmied, Helmut Bieler, Rosa Heinz, Robert Mehr, Heinz Brantl, Johann Payer, Ilona Tröls-Holzweber, David Brenner, Christa Vladyka, Anton Gaál, Hans Czidlik, Hans Lindenberger, Gerhard Praschak, Alfred Dallinger, Markus

Zelisko, Elisabeth Grimling, Friedrich Ficker, Georg Papai, Wolfgang Sodl, Johann Werndl, Aline Furtmüller, Leopold Mistinger, Ilse Mertel, Emil Baumgärtel, Reimar Gradischnik, Stefan Schemer, Renate Brauner, Josef Reixenartner, Herbert Sivec, Franz Christoph, Leopold Petznek, Anton Leikam, Jakob Bertsch, Josef Seilinger, Hans Draxler, Andrea Kuntzl, Peter Kostelka, Karl Boden, Josef Kerschbaum, Oskar Weihs, Elke Sader, Johann Wedl, Kurt Wagner, Gerhard Skiba, Hans Kerstnig, Annemarie Zdarsky, Rudolf Köppl, Josef Weiszl, Karl Knechtelsdorfer, Erich Kalteis, Gerhard Klausberger, Theodora Konecny, Nina Abrahamczik, Franz Peyerl, Josef Tomschik, Hans Thirring, Emil Fucik, Erich Trummer, Richard Freund, Franz Schürer, Johann Moser, Leo Habringer, Adolf Czettel, Michael Lindner, Franz Gratzer, Peter Ambrozy, Gerhard Weißenberg, Gertrude Perl, Adolf Schachner, Maria Enser, Josef Pfaller, Leopold Wimmer, Bertha Braunthal, Richard Strasser, Gerd Brand, Anton Grabenhofer, Georg Kerschbaumer, Georg Emmerling, Elisabeth Blanik, Hans Heigl, Christian Puswald, Martin Wabl, Peter Binder, Hans Ausserwinkler, Josef Derbl, Walter Wimmer, Karl Wrba, Franz Blabolil, Anton Ebner, Klaus Konrad, Stefan Hirsch, Herbert Thumpser, Ingrid Holzhammer, Franz Fürst, Franz Hums, Paul Peter, Leopold Wagner, Reinhold Suttner, Alwin Riedmann, Arnold Riese, Ludwig Hofmann, Karl Antal Mühlberger, Gudrun Mosler-Törnström, Walter Renner, Julius Fischer, Laura Rudas, Kathrin Gaal, Heinrich Widmayer, Bruno Frei, Heidrun Walther, Günter Leikam, Erich Hofstetter, Heinrich Schneidmadl, Harald Hofmann, Erich Hohenberger, Franz Dittelbach, Michael Genner, Anton Anderl, David Redecsy, Stephan Radinger, Klaus Mezgolits, Peter Pfeifenberger, Stefan Schennach, Michael Lampel, Klemens Friemel, Franz Hechenblaickner, Josef Auer, Marianne Gusenbauer-Jäger, Alois Mosler, Franz Prokop, Johann Buttinger, Erich Farthofer, Franz Wendl, Ingo Baumgartner, Anton Hueber, Erika Stubenvoll, Franz Müller, Michael Ausserwinkler, Renate Angerer, Karl Mantler, Fritz Koubek, Emmerich Frömel, Otto Mitterer, Franz Stampf, Daniela Gruber-Pruner, Josef Leitner, Ewald Sacher, Max Strache, Marcus Gremel, Anton Figl, Rolf Neidhart, Lotte Pirker, Ingrid Smejkal, Sonja Toifl-Campregher, Karl Wöllert, Stephanie Graf, Wanda Brunner, Robert Weisz, Gerald Tychtl, Nicole Solarz, Friedrich Schatzmayer, Marie Bock, Theodor Meißner, Josef Staudinger, Norbert Darabos, August Sigl, Karl Fischer, Franz Böröczky, Ferdinand Vouk, Hans Suchard, Johann Leichin, Maria Mosbacher, Herbert Platzer, Hans Lagger, Franz Polzer, Josef Steiner, Johann Eibensteiner, Josef Medl, Franziska Krämer, Adolf Kräutl, Paul Richter, Hugo Mayer, Friedrich Knotzer, Hilde Schaumberger, Horst Freiberger (Politiker), Friedrich Robak, Christian Illedits, Günther Nenning, Kurt Gaßner, Anton Probst, Josef Graf, Fritz Hochmair, Wolfgang Beer, Paul Koller, Hans Pramer, Josef Panzenböck, Alois Stacher, Walter Brauneis, Josef Wille, Rudolf Kirchschläger, Rudolf Tremmel, Ewald Nowotny, Josef Adlmannseder, Alois Wimberger, Helmut Zimmermann, August Forstner, Anton Proksch, Heinz Kapaun, Heinz Schaden, Robert Preußler, Veronika Schuster, Hans Litschauer, Oskar Grünwald, Gabriele Binder-Maier, Heinrich Stern, Franz Zoller, Dieter Antoni, Karl Bregartner, Otto Bruckner, Johann Astl, Johann Giefing, Günter Dietrich, Ernst Zodl, Heinz Gärtner, Hugo Schanovsky, Oskar Huemer, Victor Adler, Franz Katzengruber, Heinrich Hackenberg, Viktor Elser, Wilhelm Haberzettl, Michael Ritsch, Stephan Tull, Johann Gion, Wolfgang Luftensteiner, Rainer Wimmer, Michael Schacherl, Viktor Petschnik, Franz Birner, Stefan Springschitz, Lothar Müller, Irmtraut Karlsson, Josef Dametz, Hermann Kepplinger, Klaus Gasteiger, Albert Schultz, Hans Matzenauer, Eduard Rösch, Max Eibegger, Hilde Wanner, Pamela Rendi-Wagner, Alfred Haberl, Hans Gross, Josef Karl Friedjung, Hannes Gschwentner, Anton Kimml, Alois Grebien, Johann Koplenig, Herbert Kaufmann, Karl Czernetz, Helga Markowitsch, Johanna Köberl, Franz Matzke, Leopold Simperl, Hans Bögl, Ludwig Hrebacka, Robert Rada, Josef Wagast, Anna Pöpperl, Heidemaria Onodi, Andrea Gessl-Ranftl, Rupert Kollinger, Hans Pexa, Franz Mayer, Anna Elisabeth Haselbach, Erwin Steinmaßl, Franz Fekete, Josef Wiedenhofer, Ernst Willner, Beate Prettner, Ingrid Riezler, Franz Schuhmeier, Josef Greußing, Werner Friedl, Andreas Staribacher, Ludmilla Parfuss, Hans Hesele, Ernst Woller, Otto Winter, Kurt Heller, Ernst Ulbrich, Anica Matzka-Dojder, Rudolf Fußi, Amalie Seidel, Maria Kren, Sonja Ramskogler, Brigitte Ederer, Josef Bachmayer, Karl Pülzl, Ricky Veichtlbauer, Wilhelm Sigmund, Eberhard Heider, Astrid Eisenkopf, Gerald Bischof, Othmar Raus, Wolfgang Motz, Peter Strasser, Norbert Grinninger, Hilde Borik, Ilse Pfeffer, Hannes Primus, Hans Wolfmayr, Siegfried Pichler, Reinhold Scheucher, Wilhelm Knottek, Othmar Nacovsky, Gertrude Aubauer, Josef Maderner, Walter Peter, Ernst Kogler, Karl Lausecker, Karl Holoubek, Hermann Wielandner, Ernst Fischer, Anton Idl, Michael Amlacher, Franz Enge, Harald Reisenberger, Kurt Scholz, Hans Bock, Willi Fuhrmann, Ilse Reinprecht, Berthold König, Josef Mayr, Franz Plasser, Johann Stippel, Alfred Teschl, Andreas Schieder, Mario Lindner, Maria Berger, Franz Krenn, Kurt Preussler, Kurt Landsmann, Anton Zimmermann, Paul Speiser, Franz Guttmann, Harald Ettl, Thomas Drozda, Walter Brunner, Hermann Buchinger, Franz Bretschneider, Peter Wittmann, Erwin Frühbauer, Franz Kurz, Helmut Zilk, Hans-Peter Bock, Wilhelm Remplbauer, Johann Maier, Ariel Muzicant, Franz Novy, Sabine Oberhauser, Gerhard Zeiler, Josef Karl, Liane Pluntz

- **ÖVP**: Barthold Stürgkh, Viktor Müllner, Walter Temmel, Rudolf Sommer, Erich Schwärzler, Engelbert Bogenreiter, Franz Moser, Ernst Gödl, Leopold Helbich, Josef Taus, Michael Mangold, Leopold Babitsch, Karl Tolde, Ludmilla Martinkovics, Johann Hattmannsdorfer, Walter Murauer, Walther Weißmann, Fridolin Stadelmann, Franz Leitenbauer, Franz Pacher, Adalbert Görcz, Hans Steinacher, Ignaz Hofmacher, Alois Puntigam, Georg Prader, Franz Josef Birnbaumer, Werner Zögernitz, Rudolf Schwarzböck, Maria Rauch-Kallat, Bettina Ellinger, Wolfgang Mayer, Martin Huber, Christian Stöckl, Ernst Winsauer, Johann Bauer, Gerhard Anzengruber, Margarete Aburumieh, Karl Schmalzbauer, Walter Vogl, Florian Kreibich, Bernhard Gansrigler, Otto Fend, Alfons Stadelmann, Heinz Becker, Johann Neumann, Eugen Fleischacker, Alfred Gasperschitz, Patricia Alber, Grete Rehor, Otto Scheff, Franz Rützler, Lambert Rohrböck, Erwin Pokorny, Ludwig Sexl, Heribert Kienberger, Engelbert Lengauer, Johann Hingsamer, Franz Stocker, Franziska

Huber, Leopold Schönbauer, Johann Breiteneder, Josef Wüger, Hans Diettrich, Notburga Schiefermair, Wolfgang Hattmannsdorfer, Eva-Maria Lipp, Hermann Haller, Richard Piaty, Georg Böhm, Paul Katzberger, Karl Sandmeier, Claudia Hirn, Franz Legerer, Josef Hummer, Annemarie Kletzl, Josef Keiblinger, Franz Hiesl, Andreas Schnider, Franz Eßl, Sigurt Bauer, Bartholomäus Hasenauer, Franz Thury, Günter Puttinger, Nikolaus Lanner, Wolfgang Schürrer, Josef Kaspar, Barbara Feldmann, Josef Riegler, Rudolf Geißler, Alfred Rohr, Alois Fehringer, Fritz Obrutschka, Ludwig Bieringer, Ernst Reinhold Kerber, Josef Wöginger, Manfred Sampl, Johann Schweinberger, Karl Wilfing (1960), Gedeon Fritz, Ruth Feldgrill-Zankel, Peter Leitner, Gottfried Kneifel, Fritz Prior, Roman Stiftner, Günther Hummer, Richard Seeber, Hans Pitschmann, Bernhard Rinner, Konrad Eberle, Dietmar Alge, Maria Christine Jachs, Wilhelm Salzer, Christoph Wolf, Georg Schreiner, Herbert Bacher, Hans Ager, Hermann Rainer, Franz Eduard Kühnel, Ignaz Battlogg, Gottfried Ertl, Armin Forstner, Johanna Mikl-Leitner, Hermann Kletzmayr, Adolf Harwalik, Karl Neubauer, August Eberhard, Josef Pröll, Christian Switak, Franz Riebenbauer, Anne Marie Wicher, Gottfried Nindl, Konrad Königswieser, Christian Posch, Rudolf Mackowitz, Theresia Fletschberger, Richard Schmitz, Franz Hietl, Ferdinand Umenberger, Johann Ober, Kurt Leitl, Manfred Hofinger, Franz Weber, Matthias Hödlmoser, Franz Burgstaller, Franz Kaindl, Hermann Lußmann, Wilhelm Neusser, Michaela Resetar, Gerhard Karner, Erich Watzl, Josef Marchsteiner, Jakob Greber, Ferdinand Geißlinger, Peter Moosbrugger, Jakob Halder, Eugen Leissing, Christine Oppitz-Plörer, Anna Eisenrauch, Josef Ober, Liese Prokop, Stefan Dobrovich, Bernhard Feuerstein, Leopold Weinmayer, Josef Knosp, Herbert Gaggl, Martina Pühringer, Franz Mayr, Georg Schmelzer, Gerhard Possart, Walter Aichinger, Klaus Madritsch, Johannes Schweisgut, Karl Habsburg-Lothringen, Johannes Schmuckenschlager, Paul Kiss, Ludwig Kowald, Reinhold Lexer, Oliver Ladurner, Hans Kohler, Sixtus Lanner, Asdin El Habbassi, Alois Gföllner, August Bachmayr, Andreas Hörtnagl, Hans Ludescher, Franz Hiller, Peter Harlander, Franz Gabriele, Cordula Frieser, Matthias Tschirf, Beate Gruber, Otmar Petschnig, Andreas Berchtold, Karl Gruber, Erika Burtscher, Tobias Udier, Willi Aberer, Helena Kirchmayr, Kurt Strohmayer-Dangl, Matthias Ellmauer, Paul Burgstaller, Leopold Hartl, Monika Vonier, Hans Georg Fuchs, Hermann Schultes, Ferdinand Reiter, Gert Wiesenegger, Franz Lechner, Helmut Krieghofer, Josef Türtscher, Franz Steindl, Manuela Khom, Anton Schlager, Isabella Leeb, Kaspar Winder, Thomas Klestil, Josef C. Böck-Greissau, Karl Zerlauth, Anton Kaufmann, Josef Luhamer, Josef Stöckler, Franz Lechner (Tirol), Erwin Hornek, Ignaz Köck, Martin Mayerl, Anton Pirchegger, Anton Raffl, Hermann Gahr, Rudolf Hämmerle, Franz Walzer, Albert Welte, Gerfrid Gaigg, Franz Gausterer, Heinrich Ginther, Martina Rüscher, Karl Simlinger, Ignaz Bruckner, Wolfgang Blenk, Robert Seeber, Karl Holper, Michael Strugl, Kurt Fiedler, Franz Prinke, Norbert Sulyok, Josef Saller, Josef Kirchknopf, Andreas Salcher, Johann Pötsch, Bettina Rausch, Walter Rauch, Jakob Wolf, Christine Marek, Johannes Hawlik, Elias Wimmer, Florian Födermayr, Karl Bocek, Lindi Kálnoky, Otmar Holzer, Vinzenz Schumy, Alois

Wührer, Helmut Sampt, Hermann Kraft, Erwin Dirnberger, Reinhold Purr, Alfred Finz, Siegfried Nagl, Andrea Kaufmann, Franz Jeglitsch, Jakob Ammann, Ludwig Hagspiel, Herta Mikesch, Wilhemine Busch, Ilse Giesinger, Johann Müller, Karin Prokop, Stefan Simma, Peter Stoppacher, Josef Schmidl-Haberleitner, Jochen Pack, Wilfried Berchtold, Helmut Klomfar, Ludwig Strobl, Josef Kraus, Wilhelmine Moser, Wolfram Vindl, Hans Langreiter, Anton Glaninger, Herbert Keßler, Ursula Stenzel, Markus Wallner, Alfred Riedl, Josef Trinkl, Franz Richau, Anneliese Junker, Georg Hoffinger, Margit Hofer, Josef Zehetner, Karl Rozum, Doris Schmidl, Kurt Neuner, Josef Hechenbichler, Rosemarie Bauer, Robert Löffler, Thomas Kaufmann, Thaddäus Steinmayr, Fritz Bock, Eduard Köck, Georg Gindl, Michael Göschelbauer, Franz Wegart, Manfred Schulz, Christoph Platzgummer, Franz Blochberger, Franz Schillhuber, Michael Sandhofer, Hans Haager, Christian Buchmann, Karl Fachleutner, August Breininger, René Lobner, Karl Bandion, Benita Ferrero-Waldner, Hans Ehgartner, Ernst Kurz, Josef Maurer, Andreas Brandstetter, Edgar Mayer, Alexander Muxel, Theresia Schiffmann, Herbert Kneifel, Gebhard Amann, Andreas Pum, Barbara Riener, Andreas Josef Mähr, Albert Hofer, Karl Freund, Alois Lugger, Josef Gritschacher, Alois Leitner, Hans-Peter Bischof, Wilhelm Gorton, Stephan Pernkopf, Erwin Pröll, Josef Andreas Feuerstein, Wilfried Stauder, Andreas Liegenfeld, Clemens Martin Auer, Günther Engelmayer, Gerhart Bruckmann, Erich Exel, Elisabeth Leitner, Anton Gindler, Heinz Feierle, Franz Ferdinand Wolf, Hans Zyla, Brigitte Jank, Alois Karisch, Karl Marberger, Josef Ganglberger, Rudolf Heinz Fischer, Franz Bauer, Leopold Theuringer, Wolfgang Gerstl, Ingrid Turković-Wendl, Josef Dienbauer, Albert Schöpf, Ludwig Herberth, Magnus Brunner, Josef Schwarzenbacher, Gabriele Fürhapter, Josef Naderer, Elisabeth Schmidt, Karin Schabus, Gerhard Tusek, Thomas Bernhard, Gerhard Köhlmeier, Josef Helm, Roderich Regler, Franz Kurzreiter, Christian Menzel, Josef Gruber, Pius Michael Prutscher, Felix Ermacora, Bernhard Ebner, Josef Schöchl, Leopold Popp, Elisabeth Pfurtscheller, Germana Fösleitner, Alfred Wahl, Franz Müllner, Johann Schnitzhofer, Leo Radakovits, Edeltraud Lentsch, Johann Haller, Dietrich Heindl, Franz Marx, Simon Illmer, Wolfgang Haunsberger, Otto Wallig, Karl Litschauer, Otto Steinegger, Georg Vetter (Rechtsanwalt), Herbert Stohs, Ernst Fink, Christian Benger, Niko Alm, Viktor Wallner, Gabriel Obernosterer, Franz Eder, Dietrich Teller, Christian Jachs, Walter Naderer, Kurt Hohensinner, Anton Wimmersberger, Johann Zeyer, Rudolf Marktschläger, Johann Hofer, Karl Schwer, Josef Steiner (1926–1990), Günther Hütl, Anton Seiner, Georg Pranckh, Jürgen Bodenseer, Karl Kurzbauer, Rudolf Raffelsberger, Franz Sauerzopf, Hans Gasser, Alfred Rainer, Rudolf Schwaiger, Reinhard Jany, Johann Waltner, Theresia Fröwis, Rudolf Grohotolsky, Franz Bernhard, Petra Bohuslav, Franz Krippner, Agnes Schierhuber, Rudolf Juen, Marcus Franz, Angela Stöckl, Hans Jörg Schelling, Stefan Behm, Erwin Koler, Walter Intemann, Franz Grandl, Armin Wolf, Johannes Ditz, Adalbert Resetar, Karl Lakowitsch, Erich Mauss, Franz Berger, Kurt Fischer, Othmar Karas, Othmar Knapp, Franz Schaffer, Nikolaus Prinz, Ernst Leimböck, Günther Platter, Karl Romsy, Manfred Schlögl, Franz Josef Naumann, Katharina

Horngacher, Josef Reischer, Ernst Schindlbacher, Erwin Altenburger, Andreas Köll, Kurt Mathis, Heinrich Gleißner, Manfred Vallaster, Gerhard Böhm, Franz Gruber, Franz Kranebitter, Vinzenz Kotzina, Josef Hamberger, Wolfgang Ulm, Anna Höllerer, Harald Wittig, Alfons Weißgatterer, Helmut Mödlhammer, Hans Maurer, Karl Bader, Willibald Eigner, Markus Figl, Alfons Übelhör, Otto Keimel, Fritz Dinkhauser, Josef Ratzenböck, Otto Gumpinger, Karl Kinzl, Johann Zach, Peter Sonnberger, Günther Wiesinger, Manfred Kainz, Rudolf Rasser, Herbert Lins, Karl Wedenig, Robert Geischläger, Franz Majcen, Franz Xaver Gruber, Josef Letmaier, Fritz Aichinger, Hans Kellner, Karl Pisa, Hermann Kröll, Stefan Szirucsek, Ernst Schoiber, Anton Sutterlüty, Lelio Spannocchi, Angela Orthner, Erich Fidesser, Rudolf Nagl, Rudolf Graf, Peter Hirsch, Othmar Tödling, Susanne Reichard, Valentin Deutschmann, Peter Polleruhs, Kurt Hackl, Helmut Mader, Johann Schwarz, Eva Maria Gattringer, Kurt Balla, Rudolf Strommer, Georg Spiegelfeld-Schneeburg, Emmerich Hirmann, Herwig Hofer, Udo Illig, Bernhard Dworak, Rudolf Sallinger, Hermann Gohm, Helmut Doppler, Josef Spiess, Alois Hötzendorfer, Anton Emminger, Thomas Steiner, Franz Rennhofer, Kurt Kraft, Wilhelm Molterer, Eduard Hamedl, Rupert Wolfgruber senior, Josef Prober, Heribert Donnerbauer, Josef Strommer, Andreas Khol, Johann Kuchner, Anton Weindl, Alexander Nemecz, Alexander Neuhuber, Josef Fill, Franz Grundwalt, Elisabeth Scheucher-Pichler, Rudolf Hiessl, Walter Entholzer, Anton Bonimaier, Karl Tizian, Irene Bereuter, Franz Mold, Eduard Heinl, Hermann Lein, Walter Fritz, Peter Rieser, Claudia Schöffmann, Eduard Rauch, Karl Dittrich, Hans Kinsky, Bernd Schilcher, Sonja Ledl-Rossmann, Siegmund Stemer, Johann Wagner, Ernst Guth, Roland Minkowitsch, Pius Fink, Franz Leisser, Stephan Fischer, Marianne Lembacher, Josef Heinrich Sommer, Wilhelm Kerl, Hans Bischof, Karl Blaimschein, Martin Schuster, Maria Fekter, Oskar Natter, Rouven Ertlschweiger, Christine Fröhlich, Norbert Kapeller, Anton Türtscher, Anton Mattle, Franz Brandl, Klaus Hofmann, Herbert Guglberger, Reinhard Rack, Franz Wenger, Edmund Freibauer, Margit Göll, Christoph Kainz, Karl Fichtinger, Romuald Niescher, Josef Straßberger, Adolf Stricker, Karl Glaser, Erich Spindelegger, Karl Eichinger, Otto Pelzelmayer, Hermann Pramendorfer, Adelheid Fürntrath-Moretti, Gerold Ratz, Gregor Hammerl, Aloisia Fischer, Peter Oberlehner, Josef Mühlbachler, Josef Rathgeb, Max Mairinger, Franz Stampler, Friedrich Schwarzott, Hans Gamper, Christian Sagartz, Werner Falb-Meixner, Peter McDonald, Hermann Gruber, Hans Bürkle, Otto Glück, Georg Franz, Hans Wöckinger, Emil Grinzinger, Hubert Hofeneder, Karl Heinz Dernoscheg, Christoph Leitl, Heinrich Schmelz, Karl Kaplan, Hans Rieder, Franz Thoma, Günther Leiner, Alfons Längle, Emmy Göber, Franz Regensburger, Hermann Geißler, Johann Loos, Leopold Steinbichler, Brigitta Pallauf, Leopold Mantler, Rudolf Trauner senior, Ulrich Ilg, Adolf Concin, August Kargl, Michael Hammer, Fritz Neugebauer, Barbara Hollomey, Karin Hakl, Franz Kroyer, Adolf Lukeschitsch, Hella Ranner, Theresia Neuhofer, Heinrich Finder, Josef Renner, Erwin Köstler, Hans Scharfetter, Heinz Peter Ritter, Helga Wieser, Karl Schneider, Gernot Blümel, Gerald Nosko, Alfons Troll, Karl Petritz, Ludwig Schleritzko, Heinz

Illigen, Lujo Tončić-Sorinj, Peter Tschernko, Franz Wieser, Maria Gstaltmeyr, Kurt Staska, Josef Mittendorfer, Karl Pisec, Bernd Stöhrmann, Friedrich Mayr-Melnhof, Josef Steinböck, Karl Stolz, Johann Penz, Ferdinand Mayer, Karl Frodl, Thomas Einwallner, Johann Fraissl, Patrizia Zoller-Frischauf, Harald Himmer, Alfred Mayer, Martin Saller, Peter Marboe, Hans Rödhammer, Alois Witrisal, Johann Kurzbauer, Franz Mayrhofer, Josef Fritz, Josef Huber, Luis Fuchs, Karl Moser, Walter Grasberger, Ottilie Liebl, Kathrin Kaltenhauser, Roland Frühstück, Franz Feldgrill, Elmar Rümmele, Georg Schwarzenberger, Andrea Kdolsky, Benedikt Bittmann, Norbert Mayer, Karl Pischl, Robert Parzer, Anton Flatz, Friedrich Heinzle, Johann Heuras, Rudolf Harramach, Hubert Huber, Josef Waidhofer, Johann Nagl, Helmut Kritzinger, Manfred Rein, Johann Winetzhammer, Franz Marent, Friedrich Hensler, Leopold Gansch, Friedrich Langthaler, Anton Brugger, Alois Scheibenreif, Vinzenz Höfinger, Josef Krainer senior, Maria Schaumayer, Anton Rohrhofer, Karl Seiringer, Peter Vadasz, Maximilian Handel, Anton Eder, Wilhelm Bleyer, Josef Lettenbichler, Maria Grander, Ludwig Weiß, Siegfried Gasser, Herwig van Staa, Michael Mautner Markhof, Josef Molterer, Franz Hörl, Michael Bachinger, Marion Lederhos, Walburga Beutl, Karl Wegerer, Ferdinand Maier, Helmut Kukacka, Hans Vollmann, Gottfried Feurstein, Wolfgang Waldner, Erwin Zajicek, Michael Hülmbauer, Walther Pessl, Annemarie Brunner, Jakob Brandstätter, Rudolf Kristofics-Binder, Josef Stampfl, Gebhard Halder, Josef Stemberger, Robert Graf, Bertl Göttl, Günter Kößl, Manfred Drennig, Bernadette Mennel, Ludwig Mayer, Sebastian Kurz, Anton Fürst, Josef Zingl, Heinrich Kofler, Alexandra Pichler-Jessenko, Hermann Gaisbichler, Otto Lehner, Hanns Koren, Maria Hosp, Josef Lentsch, Alois Grauß, Alois Mock, Josef Gilschwert, Werner Groiß, Heinrich Ott, Johann Bacher, Anton Frisch, Sandra Holasek, Rudolf Gehring, Erich Putz, Rudolf Kait, Jürgen Maier, Josef Martinz, Werner Gradwohl, Franz Breitwieser, Nikolaus Hetfleisch, Johannes Prochaska, Barbara Schöbi-Fink, Barbara Eibinger, Stefan Zangerl, Anton Bayr, Konrad Landgraf, Walter Suppan, Wolfgang Rümmele, Balthasar Gierlinger, Franz Karasek, Martin Nehrer, Wilfried Haslauer junior, Margit Kraker, Anton Hüttmayr, Michael Walla, Oskar Mikula, Konrad Streiter, Albert Steidl, Leopold Weiss, Helga Rabl-Stadler, Josef Hans, Andrea Fraunschiel, Franz Fischler, Karl Titze, Heribert Steinbauer, Christine Haberlander, Hermann Ach, Franz Fartek, Hugo Westreicher, Sandra Kern, Heinrich Amtmann, Armin Rhomberg, Hans Mayr, Franz Lorenzoni, Alois Cipin, Ines Schneider, Thomas Winsauer, Otto Bernau, Odo Wöhry, Siegfried Dohr, Fritz Pelikan, Alfred Dirnberger, Erich Holzinger, Klaus Schneeberger, Gabriele Nußbaumer, Gerhard Schödinger, Wolfgang Sobotka, Johann Karall, Josef Knoll, Johannes Miller, Richard Hogl, Lorenz Karall, Michaela Gansterer, Josef Tazreiter, Frieda Mikola, Sylvester Leer, Manuela Hack, Matthias Krempl, Herta Wimmler, Kurt Jungwirth, Franz Fiedler, Lukas Huber, Johann Erhardt, Susanne Wegscheider, Wolfgang Eisl, Josef Moosbrugger, Karl Rigl, Johannes Peinsteiner, Alfred Brader, Gerhard Schäffer, Franz Xaver Zweimüller, Johanna Priglinger, Hans Sassmann, Fritz Grillitsch, Josef Klaus, Anton Erber, Franz Gschnitzer, Lois Weinberger, Nadine Paunovic, Franz Grubhofer, Dorothea

Schittenhelm, Christian Fiegl, Elisabeth Gehrer, Karl Hubinger, Heribert Oswald, Jochen Danninger, Christian Pongruber, Markus Malle, Peter Mayer, Raimund Grilc, Clemens Steindl, Edmund Grünsteidl, Sabine Schwarz, Rudolf Mandl, Franz Altmann, Werner Miedl, Karl Homole, Otto Friedrich Winter, Johann Schrotter, Werner Huber, Karl Friedl, Walter Tancsits, Kaspar Schwärzler, Jürgen Höckner, Franz Berl, Gerald Schöpfer, Elisabeth Rechnitzer, August Wöginger, Sonja Zwazl, Isidor Harsieber, Alois Huemer, Johannes Rauch, Notburga Astleitner, Roswitha Stadlober, Johann Wolf, Wolfgang Großruck, Gottfried Jaud, Karl Weichselbaumer, Ingrid Tichy-Schreder, Josef Reither, Silvia Grünberger, Franz Korinek, Rudolf Janzsa, Annemarie Adlassnig, Siegfried Sattlberger, Ferdinand Prirsch, Ernst Strachwitz, Vincenz Liechtenstein, Robert Lutschounig, Franz Schöberl, Josef Dengler, Siegmund Burger, Christoph Kaufmann, Bernhard Hirczy, Albert Leibenfrost, Johann Weidenholzer, Gottfried Wunder, Elfriede Blaickner, Johann Rennhofer, Alois Margreiter, Albert Handle, Alfred Hoch, Ernst Fritz, Hans Roth, Rudolf Marwan-Schlosser, Reinhold Lopatka, Martin Schreiber, Erwin Schauer, Kurt Bergmann, Franz Morak, Friedrich Hahn, Anton Nigl, Lorenz Landl, Fritz Eckert, Franz Rabl, Felix Wachter, Josef Eisl, Hans Gesellmann, Franz Stöhr, Arno Gasteiger, Wolfgang Schüssel, Josef Mölschl, Andrea Wolfmayr, Hermann Hartleb, Regina Aspalter, Wilhelm Mandl, Sirvan Ekici, Erwin Machunze, Josef Kast, Adolf Leskovar, Hans Heger, Kurt Buchinger, Maria Haidinger, Franz Glaser, Franz Haunschmidt, Karl Aichhorn, Johann Staggl, Helmut Frauscher, Wilhelm Heißenberger, Elisabeth Manhal, Arnold Weixelbaumer, Jörg Iro, Andrea Gottweis, Erwin Schranz, Johannes Hahn, Martina Diesner-Wais, Hans Kottulinsky, Josef Plank, Manfred Krendl, Julius Nikles, Franz Schütz, Sepp Hobiger, Ernest Gabmann, Josef Peintner, Josef Hilgarth, Christoph Winder, Matthias Kucera, Rudolf Reisetbauer, Hans Treitler, Edmund Hainisch, Hans Kulhanek, Karl Maitz, Astrid Stadler, Matthias Weghofer, Josef Mayer, Peter Haubner, Matthias Scheiber, Johann Rockenschaub, Hermann Schützenhöfer, Oswald Schobel, Ludwig Fahrnberger, Waltraud Schwammer, Günter Lampert, Karlheinz Kopf, Josef Keßler, Franz Hadriga, Günther Goller, Paula Eisenmann, Leopold Kern, Josef Pühringer, Johann Pabst, Wendelin Weingartner, Richard Kaiser, Josef Reiter, Franz Lach, Johann Hagenauer, Hans Schmid, Erika Weinzierl, Franz Kortschak, Andreas Sprenger, Liv Sprenger, Sonja Ottenbacher, Hans Schmidinger, Karlheinz Rüdisser, Werner Winsauer, Johann Ungersböck, Peter Buchleitner, Michael Praßl, Dolores Bauer, Franz Weinberger, Georg Wurmitzer, Leopold Thurner, Leopold Kunschak, Rupert Roth, Franz Kampichler, Hans Ferlitsch, Otto Hofmann-Wellenhof, Emmerich Haas, Simon Wiesenthal, Franz Greßl, Karl Url, Johannes Bodner, Karl Wilfing, Eva Maria Waibel, Christian Bernhard, Theodor Pritsch, Barbara Schwarz, Michael Killisch-Horn, Hans Hafner, Sonja Stiegelbauer, Herbert Schambeck, Walter Nettig, Rudolf Sohm, Gertrude Kubiena, Gerhard Fallent, Franz-Joseph Huainigg, Jakob Mädl, Martin Michalitsch, Wendelin Ettmayer, Christopher Drexler, Johann Hautzinger, Richard Entner, Julius Raab, Franz Schausberger, Ernst Robetschek, Karl Loidl, Beate Palfrader, Erwin Wenzl, Alois Wölfler, Franz Rechnitzer, Martin Wetschka, Arthur Mussil, Karl Lugmayer, Aloys Oberhammer, Josef Jandrasits, Martin Purtscher, Eva-Maria Himmelbauer, Franz Hasiba, Georg Ludescher, Martin Gassner, Rudolf Eder, Michael Neureiter, Karl Schmitzer, Alfred Kranich, Manfried Welan, Angelika Winzig, Bernhard Baier, Johann Kraker, Johann Endl, Wilhelm Riedl, Theodor Eggendorfer, Johann Schuster, Michaela Langer-Weninger, Karl Hofstetter, Otto Götzl, Matthias Bierbaum, Hans Lechner, Elisabeth Köstinger, Christian Brünner, Carl Heinz Bobleter, Florian Denk, Eduard Ludwig, Hannes Missethon, Hans Tuppy, Günter Stummvoll, Georg Leberbauer, Herbert Paierl, Fritz Mitterhauser, Arthur Rauth, Roman Gutscher, Maria Stangl, Karl Wimmer, Hans Pernter, Ferdinand Riefler, Andreas Maurer, Manfred Mautner Markhof junior, Felix Bergsmann, Gerhard Hirschmann, Heinz Gach, Thomas Katsich, Eva Bassetti-Bastinelli, Walter Hauser, Josef Rupp, Johann Buchinger, Bertram Jäger, Heinrich Neisser, Heribert Bogensperger, Johann Singer, Gottfried Schröckenfuchs, Alois Anzenberger, Michaela Mojzis, Rainer Gögele, Waltraud Ebner, Franz Schaber, Adolf Hosp, Hubert Rauch, Josef Brandmayr, Josef Schatz, Lola Solar, Josef Schrefel, Hans Stefan Hintner, Albert Hochleitner, Robert Lichal, Josef Balber, Ernst Kolb, Andreas Ottenschläger, Johanna Bayer, Felix Hurdes, Gerhard Jellasitz, Hubert Waibel, Marianne Hackl, Willi Sauer, Sebastian Eder, Martin Forsthuber, Doraja Eberle, Matthias Duscher, Wilhelm Gantner, Erich Kabesch, Lukas Schnitzer, Monika Reis, Friedrich Niederl, Manfred Rünzler, Kurt Lentsch, Hermann Kuenz, Herwig Hösele, Friedrich Bernhofer, Harald Schwarzbauer, Gerlinde Rogatsch, Josef Edenhauser, Beatrix Karl, Franz Perhab, Gottfried Brandl, Karl Trabitsch, Hermann Fritz, Leopold Guggenberger, Norbert Sieber, Eva-Maria Posch, Johann Wildt (Niederösterreich), Heinrich Schaidreiter, Martina Pecher, Vinzenz Schwärzler, Alfred Frauscher, Hermann Laferl, Franz Hetzenauer, Adolf Tiller, Paul Rübig, Luis Bassetti, Helga Lusetzky, Hans Ganahl, Rosa Gföller, Gregor Hoch, Marilies Flemming, Andreas Hanger, Anton Galehr, Elisabeth Olischar, Anna-Maria Hochhauser, Doris Hummer, Josef Wallner, Karl Donabauer, Josef Ferschner, Ingeborg Rinke, Günter Widder, Hans Sebinger, Waltraud Klasnic, Michael Obermoser, Ulrike Baumgartner-Gabitzer, Robert Stratil, Wolfgang Stanek, Greti Schmid, Adolf Traußnig, Josef Reich, Claudia Durchschlag, Raoul Bumballa, Karl Lackner, Karl-Albert Eckmayr, Walter J. Mayr, Theodor Cerny, Franz Strobl, Peter Michael Ikrath, Paul Fasching, Karl Vonwald, Josef Haslinger, Josef Anton Fässler, Johannes Tratter, Anton Karl, Margarethe Ottillinger, Eugen Margarétha, Gilbert Frizberg, Harald Sonderegger, Andreas Zakostelsky, Johann Halbritter, Johann Grabenhofer, Hermann Withalm, Anton Weidinger, Franz Stoll, Franz Karl Auersperg, Ferdinand Tiefnig, Ottilie Rochus, Maria Hampel-Fuchs, Maria-Luise Egerer, Josef Dinkhauser, Peter Morwitzer, Franz Koller junior, Lukas Mandl, Nikolaus Berlakovich, Fritz Polcar, Josef Geisler, Heinz Bereuter, Gerta Schönbichler, Walter Schwimmer, Peter Eisenschenk, Erich Edegger, Johann Seitinger, Friedrich Hoess, Franz Sodomka, Georg Rosner, Gabriele Tamandl, Ernst Grundemann-Falkenberg, Johann Habeler, Günter Vetter, Stephan Michael Tauschitz, Hans Tinhof, Eduard Ehrenhöfler, Helmut Schreiner, Leopold Weismann, Leopold Eichinger, Josef Höchtl, Gustav Hack, Konrad Steindl, Josef Rauch, Eleonora Hiltl, Johann

Blöchl, Anton Pertl, Rudolf Trauner junior, Ignaz Tschurtschenthaler, Günther Anton Moosbrugger, Karin Praniess-Kastner, Erika Adensamer, Josef Edlinger, Hans Schumi, Johann Eichinger, Klaus Hubert Auer, Michael Spindelegger, Friedrich Fuchs, Siegfried Egger, Johann Tesar, Franz Neudorfer, Michaela Hinterholzer, Monika Riha, Florian Riedl, Kurt Kaufmann, Johann Gregorich, Werner Roßmann, Hans Katschthaler, Franz Schlegl, Martin Müller, Rudolf Staudinger, Ursula Plassnik, Richard Kaan, Gerald Loacker, Josef Sampl, Johann Kammerhofer, Julius Hahn, Franz Leopold, Franz Brunner, Wolfgang Aigner, Heinrich Drimmel, Ernst Reinhold Lasnik, Karl Honeder, Bernhard Tilg, Karl Lueger, Franz Schedl, Leopold Goëss, Walter Heinzinger, Josef Kennerknecht, Johannes Peterlik, Martin Wex, Heinrich Gerstbach, Franz Wachter, Karl Brunner, Carl Lipp, Josef Ausweger, Helmut Batlogg, Georg Resch, Josef Steinkogler, Josef Prechtl, Kristina Edlinger-Ploder, Anton Ehrenfried, Franz Riel, Katharina Cortolezis-Schlager, Wilhelm Thomas, Kathrin Nachbaur, Johann Schweinhammer, Hubert Pirker, Karl Kokesch, Paul Kaufmann, Isidor Grießner, Johann Weissenböck, Hans Großauer, Johann Deutsch, Rudolf Gurtner, Robert Rapatz, Martina Jöbstl, Günther Köberl, Richard Nathschläger, Gerhard Koppensteiner, Ulrich Fitz, Leopold Jochberger, Alfred Maleta, Heribert Gram, Johann Haider, Walter Riedl, Helga Braunrath, Christina Steinmetzer, Josef Salcher, Friedrich Reisinger, Karl Schleinzer, Heinrich Matza, Carina Felzmann, Guntram Lins, Josef Preindl, Karl Nindl, Eva Petrik, Karl Flödl, Reinhard Herok, Ernst Wiesinger, Markus Ulram, Martin Drescher, Erika Danzinger, Hans Klupper, Georg Keuschnigg, Adolf Pinegger, Monika Lugmayr, Felix Kern, Lorenz Konzett, Josef Rehrl, Georg Zimmermann, Franz Dworak, Josef Molzer, Reinhold Mitterlehner, Josef Robl, Karl Baueregger, Manfred Scheich, Elmar Schallert, Stefan Knafl, Sebastian Mitterer, Johann Staudinger, Alois Dienstleder, Eduard Hartmann, Markus Linhart, Werner Reichart, Max Juen, Gerhard Jandl, Johann Wildt, Barbara Schwaighofer, Paul Landmann, Arno Abler, Gottfried Kapferer, Wolfgang Schmitz, Edgar Siegele, Franz Soronics, Josef Scherrer, Elfriede Salzgeber, Johann Farnleitner, Harald Wögerbauer, Bertl Emberger, Alfred Schöls, Johannes Zweytick, Jakob Mayr, Josef Feuerstein, Verena Remler, Marga Hubinek, Julius Grabner, Anton Doppler, Eduard Wallnöfer, Karl Bohle, Franz Flicker, Grete Pirchegger, Alfred Knaller, Kurt Korbatits, Leopold Maderthaner, Herbert Weiß, Friedrich Gugg, Alfons Gorbach, Johann Diwold, Klaus Fürlinger, Peter Krauland, Manfred Juraczka, Herbert Kohlmaier, Walter Mayr, Alfred Römer, Josef Fink, Franz Xaver Böhm, Elisabeth Greiderer, Hannes Krasser, Engelbert Schaufler, Gabriel Wagner, Xaver Muther, Kurt Waldheim, Walter Renner, Helga Machne, Josef Matt, Ernst Herzig, Angela Fichtinger, Stefan Weirather, Peter Rodek, Maximilian Hiegelsberger, Martin Strimitzer, Walter Zimper, Hermann Hagen, Gertrude Brinek, Leopold Husinsky, Albert Römer, Julian Fässler, Karl Lafer, Christian Poglitsch, Maria Elisabeth Möst, Georg Seidl, Paul Rittsteuer, Bernhard Görg, Bernd Toms, Otto Kranzlmayr, Wolfgang Dax, Georg Griessner, Reinhold Polster, Otto Mitterer, Gabriele Lackner-Strauss, Aurelia Gieler, Alois Partl, Andrea Eder-Gitschthaler, Michael Graff, Wolfgang Kasic, Claudia Schmidt, Johann Tschida, Wilfried Haslauer senior, Jakob Auer, Ernst Rauchwarter, Georg Fiedler, Leo Pallwein-Prettner, Josef Winkler, Wolfgang Saliger, Hans Sidonius Becker, Fritz Kaltenegger, Gabriele Juen, Johann Pehm, Johann Steinböck, Johann Gassner, Gustav Kapsreiter, Christine Werber, Josef Stockinger, Gerhard Jelinek, Ferdinand Manndorff, Eduard Pumpernig, Harald Scheucher, Michaela Steinacker, Fritz Hinterndorfer, Johann Ledolter, Oskar Mayer, Adolf Brunner, Anton Steixner, Oswald Klikovits, Johannes Hibler, Heinz Kirchmair, Viktor Sigl, Franz Rupp, Andrä Rupprechter, Hermann Zittmayr, Gerd Stepantschitz, Heinrich Reisinger, Andreas Viehauser, Stephan Koren, Heinrich Wildfellner, Oswald Haberzettl, Martin Preineder, Ludwig Jedlicka, Fritz Astl, Gallus Schmid, Siegfried Eberdorfer, Walter Macher, Hildegard Schorn, Annemarie Dengg, Franz Wolfinger, Otto Keller, Johann Rädler, Leopold Hofinger, Stefan Schmuckenschlager, Franz Latzka, Rudolf Gradinger, Christoph Vavrik, Franz Romeder, Uta Barbara Pühringer, Werner Suppan, Karl Schebesta, Franz Nimmervoll, Franz Murer, Erwin Rasinger, Josef Münzenrieder, Josef Hörl, Josef Scheidl, Anton Peltzmann, Ingrid Korosec, Walter Kornexl, Ludwig Steiner, Johanna Preinstorfer, Manfred Dörler, Milan Linzer, Franz Stangler, Sissy Roth-Halvax, Achill Rumpold, Alois Derfler, Theodor Piffl-Perčević, Walter Lingg, Helmut Lindenthaler, Franz Tschögl, Herlinde Rothauer, Heinrich Wille, Emil Schneider, Adolf Vögel, Rudolf Thienel, Alois Paulitsch, Anton Haller, Erwin Gruber, Anton Schwaiger, Anna Franz, Bernd Schönegger, Karl Dewanger, Josef Wiesler, Alois Stimakovits, Karl Kummer, Gustav Vetter, Christa Gangl, Bernd Petrisch, Friedrich Ofenauer, Sophie Schulz, Josef Ernecker, Andreas Jilly, Franz Beutl, Ferdinand Eberle, Leopold Krottendorfer, Johann Höfinger, Friedrich König, Rudolf Friewald, Leopold Breinschmid, Anna Hosp, Theodor Seykora, Anton Kecht, Josef Ofenböck, Josef Dechant, Alfred Gerstl, Josef Krainer junior, Daniela Gutschi, Josef Frühwirth, Patrik Fazekas, Franz Bartik, Werner Haubenburger, Walter Lehner, Harald Mahrer, Werner Fasslabend, Peter Csar, Herbert Nowohradsky, Martin Bartenstein, Johann Wollinger, Wilhelm Reichart, Christof Zernatto, Kurt Lettner, Otto Amann, Jürgen Weiss, Daniel Steinhofer, Erhard Busek, Andreas Nedwal, Christian Gantner, Hermann Hauer, Karl Lengheimer, Regina Heiß, Paul Raab, Franz Linsbauer, Bernhard Ederer, Hans Tschiggfrey, Anton Kasser, Rudolf Fußi, Leopold Figl, Georg Strasser, Roman Resch, Hans Guggi, Rudolf Hirsch, Konrad Blank, Edgar Schober, Ernst Strasser, Therese Lukasser, Franz Simoner, Friedrich Platzer, Ferdinand Graf, Ridi Steibl, Franz Hell, Christian Dörfel, Wolfgang Gmachl, Ingrid Gady, Johann Hofbauer, Gertrude Aubauer, Inge Sulzer, Josef Schwaiger, Dieter Lukesch, Siegfried Ludwig, Edeltraud Gatterer, Johann Josef Dengler, Edda Egger, Leopold Fischer, Josef Schmall, Doris Schulz, Werner Amon, Hubert Lang (Politiker), Peter Mandorfer, Norbert Walter, Josef Mayr, Franz Kummer, Alexander Haas, Anton Gangl, Bernhard Hofmann-Wellenhof, Herbert Sausgruber, Elisabeth Wappis, Ferdinand Hueter, Eduard Heilingsetzer, Angelika Neuhold, Josef Durry, Michael Haslinger, Josef Lackner, Johann Mayer, Johann Pum, Fritz Reuter (Mediziner), Eduard Moser, Eduard Ulmer, Franz Windisch, Josef Arthold, Michaela Eva Bartel, Karl Etlinger, Helmut Heidinger, Anton Reitzl, Thomas Stelzer, Karl Heincz, Gerhard Ammann

- **FPÖ**: Walter Gatt, Fritz Amann, Heinz-Christian Strache, Gerulf Murer, Benno Sulzberger, Martin Graf, Hermann Brückl, Bernd Brugger, Johann Buchner (Politiker), Peter Samt, Heinz Hochsteiner, Peter Wurm, Fritz Schuler, Heinrich Zechmann, Udo Landbauer, Florian Kasseroler, Markus Fauland, Barbara Schöfnagel, Dietmar Schmittner, Jakob Pistotnig, Lutz Weinzinger, Stefan Salzl, Walter Grabher-Meyer, Hubert Fuchs, Gerd Klamt, Ernst Rothenwänder, Monika Mühlwerth, Gerry Mikscha, Hans Sperger, Karl Wiedner, Daniela Raschhofer, Engelbert Weilharter, Ludwig Buchinger, Gerald Brettner-Messler, Anton Stephan, Ludwig Rader, Peter Westenthaler, Otto Rudolf Braun, Waltraud Dietrich, Gerald Ebinger, Franz Graf, Bernhard Themessl, Ulrike Wall, Paul Tremmel (Politiker), Mario Canori, Jörg Kandutsch, Marlene Svazek, Peter Fichtenbauer, Rudolf Kopf, Martin Huber, Wolfgang Aigner, Peter Handlos, Erich Königsberger, Alfred Eß, Manfred Hofbauer, Beate Hartinger, Georg Grünbart, Peter Sichrovsky, Peter Mitterer, Michael Schmid, Susanne Winter, Susanne Preisinger, Gerulf Stix, Herbert Madejski, Reinhold Huber, Géza Molnár, Josef Brunner, Gerhard Dörfler, Gerhard Bauer, Karl Schweitzer, Dagmar Belakowitsch-Jenewein, Hans Pretterebner, Rupert Schiffrer, Josef Bucher, Christian Leyroutz, Peter Ortner, Robert Wenitsch, Gustav Zeillinger, Peter Harring, Helmut Naderer, Hans Sevignani, Thomas Schellenbacher, Cornelia Michalke, Bruno Hummer, Ilse Benkö, Helmut Haas, Evelyn Kattnigg, Alois Baldinger, Alois Preiszler, Ernest Windholz, Marco Triller, Ernst Ammering, Ilse Hans, Günther Kumpitsch, Eduard Nicka, Martin Zerlauth, Robert Aspöck, Hermann Mentil, Anton Kogler, Michael Kreißl, Robert Scheuch, Mathias Venier, Günther Steinkellner, Carmen Schimanek, Ernst Blum, Dieter Egger, Gerhard Sailer, Sepp Baumgartner, Alois Pumberger, Leopold Schöggl, Georg Lakner, Werner Karg, Helmut Günther, Bernhard Gritsch, Eduard Nicka (Politiker, 1911), Kurt Scheuch, Franz Lafer, Wilhelm Weinmeier, Wolfgang Nußbaumer, Mario Kunasek, Klaus Mahnert, Axel Kassegger, Johann Gudenus, Michael Krüger, Josef Moser, Christian Scheider, Susanne Riess, Alexander Pawkowicz, Harald Vilimsky, Barbara Rosenkranz, Theresia Zierler, Klara Motter, Herbert Kober, Michael Stumpf, Gerhard Hetzl, Ernst Hagen, Edwin Rambossek, Wendelin Mölzer, Gerd Krusche, Gerald Grosz, Jörg Freunschlag, Gerhard Höggerl, Franz Mandl, Dieter Böhmdorfer, Mares Rossmann, Petra Steger, Herbert Kickl, Martina Schenk, Hans Jörg Schimanek, Mathias Reichhold, Helmut Peter, André d'Aron, Karl Vonbank, Franz Marchat, Walter Scherb, Margot Hofer, Reinhard Firlinger, Elfriede Madl, Kurt List, Adalbert Cramer, Roland Zellot, Jutta Arztmann, Sabine Binder, Walter Ebner, Edith Haller, Peter Lechenauer, Christian Hrubesch, Wolfgang Jung, Karl Kowarik, Anneliese Klein, Heidemarie Unterreiner, Erich Slupetzky, Georg Mautner Markhof, Josef Nagler, Wolfgang Zanger, Johann Herzog, Johann Gallo, Christa Kratohwil, Harald Jannach, Richard Rezar, Thomas Barmüller, Karl-Werner Rüsch, Paul Peter, Andreas Eisl, Veronika Matiasek, Christian Höbart, Hannes Amesbauer, Norbert Gugerbauer, Josef Blasisker, Anton Reinthaller, Martin Strutz, Gabriel Wagner, Elwin Blum, Arthur Tiefenthaler, Kurt Ruthofer, Siegfried Kampl, Alois Gradauer, Werner Melter, Hans-Dieter Grabher, Gerhard Kurzmann, Albert Schmidt, Max Stendebach, David Schießl, Konstanze Röhrs, Rudolf Jussel, Georg Pranckh (FPÖ), Detlev Neudeck, Erich Schreiner, Elisabeth Ullmann, Robert Thaller, Johannes Hübner, Patrick Ortlieb, Karin Landauer, Peter Zwanziger, Uwe Scheuch, Christian Ragger, Wilhelm Kindl, Heinz Anton Marolt, Franz Pirolt, Anton Blünegger, Johann Römer, Markus Wiesler, Alois Huber, Wolfgang Spadiut, Wilhelm Niederhuemer, Walter Aichinger, Alexander Götz, Erwin Angerer, Thomas Ram, Josef Gohm, Johann Böhler, Helmut Krünes, Ilse Burket, Erich Tadler, Rüdiger Schender, Hans Kronberger, Andreas Karlsböck, Rosemarie Blattl, Gilbert Trattner, Hans Müller, Fritz Franke, Holger Bauer, Hermann Böhacker, Helga Knicek, Christof Bitschi, Walter Leitner, Erich Hafner, Friedrich Rödler, Jörg Haider, Horst Nußbaumer, Peter Rosenstingl, Albert Danner, Norbert Steger, Anneliese Kitzmüller, Walter Rosenkranz, Willfried Gredler, Walter Bauer, Kornelia Spiß, Richard Heis, Franz Schwager, Ewald König, Harald Ofner, Helena Ramsbacher, Liane Höbinger-Lehrer, Gerald Grebenjak, Gunter Hadwiger, Sepp Wieser, Josef Franz Fritzenwallner, Hildtraud Wieser, Friedrich Kuchar, Siegfried Neyer, Ludwig Nosko, Alois Bachinger, Paul Stadler, Gerhard Fallent, Udo Grollitsch, Helene Partik-Pablé, Karl Schwab, Wilma Warmuth, Manfred Haimbuchner, Helfried Pfeifer, Kriemhild Trattnig, Hans Helmut Moser, Franz Linser, Helga Kügerl, Ursula Haubner, Gottfried Waldhäusl, Elisabeth Sickl, Fritz Ledermann, Wolfgang Ilgenfritz, Reinhart Gaugg, Heike Trammer, Matthias Krenn, Andreas Sodian, Manfred Haidinger, Volker Kier, Lothar Gintersdorfer, Leopold Mayerhofer, Gernot Rumpold, Walter Rauch, Anton Mahdalik, Maximilian Linder, Christoph Waibel, Rudolf Stark, Monika Forstinger, Jürgen Hatzenbichler, Walter Prinz, Wilhelm Brauneder, Edith Mühlberghuber, Manfred Beck, Magda Bleckmann, Günter Pröller, Anna Elisabeth Achatz, Günter Willegger, Renate Kanovsky-Wintermann, Rainer Pawkowicz, Christian Hafenecker, Ludwig Schandl, Maximilian Krauss, Anton Wattaul, Helmuth Josseck, Friedrich Wiedermann, Bernd Lindinger, Dieter Langer, Walter Zechmeister, Reinhard Eugen Bösch, Henriette Frank, Maximilian Hofmann, Josef Brunmair, Elmar Lichtenegger, Herbert Haupt, Bernhard Vock, Gerhard Schmid, Johann Tschürtz, David Lasar, Ewald Stadler, Karin Praxmarer, Robert Lugar, Dominik Nepp, Hermann Eigruber, Elmar Podgorschek, Harald Trettenbrein, Hans-Jörg Jenewein, Jutta Wochesländer, Hubert Gorbach, Christian Lausch, Reinhard Pisec, Ursula Stenzel, Wolfgang Haider, Karl-Heinz Grasser, Adolf Nesset, Bernhard Rösch, Walter Meischberger, Norbert Staffaneller, Erich Reiter, Joachim Weixlbaumer, Rainer Widmann, Franz Koloini, Liane Moitzi, Norbert Burger, Günther Barnet, Franz Hornegger, Anton Böhm, Thomas Schererbauer, Silke Lackner, Herbert Schweiger, Karl Leutgöb, Markus Abwerzger, Erich Silla, Bernhard Gauster, Elisabeth Kaufmann-Bruckberger, Karlheinz Klement, Hans Achatz, Harald Fischl, Kurth-Bodo Blind, Sepp Hainzl, Robert Egghart, Christian Wehrschütz, Josef Kleindienst, Friedrich Peter, Othmar Meißl, Harald Stefan, Heide Schmidt, Walter Ratt, Josef Trenk, Volker Winkler, Erich Rappold, Josef Lauritsch, Helmut Haigermoser, Adolf Stark, Georg Mayer, Eduard Schock, Gerhard Hager, Rupert Doppler, Gerald Hauser, Hans Schöll, Michael Gruber, Friedrich Probst, Wolfgang Haberler, Gernot Darmann, Eduard Mainoni, Gregor Munzenrieder, Alois Kremnitzer, Andreas Schöppl, Evelyn Freigaßner, Gerhard Deimek, Daniel Allgäuer, Lukas Essl, Alexander Nerat, Max Schöringhumer, Manfred Stromberger, Peter Böhm, Michael Raml, Roland Hausmann, Johann Richter, Rudolf Kroiß, Siegfried Dillersberger, Hermann Reindl, Wilhelm Reichart, Maximilian Walch, Emil van Tongel, Tassilo Broesigke, Günther Kleinhanns, Josef Meisinger, Nicole Hosp, Herbert Aspöck, Gerhard Kovasits, Georg Hanreich, Lisa Frühmesser, Hilmar Kabas, Franz Obermayr, Werner Strohmaier, Karl Pfeifenberger, Karl Schnell, Wilhelm Kos, Irina Schoettel-Delacher, Elke Achleitner, Hermann Jantschgi, Sylvia Papházy, Werner Herbert, Helmuth Weiss, Hubert Knaus, Roman Haider, Andreas Mölzer, Marlies Steiner-Wieser, Werner Königshofer, Mario Eustacchio, Reinhard Poglitsch, Barbara Kappel, Herbert Graf, Otto Scrinzi, Andreas Rabl, Gerald Deutschmann, Ulrike Haunschmid, Dietger Mader, Arnd Meißl, Thomas Prinzhorn, Armin Blind, Jessi Lintl, Friedhelm Frischenschlager, Anton Frisch, Josef Hintermayer, Josef A. Riemer, Edmund Tauchner, Herbert Scheibner, Norbert Hofer, Klaus Peter Nittmann, Barbara Kolm, Johanna Trodt-Limpl, Karl Baron, Herbert Eisenstein, Ute Apfelbeck, Alexander Petschnig, Michael Fischer, Horst Schender, Klaus Wittauer, Bernhard Gratzer, Andrea-Michaela Schartel, Christian Allesch, Reinhart Waneck, Wilhelm Grissemann, Herwig Mahr, Franz Koller, Walter Gau, Ambros Nesensohn, Karl Schneider-Manns Au, Michael Rockenschaub, Franz Eitzinger, Erwin Hirnschall, Peter Bahn, Christoph Längle, Heinz-Peter Hackl, Albert Gunzer, Markus Steiner, Anton Knerzl, Wolfgang Klinger, Josef Schelling, Josef Lobnig, Kurt Lukasek, Christof Neuner, Rosa Ecker, Sigisbert Dolinschek, Hedwig Staller, Gerhard Haslinger, Hartmann Lautenschlager, Dietbert Kowarik, Peter Kapral, Günter Schönhart, Albert Royer, Brigitte Povysil, Wilfried Tilg, John Gudenus, Karin Stampfel, Philipp Schrangl, Helga Moser, Manfred Krüttner, Irene Vogt, Marialuise Mittermüller, Hubert Kinz, Helmut Prasch, Robert Bösch, Josef Jury, Hellmut Geringer, Silvia Benzer, Erich Karassowitsch, Werner Neubauer, Christian Cramer

- **Grüne**: Christian Burtscher, Sigrid Maurer, Matthias Köchl, Albert Steinhauser, Maria Vassilakou, Josef Brugger, Waltraut Antonov, Barbara Huemer, Vahide Aydın, Nicole Schreyer, Aygül Berîvan Aslan, Judith Schwentner, Doris Schmidauer, Gebhard Mair, Monika Langthaler, Sabine Gretner, Andreas Angerer, Kimbie Humer-Vogl, Michaela Sburny, Alexander Van der Bellen, Christine Heindl, Simon Hofbauer, Peter Hagenauer, Wolfgang Zinggl, Eva Glawischnig, Marie Ringler, Friedrun Huemer, Marie-Edwige Hartig, Severin Renoldner, Ingrid Lechner-Sonnek, Maria Scheiber, Emmerich Weiderbauer, Alev Korun, Dieter Brosz, Karl Smolle, Margarethe Krojer, Christian Hörl, Doris Kammerlander, Angela Lindner, Barbara Zwerschitz, Günter Kenesei, Brigid Weinzinger, Kurt Grünewald, Şenol Akkılıç, Josef Buchner, Heidemarie Rest-Hinterseer, Monika Vana, Ewa Dziedzic, Christoph Chorherr, Birgit Meinhard-Schiebel, Pius Strobl, Brigitte Flinspach, Walter Geyer, Astrid Kuttner, Faika El-Nagashi, Friedrich Kofler, Helene Jarmer, Mercedes

Echerer, Christine Baur, Franz Breiner, Ingrid Felipe, Madeleine Reiser, Peter Pilz, Thomas Blimlinger, Stefan Kaineder, Cyriak Schwaighofer, Martina Berthold, Edith Zitz, Kaspanaze Simma, Andreas Wabl, Sabine Mandak, Wolfgang Pirklhuber, Lisa Rücker, Günther Nenning, Josef Scheinast, Jennifer Kickert, Severin Mayr, Martin Margulies, Michel Reimon, Martin Fasan, David Ellensohn, Julian Schmid, Ruperta Lichtenecker, Heidemarie Cammerlander, Bettina Hradecsni, Joško Vlasich, Sandra Schoch, Willibald Gföhler, Karoline Hochreiter, Bernhard Ernst, Marco Schreuder, Sabine Jungwirth, Martina Wurzer, Janine Wulz, Harald Walser, Theresia Haidlmayr, Adi Gross, Maria Wageneder, Johannes Rauch, Stefan Wallner, Ingrid Puller, Birgit Schatz, Jutta Kräutler-Berger, Rudolf Anschober, Heinrich Schellhorn, Amrita Enzinger, Eva Mückstein, Elisabeth Kerschbaum, Heidi Reiter, Sigrid Pilz, Tanja Windbüchler-Souschill, Daniel Zadra, Terezija Stoisits, Efgani Dönmez, Maria Buchmayr, Freda Meissner-Blau, Christoph Metzler, Helga Erlinger, Eva Lachkovics, Katharina Wiesflecker, Ahmet Demir, Johannes Voggenhuber, Sonja Puntscher Riekmann, Peter Kraus, Werner Kogler, Lambert Schönleitner, Claudia Smolik, Ulrike Lunacek, Friedensreich Hundertwasser, Ulrich Habsburg-Lothringen, Angelika Hörmann, Susanne Jerusalem, Madeleine Petrovic, Manfred Srb, Rolf Holub, Astrid Rössler, Gustav Götz, Hermann Weratschnig, Lara Köck, Herbert Fux, Stefan Schennach, Marijana Grandits, Christiane Brunner, Holda Harrich, Sandra Krautwaschl, Herbert Thalhammer, Eva Konrad, Helga Krismer-Huber, Sigfrid Peter, Ursula Schwarzl, Martin Rutter, Bernd Bösch, Birgit Hebein, Bruno Rossmann, Gabriela Moser, Rupert Fuchs, Fritz Zaun, Jürgen Himmelbauer, Eva Lichtenberger, Georg Willi, Karin Fritz, Ulrike Schwarz, Elisabeth Wiesmüller, Karl Öllinger, Sonja Pitscheider, Barbara Sieberth, Daniela Musiol, David Stögmüller, Doris Eisenriegler, Rüdiger Maresch

- **NEOS**: Matthias Strolz, Rainer Hable, Bettina Emmerling, Markus Ornig, Michael Bernhard (Politiker), Stefan Gara, Angelika Mlinar, Beate Meinl-Reisinger, Niko Alm, Nikolaus Scherak, Sepp Schellhorn, Christoph Vavrik, Karin Doppelbauer, Martina Pointner, Gerald Loacker, Claudia Gamon

8.2 Grouping of content categories

Most of the articles had a meta-property called internal department ("ressort") which can be understood as a proxy for topic. Thus, we refer to it in the thesis as "content category". The listed categories varied across newspapers. We tried to group them to make them usable for further analysis. The grouping is presented below.

wirtschaft & börse	politik, inland & ausland	sport, fussball und auto
regional service wirtschaft	ausland inland medien unzuordenbar	auto
chronik lokal regional wirtschaft	politik sport unzuordenbar	regional sport
lokal wirtschaft	politik	auto sport
wirtschaft	inland ausland leben kultur medien	sport
regional wirtschaft	chronik ausland	sport fussball
service wirtschaft	inland wirtschaft	unzuordenbar sport
unzuordenbar wirtschaft	ausland unzuordenbar	auto unzuordenbar
sport wirtschaft	ausland leben	lokal sport unzuordenbar
lokal regional wirtschaft	inland unzuordenbar	chronik sport
	leserbriefe unzuordenbar politik	auto unzuordenbar sport
	ausland chronik lokal unzuordenbar	sport unzuordenbar
	meinung unzuordenbar	chronik seite seite1 sport
	ausland unzuordenbar wirtschaft	auto wirtschaft
	meinung	
	unzuordenbar politik	
	meinung leserbriefe	
	chronik inland	
	politik wirtschaft	
	chronik lokal politik	
	inland ausland	
	inland	
	leserbriefe	
	ausland	
	meinung seite titelseite	
	inland ausland wirtschaft	
	politik unzuordenbar	
	inland politik	
	ausland politik	

kultur, leben & reise	wissenschaft	schule	allgemein & thema	karriere
leben kultur medien	unzuordenbar wissenschaft	schule	allgemein	karriere
leben unzuordenbar	wissenschaft	karriere schule	kultur thema	karriere wirtschaft
medien kultur service	chronik lokal wissenschaft	schule unzuordenbar	leben thema	
kultur unzuordenbar			thema	
kultur wissenschaft			unzuordenbar thema	
chronik leben lokal unzuordenbar			inland thema	
medien kultur unzuordenbar				
kultur lokal regional				
chronik leben				
kultur unzuordenbar sport				
kultur service				
kultur medien service				
leben lokal				
unzuordenbar reise				
reise				
chronik kultur lokal				
kultur unzuordenbar seite				
kultur leben				
leben wirtschaft				
medien kultur				
kultur medien unzuordenbar				
kultur				
kultur regional				
kultur medien				
chronik kultur lokal regional				
kultur sport				
leben				
kultur leben medien				
leben regional				
leben service				
kultur reise				

List of Figures

List of Tables

Acronyms

ACDH Austrian Center for Digital Humanities. 2, 14, 18, 20, 29, 33

AMC Austrian Media Corpus. 2–4, 16, 21, 29, 31, 38, 53, 82

GB Giga Byte. 14

IE Information Extraction. 16

IWNLP Inverse Wiktionary for Natural Language Processing. 15, 23

LDA Latent Dirichlet Allocation. 8

NER Named Entity Recognition. 14, 18, 19, 27, 28, 31

NLP Natural Language Processing. 2–4, 10, 11, 13–17, 19–22, 24, 25, 35, 73, 80, 82, 109

PCA Principle Component Analysis. 7

POS Part-of-Speech. 21

RAM Random-access memory. 14

SVM Support Vector Machine. 8

Bibliography

[AAB+03] Stefanie Albert, Jan Anderssen, Regine Bader, Stephanie Becker, Tobias Bracht, Sabine Brants, Thorsten Brants, Vera Demberg, Stefanie Dipper, Peter Eisenberg, et al. Tiger Annotationsschema. 2003.

[AG17] Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. Working Paper 23089, National Bureau of Economic Research, January 2017.

[Alr21] Hend Alrasheed. Word synonym relationships for text analysis: A graph-based approach. *PLOS ONE*, 16(7):e0255127, July 2021.

[AM18] Maria Antoniak and David Mimno. Evaluating the Stability of Embedding-based Word Similarities. *Transactions of the Association for Computational Linguistics*, 6:107–119, February 2018.

[AR23] Matin N. Ashtiani and Bijan Raahemi. News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, 217:119509, May 2023.

[ATMM17] Alessio Palmero Aprosio, Sara Tonelli, Stefano Menini, and Giovanni Moretti. Using Semantic Linking to Understand Persons' Networks extracted from Text. *Frontiers in Digital Humanities*, 4:22, 2017.

[AV09] Martin Atkinson and Erik Van der Goot. Near Real Time Information Mining in Multilingual News. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 1153–1154, New York, NY, USA, 2009. ACM.

[Aza16] Julia R. Azari. How the News Media Helped to Nominate Trump. *Political Communication*, 33(4):677–680, October 2016.

[BB12] H. Russell Bernard and Harvey Russell Bernard. *Social Research Methods: Qualitative and Quantitative Approaches*. Sage, 2012.

[BBS18] Saroj Kr Biswas, Monali Bordoloi, and Jacob Shreya. A graph based keyword extraction model using collective node weight. *Expert Systems with Applications*, 97:51–59, 2018.

[BDE+04] Sabine Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen-Schirra, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith, and Hans Uszkoreit. TIGER: Linguistic interpretation of a German corpus. *Research on language and computation*, 2(4):597–620, 2004.

[Ben16] W. Lance Bennett. *News: The Politics of Illusion, Tenth Edition*. University of Chicago Press, September 2016.

[BGLL08] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008.

[BHJ09] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. *Icwsm*, 8:361–362, 2009.

[BL12] Roi Blanco and Christina Lioma. Graph-based term weighting for information retrieval. *Information Retrieval*, 15(1):54–92, February 2012.

[BMM15] Slobodan Beliga, Ana Meštrović, and Sanda Martinčić-Ipšić. An overview of graph-based keyword extraction methods and approaches. *Journal of information and organizational sciences*, 39(1):1–20, 2015.

[BNJ03] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, January 2003.

[BT17] Christiaan Burggraaff and Damian Trilling. Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*, page 1464884917716699, 2017.

[BVd16] Bjorn Burscher, Rens Vliegenthart, and Claes H. de Vreese. Frames Beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5):530–545, 2016.

[CCV+15] Esteban Castillo, Ofelia Cervantes, Darnes Vilarino, David Báez, and Alfredo Sánchez. UDLAP: Sentiment analysis using a graph-based representation. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 556–560, 2015.

[CG17] Manuela Caiani and Simona Guerra. *Euroscepticism, Democracy and the Media: Communicating Europe, Contesting Europe*. Springer, February 2017.

[CHL22] Honglin Chen, Xia Huang, and Zhiyong Li. A content analysis of Chinese news coverage on COVID-19 and tourism. *Current Issues in Tourism*, 25(2):198–205, January 2022.

[CLA+06] Silvia MG Caldeira, TC Petit Lobao, Roberto Fernandes Silva Andrade, Alexis Neme, and JG Vivas Miranda. The network of concepts in written texts. *The European Physical Journal B-Condensed Matter and Complex Systems*, 49(4):523–529, 2006.

[CN06] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5):1–9, 2006.

[CRC15] Niall J. Conroy, Victoria L. Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015.

[CRRS21] Sergio Consoli, Diego Reforgiato Recupero, and Michaela Saisana, editors. *Data Science for Economics and Finance: Methodologies and Applications*. Springer Nature, 2021.

[CS00] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

[DC05] Jana Diesner and Kathleen M Carley. Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In *Causal Mapping for Research in Information Technology*, pages 81–108. IGI Global, 2005.

[DCT12] Jana Diesner, Kathleen M. Carley, and Laurent Tambayong. Extracting socio-cultural networks of the Sudan from open-source, large-scale text data. *Computational and Mathematical Organization Theory*, 18(3):328–339, 2012.

[DG07] Loraine Daston and Peter Galison. *Objectivity*. Zone Books, New York, 2007.

[DJHM13] Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th International Conference on Semantic Systems*, pages 121–124. ACM, 2013.

[DK20] Catherine D'Ignazio and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. The MIT Press, Cambridge, Massachusetts, 2020.

[DMPR14] Matej Durco, Karlheinz Moerth, Hannes Pirker, and Jutta Ransmayr. Austrian Media Corpus 2.0. 2014.

[DSMM23] Kevin Durrheim, Maria Schuld, Martin Mafunda, and Sindisiwe Mazibuko. Using word embeddings to investigate cultural biases. *British Journal of Social Psychology*, 62(1):617–629, 2023.

[EBSW08] Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, 2008.

[EHLB23] Olga Eisele, Tobias Heidenreich, Olga Litvyak, and Hajo G. Boomgaarden. Capturing a News Frame – Comparing Machine-Learning Approaches to Frame Analysis with Different Degrees of Supervision. *Communication Methods and Measures*, 17(3):205–226, July 2023.

[Eis19] Jacob Eisenstein. *Introduction to Natural Language Processing*. MIT Press, October 2019.

[Ekm99] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, pages 45–60, 1999.

[ER04] Günes Erkan and Dragomir R Radev. Lexrank: Graph-Based lexical centrality as salience in text summarization. *journal of artificial intelligence research*, 22:457–479, 2004.

[FAL+13] Ilias Flaounas, Omar Ali, Thomas Lansdall-Welfare, Tijl De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. Research Methods in the Age of Digital Journalism. *Digital Journalism*, 1(1):102–116, February 2013.

[FAT+11] Ilias Flaounas, Omar Ali, Marco Turchi, Tristan Snowsill, Florent Nicart, Tijl De Bie, and Nello Cristianini. NOAM: News Outlets Analysis and Monitoring System. In *2011 ACM SIGMOD International Conference on Management of Data, Proceedings*, pages 1275–1277. Association for Computing Machinery (ACM), 2011.

[FMS+20] Fabian Falck, Julian Marstaller, Niklas Stoehr, Sören Maucher, Jeana Ren, Andreas Thalhammer, Achim Rettinger, and Rudi Studer. Measuring Proximity Between Newspapers and Political Parties: The Sentiment Political Compass. *Policy & Internet*, 12(3):367–399, 2020.

[Fre82] Linton C. Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304, October 1982.

[Gar21] Muskan Garg. A survey on different dimensions for graphical keyword extraction techniques. *Artificial Intelligence Review*, 54(6):4731–4770, August 2021.

[GB17] Esther Greussing and Hajo G. Boomgaarden. Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis. *Journal of Ethnic and Migration Studies*, 43(11):1749–1774, 2017.

[Gil14] Tarleton Gillespie. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, 167, 2014.

[GJH17] Gabriel Grill, Julia Neidhardt, and Hannes Werthner. Network analysis on the austrian media corpus. In *VSS 2017-Vienna Young Scientists Symposium*, pages 128–129, 2017.

[GMK+18] Manisha Gahirwal, Sanjana Moghe, Tanvi Kulkarni, Devansh Khakhar, and Jayesh Bhatia. Fake News Detection. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(1):817–819, March 2018.

[GOB+12] Brynjar Gretarsson, John O'donovan, Svetlin Bostandjiev, Tobias Höllerer, Arthur Asuncion, David Newman, and Padhraic Smyth. Topicnets: Visual analysis of large text corpora with topic modeling. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(2):23, 2012.

[GP15] Marta Gruszecka and Michal Pikusa. Using Text Network Analysis in Corpus Studies–A Comparative Study on the 2010 TU-154 Polish Air Force Presidential Plane Crash Newspaper Coverage. *International Journal of Social Science and Humanity*, 5(2):233, 2015.

[GSG17] Johanna Geiß, Andreas Spitz, and Michael Gertz. NECKAr: A named entity classifier for Wikidata. In *International Conference of the German Society for Computational Linguistics and Language Technology*, pages 115–129. Springer, 2017.

[Hab91] Jurgen Habermas. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press, August 1991.

[Ham23] Felix Hamborg. *Revealing Media Bias in News Articles: NLP Techniques for Automated Frame Analysis*. Springer Nature, 2023.

[HDG19] Felix Hamborg, Karsten Donnay, and Bela Gipp. Automated identification of media bias in news articles: An interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415, December 2019.

[Hen04] Svetlana Hensman. Construction of conceptual graph representation of texts. In *Proceedings of the Student Research Workshop at HLT-NAACL 2004*, pages 49–54. Association for Computational Linguistics, 2004.

[HFC12] Elena Hensinger, Ilias Flaounas, and Nello Cristianini. Modelling and predicting news popularity. *Pattern Analysis and Applications*, 16(4):623–635, December 2012.

[HH16] Kristoffer Holt and André Haller. The Populist Communication Paradox of PEGDIA : Between "Lying Press" and Journalistic Sources. 2016.

[HJ15] Matthew Honnibal and Mark Johnson. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, 2015.

[HJ17] Martin Haselmayer and Marcelo Jenny. Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & quantity*, 51(6):2623–2646, 2017.

[HM17] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 2017.

[HMPR04] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design Science in Information Systems Research. *MIS Q.*, 28(1):75–105, March 2004.

[HMW14] Johannes Hoffart, Dragan Milchevski, and Gerhard Weikum. AESTHETICS: Analytics with Strings, Things, and Cats. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 2018–2020, New York, NY, USA, 2014. ACM.

[HNA23] Mohammed Haqi Al-Tai, Bashar M. Nema, and Ali Al-Sherbaz. Deep Learning for Fake News Detection: Literature Review. *Al-Mustansiriyah Journal of Science*, 34(2):70–81, June 2023.

[HNI21] Sameera Horawalavithana, Kin Wai Ng, and Adriana Iamnitchi. Drivers of Polarized Discussions on Twitter during Venezuela Political Crisis. In *Proceedings of the 13th ACM Web Science Conference 2021*, WebSci '21, pages 205–214, New York, NY, USA, June 2021. Association for Computing Machinery.

[HPSE15] Stefan Heindorf, Martin Potthast, Benno Stein, and Gregor Engels. Towards vandalism detection in knowledge bases: Corpus construction and analysis. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 831–834. ACM, 2015.

[HSSC08] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[Hun07] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[iSK04] Ramon Ferrer i Cancho, Ricard V. Solé, and Reinhard Köhler. Patterns in syntactic dependency networks. *Physical Review E*, 69(5):051915, 2004.

[JBL16] Ridong Jiang, Rafael E Banchs, and Haizhou Li. Evaluating and Combining Named Entity Recognition Systems. *ACL 2016*, page 21, 2016.

[Jen20] Klaus Bruhn Jensen. *A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies.* Routledge, December 2020.

[JLS+16] Sen Jia, Thomas Lansdall-Welfare, Saatviga Sudhahar, Cynthia Carter, and Nello Cristianini. Women Are Seen More than Heard in Online Newspapers. *PLOS ONE*, 11(2):e0148434, February 2016.

[JM09] Daniel Jurafsky and James H. Martin. *Speech and Language Processing (2Nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009.

[Jol86] Ian T. Jolliffe. Principal component analysis and factor analysis. In *Principal Component Analysis*, pages 115–128. Springer, 1986.

[JOP01] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: Open source scientific tools for Python, 2001.

[JS07] Wei Jin and Rohini K. Srihari. Graph-based text representation and knowledge discovery. In *Proceedings of the 2007 ACM Symposium on Applied Computing*, pages 807–811. ACM, 2007.

[JVHB14] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, 9(6):e98679, June 2014.

[JvW16] Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1):89–106, January 2016.

[Kel16] Douglas Kellner. *American Nightmare: Donald Trump, Media Spectacle, and Authoritarian Populism*. Springer, November 2016.

[KMK16] Muhammad Tahir Khan, Yukun Ma, and Jung-jae Kim. Term Ranker: A Graph-Based Re-Ranking Approach. In *FLAIRS Conference*, pages 310–315, 2016.

[KNMN23] Brian Felipe Keith Norambuena, Tanushree Mitra, and Chris North. A Survey on Event-Based News Narrative Extraction. *ACM Computing Surveys*, 55(14s):300:1–300:39, July 2023.

[Kon14] Natalia Konstantinova. Review of Relation Extraction Methods: What Is New Out There? In *Analysis of Images, Social Networks and Texts*, Communications in Computer and Information Science, pages 15–28. Springer, Cham, April 2014.

[KRK20] Thomas Kellermeier, Tim Repke, and Ralf Krestel. Mining Business Relationships from Stocks and News. In Valerio Bitetta, Ilaria Bordino, Andrea Ferretti, Francesco Gullo, Stefano Pascolutti, and Giovanni Ponti, editors, *Mining Data for Financial Applications*, Lecture Notes in Computer Science, pages 70–84, Cham, 2020. Springer International Publishing.

[KRP+16] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E. Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B. Hamrick, Jason Grout, and Sylvain Corlay. Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90, 2016.

[KSR16] Roman Klinger, Surayya Samat Suliya, and Nils Reiter. Automatic Emotion Detection for Quantitative Literary Studies – A case study based on Franz Kafka's "Das Schloss" and "Amerika". In *Digital Humanities 2016: Conference Abstracts*, pages 826–828, Kraków, Poland, July 2016. Jagiellonian University and Pedagogical University.

[KTR21] Anne C. Kroon, Damian Trilling, and Tamara Raats. Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage. *Journalism & Mass Communication Quarterly*, 98(2):451–477, June 2021.

[LC15] Matthias Liebeck and Stefan Conrad. IWNLP: Inverse Wiktionary for Natural Language Processing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 414–418. Association for Computational Linguistics, 2015.

[LKS05] Levon Lloyd, Dimitrios Kechagias, and Steven Skiena. Lydia: A System for Large-Scale News Analysis. In *String Processing and Information Retrieval*, Lecture Notes in Computer Science, pages 161–166. Springer, Berlin, Heidelberg, November 2005.

[LS13] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention*, volume 2, pages 1–49. Citeseer, 2013.

[LSTC17] Thomas Lansdall-Welfare, Saatviga Sudhahar, James Thompson, and Nello Cristianini. The Actors of History: Narrative Network Analysis Reveals the Institutions of Power in British Society Between 1800-1950. In *International Symposium on Intelligent Data Analysis*, pages 186–197. Springer, 2017.

[MA11] K. J. Millman and M. Aivazis. Python for Scientists and Engineers. *Computing in Science Engineering*, 13(2):9–12, March 2011.

[MA23] Shawn McCarthy and Gita Alaghband. Enhancing Financial Market Analysis and Prediction with Emotion Corpora and News Co-Occurrence Network. *Journal of Risk and Financial Management*, 16(4):226, April 2023.

[McK10] Wes McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[MM11] Gautam Mitra and Leela Mitra. *The Handbook of News Analytics in Finance*. John Wiley & Sons, July 2011.

[Moo22] Paul Mooney. 2022 Kaggle Machine Learning & Data Science Survey, 2022.

[MPS14] Jan Michelfeit, Jan Pomikálek, and Vít Suchomel. Text tokenisation using unitok. In *8th Workshop on Recent Advances in Slavonic Natural Language Processing, Brno, Tribun EU*, pages 71–75, 2014.

[MPY22] Milagros Miceli, Julian Posada, and Tianling Yang. Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–14, 2022.

[MSA22] Shubha Mishra, Piyush Shukla, and Ratish Agarwal. Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets. *Wireless Communications and Mobile Computing*, 2022:e1575365, March 2022.

[MSB+14] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014.

[MT04] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004.

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, January 2003.

[NS07] David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Lingvisticæ Investigationes*, 30(1):3–26, January 2007.

[NSZ+15] Vlad Niculae, Caroline Suen, Justine Zhang, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. QUOTUS: The Structure of Political Media Coverage as Revealed by Quoting Patterns. *arXiv:1504.01383 [physics]*, April 2015.

[OAO22] Marc Gallofré Ocaña, Tareq Al-Moslmi, and Andreas L. Opdahl. Knowledge Graph Semantic Annotation and Population with Real-Time Events Data from GDELT. In *2022 IEEE 24th Conference on Business Informatics (CBI)*, volume 02, pages 65–72, June 2022.

[OB20] Ahmed Hamza Osman and Omar Mohammed Barukub. Graph-Based Text Representation and Matching: A Review of the State of the Art and Future Challenges. *IEEE Access*, 8:87562–87583, 2020.

[OBY98] Yukio Ohsawa, Nels E. Benson, and Masahiko Yachida. KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In *Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum On*, pages 12–18. IEEE, 1998.

[ÖCB08] Arzucan Özgür, Burak Cetin, and Haluk Bingol. Co-occurrence network of reuters news. *International Journal of Modern Physics C*, 19(05):689–702, 2008.

[Oli06] Travis E. Oliphant. *A Guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[Oli07] T. E. Oliphant. Python for Scientific Computing. *Computing in Science Engineering*, 9(3):10–20, May 2007.

[Par11] Dmitry Paranyushkin. Identifying the pathways for meaning circulation using text network analysis. *Berlin: Nodus Labs. Retrived at: http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis*, 2011.

[Par19] Dmitry Paranyushkin. InfraNodus: Generating Insight Using Text Network Analysis. In *The World Wide Web Conference*, WWW '19, pages 3584–3589, New York, NY, USA, May 2019. Association for Computing Machinery.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[PDM12] Slav Petrov, Dipanjan Das, and Ryan McDonald. A universal part-of-speech tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2089–2096, 2012.

[PG07] Fernando Perez and Brian E. Granger. IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, 9(3):21–29, 2007.

[Pic16] Robert G Picard. Are journalists reporting—or publicising—populism? *European Journalism Observatory*, 28:2016, 2016.

[PK15] Kasun Perera and Damitha Karunarathne. KeyGraph and WordNet hypernyms for topic detection. In *Computer Science and Software Engineering (JCSSE), 2015 12th International Joint Conference On*, pages 303–308. IEEE, 2015.

[PNPV23] Phu Pham, Loan T. T. Nguyen, Witold Pedrycz, and Bay Vo. Deep learning, graph-based text representation and classification: A survey, perspectives and challenges. *Artificial Intelligence Review*, 56(6):4893–4927, June 2023.

[PU16] Thomas Proisl and Peter Uhrig. SoMaJo: State-of-the-Art tokenization for German web and social media texts. In *Proceedings of the 10th Web as Corpus Workshop*, pages 57–62, 2016.

[RQH10] Robert Remus, Uwe Quasthoff, and Gerhard Heyer. SentiWS-A Publicly Available German-Language Resource for Sentiment Analysis. In *LREC*, 2010.

[RS16] Samuel Rönnqvist and Peter Sarlin. Bank networks from text: Interrelations, centrality and determinants. Working Paper Series 1876, European Central Bank, 2016.

[SAG17] Andreas Spitz, Satya Almasian, and Michael Gertz. EVELIN: Exploration of Event and Entity Links in Implicit Networks. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, pages 273–277, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee.

[Sch95] Anne Schiller. Vorlaufige Guidelines fur das Tagging deutscher Textcorpora mit STTS. In *Seminar Fur Sprachwissenschaft, 1995*, 1995.

[Scu10] David Sculley. Web-scale k-Means clustering. In *Proceedings of the 19th International Conference on World Wide Web*, pages 1177–1178. ACM, 2010.

[SDFFC15] Saatviga Sudhahar, Gianluca De Fazio, Roberto Franzosi, and Nello Cristianini. Network analysis of narrative content in large corpora. *Natural Language Engineering*, 21(01):81–112, January 2015.

[Sea16] Jean Seaton. Brexit and the Media. *The Political Quarterly*, 87(3):333–337, July 2016.

[Seg21] Elad Segev. *Semantic Network Analysis in Social Sciences*. Routledge, November 2021.

[SERS22] Julian Schelb, Maud Ehrmann, Matteo Romanello, and Andreas Spitz. ECCE: Entity-centric Corpus Exploration Using Contextual Implicit Networks. In *Companion Proceedings of the Web Conference 2022*, WWW '22, pages 278–281, New York, NY, USA, August 2022. Association for Computing Machinery.

[SG15] Andreas Spitz and Michael Gertz. Breaking the News: Extracting the Sparse Citation Network Backbone of Online News Articles. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 274–279, New York, NY, USA, 2015. ACM.

[SG18] Andreas Spitz and Michael Gertz. Exploring Entity-Centric Networks in Entangled News Streams. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 555–563, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.

[SHB21] Fynn Schröder, Hans Ole Hatzel, and Chris Biemann. Neural end-to-end coreference resolution for German in different domains. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pages 170–181, 2021.

[SK14] Sheetal S. Sonawane and Parag A. Kulkarni. Graph based representation and analysis of text document: A survey of techniques. *International Journal of Computer Applications*, 96(19), 2014.

[SKH+21] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Kumar Paritosh, and Lora Mois Aroyo. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[SO21] Geoffrey Supran and Naomi Oreskes. Rhetoric and frame analysis of ExxonMobil's climate change communications. *One Earth*, 4(5):696–719, May 2021.

[SPW15] Junseop Shim, Chisung Park, and Mark Wilding. Identifying policy frames through semantic network analysis: An examination of nuclear energy policy across six countries. *Policy Sciences*, 48(1):51–83, 2015.

[SR13] Hassan Sayyadi and Louiqa Raschid. A graph analytical approach for topic detection. *ACM Transactions on Internet Technology (TOIT)*, 13(2):4, 2013.

[SSG18] Andreas Spitz, Jannik Strötgen, and Michael Gertz. Predicting Document Creation Times in News Citation Networks. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 1731–1736, Republic and Canton of Geneva, CHE, April 2018. International World Wide Web Conferences Steering Committee.

[SSVW09] Rico Sennrich, Gerold Schneider, Martin Volk, and Martin Warin. A new hybrid dependency parser for German. *Proceedings of the German Society for Computational Linguistics and Language Technology*, pages 115–124, 2009.

[SVC15] Saatviga Sudhahar, Giuseppe A. Veltri, and Nello Cristianini. Automated analysis of the US presidential elections using Big Data and network analysis. *Big Data & Society*, 2(1):2053951715572916, February 2015.

[SVRS22] Alfonso Semeraro, Salvatore Vilella, Giancarlo Ruffo, and Massimo Stella. Emotional profiling and cognitive networks unravel how mainstream and alternative press framed AstraZeneca, Pfizer and COVID-19 vaccination campaigns. *Scientific Reports*, 12(1):14445, August 2022.

[SVS13] Rico Sennrich, Martin Volk, and Gerold Schneider. Exploiting synergies between open resources for german dependency parsing, pos-tagging, and morphological analysis. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 601–609, 2013.

[TFA16] Chenhao Tan, Adrien Friggeri, and Lada Adamic. Lost in Propagation? Unfolding News Cycles from the Source. In *Tenth International AAAI Conference on Web and Social Media*, March 2016.

[THH20] Yassin Taskin, Tobias Hecking, and H. Ulrich Hoppe. ESA-T2N: A Novel Approach to Network-Text Analysis. In Hocine Cherifi, Sabrina Gaito, José Fernendo Mendes, Esteban Moro, and Luis Mateus Rocha, editors, *Complex Networks and Their Applications VIII*, Studies in Computational Intelligence, pages 129–139, Cham, 2020. Springer International Publishing.

[Tri13] Craig Trim. The Art of Tokenization (Language Processing), 2013.

[Tri23] Francesca Tripodi. Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 25(7):1687–1707, July 2023.

[TRv15] Vincent A. Traag, Ridho Reinanda, and Gerry van Klinken. Elite co-occurrence in the media. *Asian Journal of Social Science*, 43(5):588–612, 2015.

[TRv16] Vincent A. Traag, Ridho Reinanda, and Gerry van Klinken. Structure of a media co-occurrence network. In *Proceedings of ECCS 2014*, pages 81–91. Springer, 2016.

[Tug16] Don Tuggener. *Incremental Coreference Resolution for German*. PhD thesis, 2016.

[VBJG16] Tommaso Venturini, Liliana Bounegru, Mathieu Jacomy, and Jonathan Gray. How to Tell Stories with Networks: Exploring the Narrative Affordances of Graphs with the Iliad. In *Datafied Society: Social Research in the Age of Big Data*. Amsterdam University Press, 2016.

[vGR13] Maarten van Gompel and Martin Reynaert. FoLiA: A practical XML format for linguistic annotation – a descriptive and comparative study. *Computational Linguistics in the Netherlands Journal*, 3:63–81, December 2013.

[VK14] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

[VPM+20] Vasiliki Voukelatou, Luca Pappalardo, Ioanna Miliou, Lorenzo Gabrielli, and Fosca Giannotti. Estimating countries' peace index through the lens of the world news as monitored by GDELT. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 216–225, October 2020.

[Wan17] William Yang Wang. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 422–426, May 2017.

[Web12] Jim Webber. A programmatic introduction to neo4j. In *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, pages 217–218. ACM, 2012.

[WF94] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*, volume 8. Cambridge university press, 1994.

[WF22] Stefanie Walter and Zoltán Fazekas. Similar citizen portrayals? Converging media-based othering in tabloids and broadsheets. *Journalism*, page 14648849221116204, August 2022.

[WK92] Jonathan J. Webster and Chunyu Kit. Tokenization as the Initial Phase in NLP. In *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*, 1992.

[WSP+16] Chris Wells, Dhavan V. Shah, Jon C. Pevehouse, JungHwan Yang, Ayellet Pelled, Frederick Boehm, Josephine Lukito, Shreenita Ghosh, and Jessica L. Schmidt. How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning. *Political Communication*, 33(4):669–676, October 2016.

[YKY08] Jen-Yuan Yeh, Hao-Ren Ke, and Wei-Pang Yang. iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network. *Expert Systems with Applications*, 35(3):1451–1462, 2008.

[YSJ14] C. S. Yadav, A. Sharan, and M. L. Joshi. Semantic graph based approach for text mining. In *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, pages 596–601, February 2014.

[YSZ+18] Shanliang Yang, Qi Sun, Huyong Zhou, Zhengjie Gong, Yangzhi Zhou, and Junhong Huang. A Topic Detection Method Based on KeyGraph and Community Partition. In *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*, ICCAI 2018, pages 30–34, New York, NY, USA, 2018. ACM.

[YTT15] Ikuya Yamada, Hideaki Takeda, and Yoshiyasu Takefuji. An end-to-end entity linking approach for tweets. In *5th Workshop on Making Sense of Microposts: Big Things Come in Small Packages,# Microposts 2015, at the 24th International Conference on the World Wide Web, WWW 2015*. CEUR-WS, 2015.

[ZCF+10] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95, 2010.

[ZSLZ22] Ning Zhou, Wenqian Shi, Renyu Liang, and Na Zhong. TextRank Keyword Extraction Algorithm Using Word Vector Clustering Based on Rough Data-Deduction. *Computational Intelligence and Neuroscience*, 2022:5649994, January 2022.

[ZZ20] Xinyi Zhou and Reza Zafarani. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, 53(5):109:1–109:40, September 2020.