# Measuring Polarization in an Online News Forum

## DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieur**

im Rahmen des Studiums

**Business Informatics**

eingereicht von

**Patrick Oliver Riemer, BSc.**
Matrikelnummer 01452119

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Dipl.-Ing. Dr.techn. Hannes Werthner
Mitwirkung: Univ.Ass. Mag.rer.nat. Dr.techn. Julia Neidhardt

Wien, 15. April 2021

Patrick Oliver Riemer | Hannes Werthner

TU WIEN Informatics

# Measuring Polarization in an Online News Forum

## DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

**Diplom-Ingenieur**

in

**Business Informatics**

by

**Patrick Oliver Riemer, BSc.**
Registration Number 01452119

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Hannes Werthner
Assistance: Univ.Ass. Mag.rer.nat. Dr.techn. Julia Neidhardt

Vienna, 15th April, 2021

Patrick Oliver Riemer | Hannes Werthner

# Erklärung zur Verfassung der Arbeit

Patrick Oliver Riemer, 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, 15. April 2021

Patrick Oliver Riemer

# Danksagung

Ich danke meinen Betreuen Dr. Julia Neidhardt und Prof. Hannes Werthner für die Betreuung dieser Arbeit und für ihr Feedback und ihre Expertise. Dr. Neidhardt hat in einem Vortrag zum Thema Soziale Netzwerkanalyse mein Interesse an diesem Forschungsgebiet geweckt und mich zu einer Diplomarbeit auf diesem Gebiet inspiriert.

Darüber hinaus möchte ich allen danken, die mich unterstützt haben, insbesondere meine Eltern, Freunde und Kollegen. Außerdem möchte ich meinem Freund Jörg dafür danken, dass er mich unermüdlich, seit dem ersten Tag, daran erinnert hat diese Arbeit fertigzustellen.

Abschließend möchte ich meiner Freundin Cathrin meinen tiefsten Dank aussprechen. Du hast immer dafür gesorgt, dass ich trotz der vielen anderen Projekte und Vorhaben nie den Fokus verloren habe. Danke für dein Verständnis, deinen Rat und deine Geduld.

# Acknowledgements

I want to thank my advisors Dr. Julia Neidhardt and Prof. Hannes Werthner, for supervising this thesis and for their feedback and expertise. When holding a lecture about Social Network Analysis Dr. Neidhardt raised my interest in the research field and inspired me to write a thesis in this area even though I already started to write a thesis in a different field.

Furthermore, I want to thank everybody who supported me, especially my parents, friends, and colleagues. Moreover, I want to thank my friend Jörg for relentlessly reminding me to finish my thesis every day since the day I started writing on it.

Finally, I want to express my deepest gratitude to my girlfriend Cathrin. You have always supported and motivated me since we know each other. You made sure that university and this thesis are always the top priority despite all the other projects and endeavours coming across my way during these years and I am glad to return this favour.

# Kurzfassung


Die Kontroversen um die US-Präsidentschaftswahlen 2020 haben gezeigt, wie polarisiert unsere Gesellschaft wirklich ist. Menschen sind in ihren ganz eigenen Filterblasen gefangen, die durch Empfehlungsalgorithmen von Facebook und Twitter erzeugt werden und sie anfälliger für Manipulation im Allgemeinen machen. Um diesen nachteiligen Auswirkungen der Polarisierung entgegenzuwirken, ist es notwendig, in einem ersten Schritt die Polarisierung in diesen Netzwerken zu identifizieren, zu analysieren und zu quantifizieren. Wir konzentrieren uns jedoch nicht auf soziale Netzwerke, da sie bereits vielfach Forschungsgegenstand waren, sondern auf Online-Nachrichtenforen, da sie für die Verbreitung von Informationen ebenso entscheidend sind. Hierfür verwenden wir die Daten des größten österreichischen Online-Nachrichtenforums “derStandard.at”.

Obwohl in der Literatur eine Vielzahl von Methoden zur Quantifizierung von Polarisierung veröffentlicht wurde, existiert bis heute kein umfassender Vergleich dieser Methoden. Wir identifizieren daher geeignete Methoden zur Erkennung und Quantifizierung von Polarisierung in sozialen Netzwerken und vergleichen und bewerten diese in Bezug auf Korrektheit und Leistungsfähigkeit. Hier zeichnet sich *Boundary Connectivity* durch Korrektheit bei gleichzeitig bester Laufzeit-Performance als gemeinsamer Nenner aus.

Außerdem evaluieren wir verschiedene Möglichkeiten, um Netzwerke in einem Online-Nachrichtenforum zu extrahieren und die Polarisierung zu messen. Wir zeigen, dass populäre und kontroverse Inhalte nicht unbedingt polarisiert sein müssen. Darüber hinaus berechnen wir die Polarisierung von 5.000 Artikeln, wodurch sich zeigt, dass einige Messmethoden, insbesondere die häufig verwendete *Modularität*, in diesem Fall keine konsistenten Ergebnisse liefern. Die Erkenntnisse, die wir aus der Analyse der Polarisierung gewonnen haben, bestätigen unsere Vermutung, dass Methoden aus der Social Network Analysis sehr von der Anwendung inhaltsbasierter Methoden, wie Natural Language Processing, profitieren könnten um ein besseres Verständnis der Inhalte selbst zu erhalten, anstatt sich nur auf strukturelle Eigenschaften zu verlassen.

Darüber hinaus analysieren wir die Polarisierung ganzer Themen, d. h. einzelner Artikel die sich auf dasselbe Thema beziehen. Wir definieren kontextunabhängige Einschränkungen für diese zusammengesetzten Netzwerke, die erfüllt sein müssen um Verzerrungen bei der Messung der Polarisierung zu minimieren. Darüber hinaus haben wir auch die Polarisierung auf Basis der Beziehungen zwischen Benutzern analysiert, d. h. Benutzer die sich gegenseitig folgen oder ignorieren, was sich als nicht polarisiert herausstellte.

# Abstract

The controversies around the 2020 U.S. presidential election have proven how polarized and separated our society really is. People are imprisoned in their very own echo chambers and filter bubbles, created by the recommendation algorithms of Facebook and Twitter, making people more vulnerable to manipulation in general. To counter these adversarial effects of polarization, it is necessary to identify, analyse and quantify polarization in social networks. We do not focus on social media networks however, as they have already been subject to relevant research numerous times, but on online news forums, as they are just as critical for spreading information. For this, we use the data from the largest Austrian online news forum "derStandard.at".

Although a broad range of methods has been published in literature to quantify polarization, no extensive comparison exists to this day. We therefore identify suitable methodologies to detect and quantify polarization in social networks and compare and evaluate them in terms of correctness and performance. Here, *Boundary Connectivity* stands out as a common denominator, being able to quantify polarization correctly while also showing the best run-time performance.

We also describe and evaluate different possibilities to extract networks and measure polarization in an online news forum. We show that popular and controversial content must not necessarily be highly polarized. Furthermore, we calculate the polarization of 5,000 articles which enables us to get a more in-depth insight into the various polarization measurements, of which some fail to produce consistent results. The frequently used *Modularity* measure for instance fails to reliably quantify polarization. The insights we gained from analysing polarization based on user-generated comments, confirms our assumption, that methods from Social Networks Analysis could benefit greatly from applying content-based methods such as Natural Language Processing, to get a better understanding of the content itself, instead of solely relying on structural properties.

Moreover, we analyze the polarization of whole topics, i.e. single articles related to the same subject. We define context-independent restrictions for these composed networks that must be fulfilled to minimize bias when measuring the polarization. We also analysed polarization based on the relationships between users, i.e. users following or ignoring each other, which turned out as not being polarised.

# Contents

CHAPTER 1

# Introduction

Undeniable the rise of *Facebook*, *Twitter* and other social media platforms have changed the way people communicate. They connect people all across the globe, regardless of nationality, social status or beliefs. And precisely this point is discussed quite intensively in literature. Never before in history has it been easier to access that much information and high-quality data. Although we have almost limitless access to diverse sources of information, it has become astonishingly easy and common to restrict ourselves to social circles sharing our own beliefs and point of views.

This phenomenon has become famous as *echo chambers* in news and literature. People "hear their own voice - an echo", i.e. in the context of social media, people mostly consume content matching their own point of view. This may lead to people exclusively consuming content, news and opinions from content creators compliant with their own particular point of view.

Various studies show that networks in social media become increasingly polarized. The adverse effects on our daily lives are not to be neglected. Polarized networks are much more likely to be manipulated by wrong information, which proved itself as increasingly dangerous in the last couple of years. In politics, polarization leads to more extreme viewpoints and intolerance to opposing ones, which might be exploited by certain political parties. All these adverse effects became evident in a number of real-world events with great significance, such as the U.S. presidential elections, the Indian elections, or the Brexit vote, where polarization, fake news and propaganda had a significant impact [Gar18, BF16, AG05].

## 1.1 Motivation and problem definition

Due to the increasing importance, relevant research concentrates on how to mitigate the adverse effects, and how to decrease polarization of social networks [GDFMGM17, MMT18]. Before it is possible to mitigate any adverse effects, though, it is necessary to identify polarized topics somehow, which in turn makes it necessary to define which conditions must be fulfilled in a network to consider it polarized. Dedicated polarization measurements identify polarized networks and quantify the strength of the polarization. This makes it possible to identify topics which are strongly or even extremely polarized, as these are the ones exposing a range of respective adverse effects caused by polarization.

However subject of measurement is not a social network itself, at least not directly, but the various graph representations of it, being used as proxies for analysis. Together with different possible dimensions inherent in the data of social media networks, such as followers and friends, likes and votes and comments and postings, different ways to construct graphs must be considered. These different kinds of data also encode different semantics and network structures.

Moreover, the scope of analysis is essential. A network in online news forums like *Der Standard* consists of several categories, which in turn span over lots of different topics, whereas a topic again is subject to several discussions, which again consists of a number of sub-discussions. It is possible to construct networks on every level of this hierarchy, which probably yields different polarization measures in the end. Though articles and the comments underneath them are not the only possibilities, as users can also up- or down-vote other users' comments. They can also follow others or ignore them, depending on their mutual relationship. This leads to the following research question:

**RQ1:** What are appropriate ways to construct networks to measure their polarization based on user interaction data in online news forums?

Although there is a number of measurements proposed in literature, the question arises how to choose among them, and if there might be the one best way to measure polarization. *Modularity* is by far the most common measure used in literature. Though there are many more to choose from, with different proposed advantages and use cases. Leading to the following research question:

**RQ2:** What are appropriate ways to measure the polarization of discussions in online news forums?

## 1.2 Methodology

The methodological approach consists of a set of planned steps, which will be realised by applying different research methods. The steps, as described in the following, are performed iteratively, continuously incorporating findings and outcomes for further refinement:

1. **Literature review**

   To get an extensive overview of the theoretical concepts related to our research, we conduct a literature review. As our research is split into two parts, network construction and measuring polarization, this literature review also consists of two parts, each focused on one of these areas. Garimella et al. 2015 [GMGM15], a comparison of different polarization measures, is the starting point of our review.

We recursively search for definitions of other polarization measures, or applied usages of such, using the citations and references of the relevant papers found previously. We justify this approach based on the rationale that, to define a new polarization measure, the authors must compare it to existing ones, or at least cite other measures. This also holds true for applications of the polarization measures found, as otherwise, the authors would violate the requirement for scientific rigour. This approach also does not require the complex search and querying of multiple databases, as our search is not based on keywords, but solely on the connections between papers through citations or references [Pau13, CRC08].

To systematically keep track of all findings in an organised way, we use the free reference management software Zotero[1] [Bas14]. Although we aim to achieve an exhaustive overview of polarization measures in literature, we refrain from conducting a systematic literature review, as following all steps and guidelines would go beyond the scope of this thesis and we are confident in the feasibility of our approach as described above [Kit04, KPBB$^+$09].

Following this approach for getting an overview of measures defined in literature, we simultaneously do not only get insights into the approaches and experiment designs used to analyse polarization, but also ways to construct networks and graphs out of extracted data for further analysis. We consider this exhaustive, as every paper concerned with network extraction, network construction and measuring polarization would reference one of the papers, defining a polarization measure, we found earlier. Therefore only one literature review is needed to achieve an extensive overview of polarization measures and network construction methodologies in literature [AKR14].

2. **Measuring polarization**

   Directly assessing all combinations of the constructed networks in a first step, with all measurements found, would exceed the scope of our final evaluation. Therefore, to limit the number of measurements beforehand, we evaluate them using publicly available social networks, such as *Zachary's karate club* [Zac77] or *2004 U.S. Political Blogosphere* [AG05], which has been used previously and frequently in literature to assess ways to measure polarization. For the verification of our implementation, we will then compare our results for these datasets with results from literature. Moreover, we discuss the characteristics, advantages and trade-offs of the various measures. We try to deduce general trends and insights and compare them also in terms of run-time and reliability - in connection with the different network structures available.

[1] https://www.zotero.org/

3. **Network construction**

   Building on the theoretical basis gained from the initial literature review, we will extract various different networks from *Der Standard*. We will examine different scopes of analysis from only one single article, up to several articles forming a topic. We will also assign different semantics to the elements of the networks, using relationships between users, comments or votes. Therefore we expect a number of different networks in the end, which we will assess in terms of the information, communication structures and polarization they expose.

4. **Polarization measurement workflow**

   Applying our findings to analyse the polarization of networks extracted from data provided by *Der Standard* is partly a data mining problem, as we have to find and extract relevant data from our datasets to construct networks from it to finally measure the polarization. Therefore our workflow is based on selected ideas and principles of data mining processes such as *Fayyad's KDD process* [Fay96] or *CRISP-DM* [WH00].

   Our objective is the creation of a semi-automatic workflow, which selects and extracts data, transforms it into graphs followed by subsequent steps necessary to result in a quantitative assessment about the polarization of the underlying data, which is also partly inspired by findings in literature [Gar18].

   Data selection is the first and most crucial step in our workflow, as the data (e.g. a topic, keyword, article, or similar) selected in this stage, is the basis for all the following steps. Therefore we are pursuing different strategies to select relevant content. On one side we select content by popularity, i.e. articles and keywords with the most comments and votes, but also by manual selection of topics which are known to be controversial and polarized, like politics or migration. To mitigate the risk of any bias, we also talk to the community management of *derStandard.at* about their thoughts and expertise as nobody is more likely to judge which content tends to be polarized, as the ones manually moderating the forum every day.

5. **Evaluation**

   - Polarization Measures

     We evaluate the polarization measures using selected networks extracted from the available data from *Der Standard*. Through qualitative assessment of the polarization of the networks, we form a ground truth. We test the different measurements against this ground truth, if they correlate with the perceived polarization and how they perform compared to the *Modularity* measure, which acts as the second *gold standard* in this case.

   - Qualitative Evaluation

     The outcomes are then interpreted qualitatively by the authors. The objective here is to interpret and assess the obtained results, especially the combinations found in the previous evaluation stage.

## 1.3 Structure of the work

The structure of this thesis is as follows: in Chapter 2, we discuss the theoretical background of methods and measurements we used in this thesis and related work. This includes the definition of polarization and controversy and their connection to social network analysis. We provide an overview and definitions for all relevant polarization measurements and an overview of their usage and references in literature. We complete the chapter with an overview of content-based methods and various polarization indicators, as well as relevant experiment approaches in literature to quantify polarization in social networks.

In Chapter 3, we outline our experiment design. We briefly discuss our implementation of the various polarization measurements and how they are compared using synthetic datasets and ones from literature. We proceed with an analysis of the data we use in our experiments, providing relevant statistics and explanations. We continue by explaining our data selection procedure, to limit the data available to relevant and manageable parts, and how we transform them into graphs or networks. The chapter is then closed with an overview of the workflow and its single steps to select data, construct and partition the network, and quantify the polarization for that network.

Chapter 4 provides our evaluation of different polarization measurements and a thorough analysis of our results. The evaluated polarization measurements are compared in terms of performance, i.e. the polarization score they deliver, standard deviation, their run time performance and a comparison to the results obtained in previous studies.

Chapter 5 provides a detailed analysis and overview of different ways to extract and transform data from online news forums to measure polarization. We analyse networks constructed from postings belonging to a single article, but also networks composed from the data of several articles of the same topic. Furthermore we analyse polarization using *follow* and *ignore* user relationships and end the chapter with a qualitative assessment of our results.

Chapter 6 concludes this thesis and we present ideas and starting points for future research and improvements to our work in Chapter 7.

# CHAPTER 2

# Related work

This chapter describes the theoretical foundations of the concepts and techniques used in this thesis. We describe the concept of polarization in section 2.1 and what differentiates it from controversy in section 2.2. We continue with explaining how polarization can be analysed using social network analysis in section 2.3. Several approaches to measure polarization were proposed in the past, which we will outline in section 2.4 and analyse their usages in literature in section 2.5. In the sections 2.6 and 2.7 we describe additional methods to analyse polarization using content based methods or to simply detect polarized discussions. In section 2.8 we will present approaches to analyse the polarization in social media used in the past, which served as motivation for our experiment design.

## 2.1 Polarization

Polarization, as defined by the Oxford Dictionary[1], is the "division into two sharply contrasting groups or sets of opinions or beliefs", which means the divergence of opinions and political attitudes to opposing extremes in social media or social networks in general. Group polarization itself is not a phenomenon bound to social media. It has been found in many diverse tasks and results in groups often making more extreme decisions than the average member in the group would make [Sun02].

We consider a topic as potentially polarizing if the existence of two or more opposing ideological extremes is possible. Consider the following question: "Should Austria welcome more refugees?" This is a controversial question, to which we might get different and conflicting answers. If this is the case and people hold opposing viewpoints, the topic is polarized. Often cited polarizing topics are often ones which are popular in the U.S. and are dividing the population there. Such as the legality of abortion, gun

[1] https://en.oxforddictionaries.com/definition/polarization

control or the political landscape in general, i.e. the split in democrats and republicans [GJCK13, CRF+11].

The existence of two or more opposing sides is crucial for the definition of polarization. Considering polar (yes-no) questions, these two sides typically correspond to either *yes*, supporting, or *no*, opposing a cause. Sometimes it might also be the case, that a third, neutral group exists which supports neither side.

## 2.2 Controversial or polarized?

Controversy and polarization – both terms are widely used in relevant literature. Often interchangeably. As defined by the Oxford Dictionary[2] controversy is a "prolonged public disagreement or heated discussion". In comparison with the definition of polarization provided above, it becomes evident that both describe related, but very different phenomena. While controversy describes a disagreement or discussion about something, polarization goes beyond that.

Polarization requires the division of the members of a discussion into sharply contrasting groups, which would precisely be what we examine in this thesis. As only such a sharp division into different groups leads to a wide range of adverse effects, such as people being easier to be influenced by "fake news" or easier to be influenced in general. Thus not only the presence of two clustered communities is essential, but also the communication structure between and inside these communities. A discussion is much more polarized, leading to more severe adverse effects if there is hardly any communication between the two opposing sides, which leads to echo-chambers. A very controversial discussion on the other side, where almost everyone communicates with everyone else, can hardly be considered polarized, as everyone is exposed to the respective opposing point of view.

One might argue though, that controversy might be a prerequisite for a topic or question to turn out polarized at some point in time. A topic being subject to a heated discussion with a large enough membership may or may not become polarized, depending on group dynamics and other factors. A topic which is not discussed controversially, rarely will become polarized. It would be possible, however, that a topic, although no subject of controversial discussions, will be polarized. The question, in this case, is if such topics would be worth considering in relevant research? As such topics would not be of general interest to society as they only affect a small minority.

We refrain from using these terms interchangeably, as they convey very different semantics. Therefore we limit our research in this thesis to topics, discussions and questions which are discussed controversially, a requirement for polarization, and are also polarized, which means, that we focus on subjects which are of interest for a large enough group of people to be discussed with such emphasis that they become polarized.

[2] https://en.oxforddictionaries.com/definition/controversy

## 2.3 Polarization and social network analysis

The polarization of a group of people and their interactions can be measured through social network analysis, which makes the analysis of social structures and phenomena possible through the use of networks and graph theory. Central to the concept of polarization is the presence of two opposing sides. Making it necessary to define three steps to measure polarization:

1. Construct network
2. Assign sides
3. Quantifying

Graphs and networks are the subject to analysis in social network analysis, and therefore crucial for any subsequent steps. Although platforms like *Twitter* or *Facebook* are well known as social "networks", data extracted from social media or other origins must be processed and transformed into a real network or graph. A graph is defined as a structure with individual subjects related to each other in some way. It consists of vertices (or nodes) representing individual subjects and edges representing a relationship between two subjects. Additionally, graphs can be distinguished into directed and undirected ones. If the relationships (the edges) of a graph are not valid in both directions (e.g. if A likes B, but B does not necessarily like A), a graph is directed, and undirected otherwise.

Transforming data extracted from any source into a graph requires to define what kind of data represents the vertices and edges. This strongly depends on the problem at hand. Vertices frequently represent the users of a platform. As edges represent some form of relationship, there are much more possibilities to define them. In the case of *Twitter*, it might be retweets or following another user. For *Facebook* sharing posts or friendships with other users may be possible. In general, the interaction must represent some kind of endorsement or disapproval. Without a clear positive or negative meaning associated with an interaction, it is not clear if two users are on the same side, or on opposing ones when seen in the context of polarization.

Before analysing the polarization of a network, it is necessary to define, or rather identify, these two (eventually) opposing sides. As most ways to measure polarization rely on the presence of a distinction of nodes into two groups. These groups are often referred to as clusters, components, modules or communities [GAS+15]. The outcome of this step is the assignment of every node in the network to one particular group. At the same time, each group represents one particular side of a discussion or opinion.

Although this could be done by hand for every node and may also achieve the best result, an automatic solution is needed, to automate the analysis process completely. In literature two conventional approaches are clustering and graph partitioning. Clustering is a well-studied data analysis problem, with many solutions like label propagation, spectral clustering and many more. A significant drawback of this approach, though, is the need for more information than being present in our scenarios. These methods work best if the network can be enriched with additional data, which is then used to assign nodes to clusters.

Therefore we will stick to experiment designs from literature with very similar settings [GMGM15, RAAJAK16, AAY+17, GMGM17, Kou18]. Numerous solutions for graph partitioning and community detection have been proposed in the past, as it is a frequent and well-studied problem [KK98, NG04, BGLL08, LLM10]. As an extensive in-depth comparison of such methods is beyond the scope of this thesis, we want to refer the interested reader to the cited literature.

Quantifying polarization is crucial to detecting and analysing polarization itself, its causes, and effects. All polarization measures defined in literature rely on a partitioned graph and on the interactions between these two sides. Some measures compare one characteristic, like betweenness or boundaries, with a significant difference indicating intense polarization. Some others rely on the strength of the clustering, i.e. the fact that high intra-cluster connectivity and low inter-cluster connectivity are common among polarized networks. As for choosing "the right" polarization measure is crucial, we compare and discuss them thoroughly in chapter 4.

The usage of graph partitioning to analyse polarization, as described above, is depicted in Figure 2.1. It shows a Barabási–Albert Graph [BA99] with n=30, m=15 on the left side and Zachary's karate club [Zac77], a typical example for polarization, on the right. Both graphs were partitioned using a graph partitioning algorithm called *METIS* [KK98], although every other partitioning algorithm should achieve similar results when aiming for two communities. The partitions are highlighted in either orange or blue.

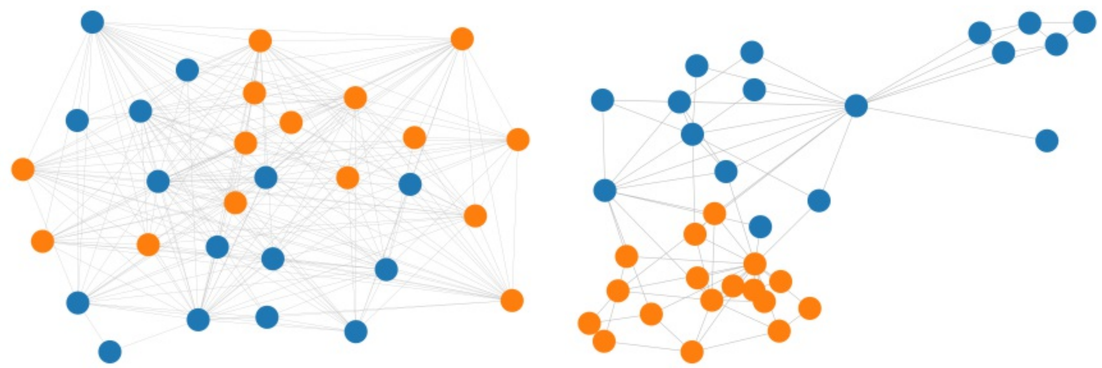


Figure 2.1: Two examples of graphs, with the left being unpolarized (Barabási–Albert Graph with n=30, m=15) and Zachary's karate club on the right.

In this simple example, it is evident that the left graph is unpolarized, whereas the right one is strongly polarized. This is based on the fact that the karate club has two communities, with very few connections between them, but proportionally more between the nodes belonging to the same community. Whereas the left graph was partitioned more or less randomly into two groups, showing no clear distinction. From a visual point of view, a graph is more polarized if nodes with the same colour sticking together, making it possible to draw a clear, imaginary line between the two communities.

Although this is easy to see and assess for small graphs with maybe even a couple of dozens of nodes, this procedure is neither doable for massive graphs, consisting of hundreds or thousands of nodes, nor does it lead to any quantifiable result, which is then achieved in the last step by using the partitioned networks as input for a polarization measure. In this case, this would, for example, lead to a *Modularity* score of 0.064 and 0.358, respectively.

## 2.4 Measurements

In this section, we give an overview of polarization measurements defined in literature. We only provide a brief description and (mathematical) definition for every measurement in this thesis and refer the interested reader to the original publications for further information and details.

### 2.4.1 Modularity

Newman and Girvan [NG04] proposed *Modularity* in 2004, as a measure for the strength of the community structure found by their proposed algorithms for discovering community structure in networks. These algorithms iteratively remove edges, identified by dedicated betweenness measures, from the network to form new communities, with these betweenness measures being recalculated after every iteration. Usually, the communities in a network are not known beforehand. However, such algorithms always produce a division into communities. Therefore a measure was needed to assess the quality of a particular division of a network into communities, which was called *Modularity*.

The *Modularity* of a community structure is the fraction of the edges that fall within the given communities minus the expected fraction if edges were distributed at random. For undirected graphs, it lies in the range -0.5 and 1, with higher numbers indicating a better division. It is defined as Equation 2.1. Where $A$ is the adjacency matrix of the graph and $A_{ij}$ the respective element of the adjacency matrix. The number of edges is given by $m$, while $k_i$ and $k_j$ are node degrees of the nodes $i$ and $j$ respectively, $\delta$ is the Kronecker delta, a function being 1 if the variables are equal, and 0 otherwise, with both variables being the community of the nodes they belong to, effectively neglecting the score of two nodes if they do not belong to the same communities.

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{2.1}$$

*Modularity* is frequently used in literature to assess the community structure and polarization of graphs and networks. It was especially used to analyse political relationships in the United States. Among them, congressional co-sponsorship (Zhang et al. [ZFT+08]) and polarization in the United States Congress in general (Waugh et al. [WPF+11]). In online social networks, it was applied to give evidence for the segregation between political groups (Conover et al. [CRF+11]). Furthermore, it is used as the basis of community detection algorithms trying to find the best partition by maximising *Modularity* [NG04, BGLL08].

Although frequently used and known to most researchers in relevant fields, it is also frequently criticised. A known limitation is the resolution limit of *Modularity*, which imposes a limit on the community size one can obtain by *Modularity* optimisation. The resolution limit is the result of the null model the *Modularity* measurement is based on. It compares the number of edges of the graph to be partitioned with the expected number of edges that would be present if the network would be a random network with the same number of nodes and where every node has the same degree as in the real graph. This random null model implicitly assumes, that every node can get attached to every other node in the graph, although this assumption is unreasonable for larger networks, as the horizon of a node (i.e. the nodes it might get attached to) gets rather small compared to the size of the whole network.

Considering a network large enough, the expected number of edges may get smaller than one, leading to the interpretation of even a single edge as a strong correlation between two clusters. In the case of *Modularity* maximisation algorithms, this leads to the merge of communities that otherwise would be distinct [FB07]. It also opposes certain limits on the application of *Modularity* for measuring polarization, as the behaviour gets different with increasing graph sizes.

Therefore it is not advisable to compare *Modularity* values across different networks with very different sizes [GJCK13]. This fact is crucial for our evaluation and comparison of the different polarization measures, as the network sizes are different in the order of magnitudes. For the analysis of content from *derStandard.at*, however, we will extract networks of roughly the same sizes, or at least within the same magnitude, making it possible in this case to compare the *Modularity* scores directly for these networks.

Guerra et al. addresses four drawbacks on mapping *Modularity* to the sociological behaviour of polarization:

1. They retrieved moderately high *Modularity* scores for datasets where they do not expect to observe any antagonism.
2. There are inconsistencies in interpreting *Modularity* scores in literature. Values considered to be high enough to be evidence of polarization in one publication ([ZFT+08]), are considered not to be associated with an evident community structure in another one ([CRF+11]).
3. Absence of links as a sign of polarization
4. Resolution Limit Problem

We will address the first point in the evaluation of the implemented measurements in chapter 4. The second point addresses the problem that there is no consensus, guideline or defined ranges, when the *Modularity* obtained is high enough to be considered polarized, non-polarized or how exactly this is coupled to network size itself. The drawback of the resolution limit problem and how it affects interpretability has already been discussed above.

The rationale behind *Modularity* is the interpretation of missing links between communities as a sign of *antagonism* between them, i.e. a separation into partitions. It compares the internal and external connectivity of both partitions with each other. However, the formulation of both, homophily (nodes tend to form connections between each other due to similarity) and antagonism (nodes avoid forming connections between each other due to differences), in one formula limits the understanding of antagonism in isolation and therefore polarization. Antagonism is not bound to the absence of links, as there may also be antagonism in almost complete graphs, where everyone is intimately connected but has very different opinions or viewpoints [GJCK13].

### 2.4.2 Betweenness Centrality Controversy (BCC)

This measure was defined by Garimella et al. [GMGM15] based on the different distributions of betweenness centrality among the partitions and the cut between them. The betweenness centrality $b(e)$ of an edge $e$ is defined as

$$b(v) = \sum_{s \neq t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}} \tag{2.2}$$

Where $\sigma_{s,t}$ is the total number of shortest paths from node $s$ to node $t$ and $\sigma_{s,t}(e)$ is the number of these paths that include edge $e$.

The intuition behind this metric is that, if the partitions are well-separated, the cut between them will consist of edges having a higher betweenness centrality, than edges inside the partitions. In this case, the shortest paths connecting vertices of the two partitions will pass through the edges in the cut, leading to higher betweenness centrality scores than in the rest of the graph. If the graph is not well partitioned (i.e. no polarization), then the cut will consist of strong ties [Gra77]. Therefore the shortest paths connecting vertices of the two partitions will pass through one of the many edges on the cut, leading to similar betweenness centrality scores similar to the rest of the graph.

For the comparison of the two distributions (edges in cut vs. rest of the graph) of edge betweenness centrality scores, the Kullback-Leibler divergence (KL) [KL51] is used. The Kullback-Leibler divergence, also called relative entropy, measures the difference between two probability distributions. In this case, the difference between the distribution of betweenness centrality scores among edges in the cut and the rest of the graph is measured. In the original experiment design, Garimella et al. sampled 10,000 scores with replacement from every distribution, as KL divergence demands two lists with the same length as input, and the cut-distribution naturally consists of fewer data points as the other distribution. Therefore kernel density estimation is used to compute the respective probability density function (PDF) for both distributions, to then sample the necessary data points from it. The *Betweenness Centrality Controversy (BCC)* score is then defined as:

$$BCC = 1 - e^{-d_{KL}} \tag{2.3}$$

Which normalises the score into a range between zero, effectively indicating no polarization at all, and one, indicating extreme polarization.

### 2.4.3 Embedding Controversy (EC)

This measurement was proposed by Garimella et al. [GMGM15] with the rationale, that the same algorithm (Gephi's *ForceAtlas2* [JVHB14]) creating layouts of graphs for their visual analysis, might also be suitable to quantify polarization. Because a force-directed embedding also maximises *Modularity*, the two-dimensional layouts produced by such algorithms indicate a layout with maximum *Modularity* [Noa09].

Consider a two-dimensional embedding $\phi(v)$ of vertices $v \in V$ and given two partitions X and Y of a graph, the *Embedding Controversy (EC)* measure is defined as:

$$EC = 1 - \frac{d_X + d_Y}{2d_{XY}} \tag{2.4}$$

Where $d_X$ and $d_Y$ is the average distance among pairs of vertices in the same partition, for X and Y respectively. Where the distance between two vertices is calculated by using the position of the points in the two-dimensional plane produced by the layout algorithm. Whereas $d_{XY}$ is the average distance among pairs of vertices across the two partitions X and Y. Resulting in *EC* being able to be paraphrased as *"1 minus the sum of intra-partition distances divided by two times the inter-partition distance"*.

The score obtained is close to zero for graphs of topics which are not polarized, and one for polarized ones–corresponding to more divergent layouts produced by the force-directed layout algorithm, which is the case for well-separated communities/partitions.

In the original paper, the two-dimensional embedding was retrieved using *ForceAtlas2*. However, Darwish [Dar19] proposes the usage of *Uniform Manifold Approximation and Projection* (UMAP)[3] to retrieve the embedding. *UMAP* is a dimension reduction technique based on manifold learning techniques and ideas that can be used for visualisation similarly to t-Stochastic Neighbor Embedding (t-SNE), one of the most popular dimensionality reduction methods besides Principal Component Analysis (PCA). The

[3]Implementation in Python by the original authors: `https://umap-learn.readthedocs.io`

method is constructed from a theoretical framework based on Riemannian geometry and algebraic topology. The algorithm consists of two phases. In the first phase, a fuzzy topological representation of the data is constructed. The second phase is optimising the low dimensional representation using cross-entropy as a loss function [MHM18].

Darwish argues that *UMAP* being more aggressive in constructing the layout or two-dimensional embedding, i.e. projecting similar users near to each other, while pushing dissimilar ones further apart than *ForceAtlas2*. Their experiments consistently show higher polarization scores over all of their datasets when using *UMAP* instead of *ForceAtlas2*.

### 2.4.4 Boundary Connectivity (BC)

The criticism of *Modularity* and the discussion of its drawbacks lead to the formulation of a new polarization measure by Guerra et al. focusing on antagonism, not defined through the absence of connections, but the tendency to connect to either group of nodes on the cut or boundary between two partitions.

A node $v$ is part of a *community boundary* $B_{i,j}$ of a community $G_i$ if it satisfies two conditions:

1. It must have at least one edge connection to community $G_j$.
2. It must have at least one edge connecting to a member of $G_i$ which is not connected to $G_j$.

Which can be formally defined as:

$$B_{i,j} = \{v_i : v_i \in G_i, \exists e_{ik} | v_k \in G_j, \exists e_{ik} | (v_k \in G_i, \nexists e_{kl} | v_l \in G_j), i \neq j\} \tag{2.5}$$

Defining nodes from $G_i$ which do not belong to $B_{i,j}$ as *internal nodes* $I_i$ leads to 4 sets: $I_i$, $B_{j,i}$, $I_j$ and $B_{i,j}$. Based on this the authors define two sets of edges $E_B$, capturing edges that connect members from $G_i$ to members from $G_j$, and $E_{int}$, capturing edges that connect boundary nodes to internal ones:

$$E_B = \{e_{mn} : v_m \in B_{i,j} \wedge v_n \in B_{j,i}\} \tag{2.6}$$

$$E_{int} = \{e_{mn} : v_m \in (B_{1,2} \cup B_{2,1}) \wedge v_n \in (I_1 \cup I_2)\} \tag{2.7}$$

Equation 2.8 depicts the comparison among the connectivity choices all nodes in $B_{i,j}$ make regarding nodes in $I_i$ or $B_{j,i}$. For each node $v$ in the community boundary $B$ the ratio between the number of edges it has in $E_{int}$ ($d_i(v)$) and the total number of edges in $E_B$ ($d_b(v)$) and $E_{int}$. Based on the null hypothesis that each node have the same

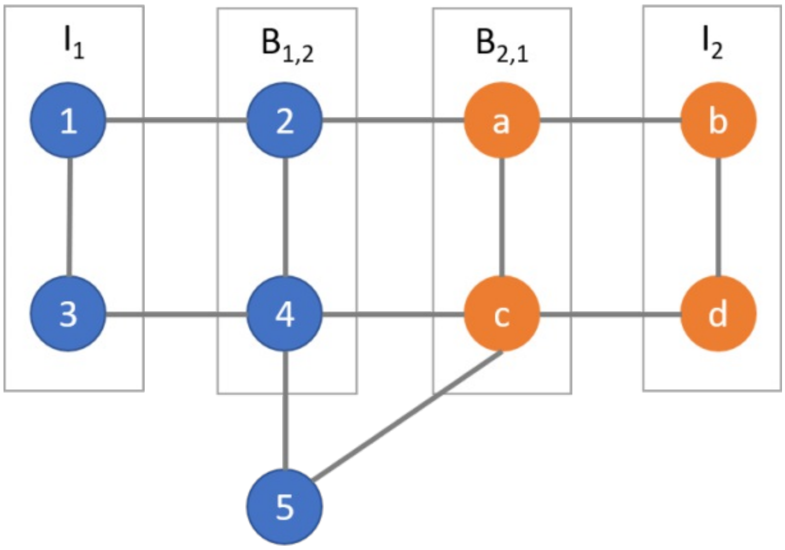



Figure 2.2: Example of a graph divided into two communities according to Guerra et al.

preferential attachment for both, internal nodes and nodes from the other community, i.e. edges are spread equally among both sets.

$$P = \frac{1}{|B|} \sum_{v \in B} \left[ \frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \right] \tag{2.8}$$

The resulting measure $P$ lies in the range -1/2 to 1/2, where a negative value does not only indicate a lack of polarization, but even that nodes in the boundary are more likely to connect to nodes of the other community, than to nodes of the same one.

An example is given in Figure 2.2, inspired by an example in the original paper. A graph is partitioned into two communities, orange and blue. Following the procedure described, we end up with the sets $I_1 = 1, 3$, $I_2 = b, d$ and $B_{1,2} = 2, 4$, $B_{2,1} = a, c$. This graph has a *Modularity* score of 0.25, indicating a reasonable level of polarization. However, calculating the score according to the formula above yields a polarization score of 0. The network does not show any signs of polarization at all. As no node in the boundary exhibits any type of antagonism since the same number of cross-group and internal edges are established.

### 2.4.5 Polarization Index (PI – Matakos et al.)

This measure, originally proposed by Matakos et al. [MTT17], is based on opinion formation theory. According to Friedkin and Johnsen [FJ90] and their opinion-formation model, every person $i$ has a persistent *internal opinion* $s_i$ and an *expressed opinion* $z_i$ which depends both on the respective internal opinions $s_i$, and the expressed opinions of their neighbours. Additionally, every edge (i,j) is associated with a weight $w_{ij}$, expressing the strength of the connection (or the influence they have on each other) between i and j. It has been shown that iteratively updating their expressed opinion, leads to the convergence of the expressed opinions to an opinion vector $z$. Where the expressed opinion of a person $i$ is the weighted average of the internal opinion and the external opinions of the neighbours $N(i)$:

$$z_i = \frac{w_{ii}s_i + \sum_{j \in N(i)} w_{ij}z_j}{w_{ii} + \sum_{j \in N(i)} w_{ij}} \tag{2.9}$$

The expressed opinion might be positive or negative. Where the absolute value of the opinion captures how extreme the opinion of the user in consideration is. Under this rationale, the absence of any polarization in the networks would express a purely neutral opinion vector, i.e. $z_i$ being 0 for every node in the network.

The measure is based on the findings of Gionis et al. [GTT13], that there is a direct connection between the opinion formation model, and random walks with absorbing nodes, i.e. every random node reaching such an absorbing node ends at such a node. Consider a typical graph $G = (V, E)$. Now an augmented graph $H$ is constructed as follows. For each vertex $v_i \in V$, a new vertex $x_i$ is added, symbolising the internal opinion of the node. Additionally a directed edge $r(v_i, x_i)$ is added. With $X$ being the set of all newly added vertices and $R$ being the set of all newly added edges, a new graph is created $H = (V \cup X, E \cup R)$.

The opinion vector $z$ is now calculated as follows. A random walk on graph $H$ starts from a vertex $v \in V$. As stated already, the nodes in $X$ are absorbing nodes, i.e. a random node reaching such a node ends at that node. For each absorbing node $x_i$ and the starting node $v_j$, the probability $P(x_i|v_j)$ can be computed. Being the probability of a random walk starting from $v_j$ terminating at node $x_i$, with equation 2.10 showing this relation. The probability $P$ can be interpreted as the probability that a person $(v_j)$ adopts the opinion of person $(v_i)$.

$$z_j = \sum_{i=1}^{n} P(x_i|v_j)s_i \tag{2.10}$$

Initially every node is assigned a specific opinion between -1 and 1, which, in our experiments, are set according to the partition a node belongs to. All nodes belonging to the first partition will be assigned -1, whereas all nodes from the second partition will be assigned an opinion of 1. Having then calculated the opinion vector $z_i$ for every person $i$ in the network, the *Polarization Index* is then the length of the whole vector $z$ under the $L^2$ norm $||z||$. To account for scaling effects of larger or smaller networks, the result is then normalised by the number of nodes in the graph, as depicted by equation 2.11.

$$\pi(z) = \frac{||z||^2}{n} \tag{2.11}$$

The measure yields results between zero and one. With low values are indicating that the opinions are rather neutral in the network, i.e. a lack of polarization. Extreme opinions of either side, positive or negative, account for large values, indicating extreme polarization across the network.

### 2.4.6 Polarization Index (MBLB – Morales et al.)

The electric dipole moment – a measure of the separation of positive and negative charges in a system, i.e. a measure of the system's overall polarity – inspired Morales et al. [MBLB15] to define the *Polarization Index*. In a network, they distinguish between two different types of nodes, *elite* and *listeners*. Where the first ones have a fixed opinion of $\pm 1$, influencing the opinion of the second ones, whose opinion is formed based on their neighbours. As *elite* nodes represent the most influential individuals in a network, the nodes with the highest degrees in a network are chosen ([GMGM15]).

At each iteration, the *listeners* in the network update their opinion based on the mean of the opinions expressed by their neighbours, normalised by the number of neighbours, as shown by equation 2.12. Where $A_{ij}$ represents the elements of the adjacency matrix which are neighbours of the current node, $X_j$ with $-1 \leq X_j \leq 1$ is the opinion of a neighbour and $k_i^{in}$ represents the in-degree of the node. This procedure is repeated continuously until $X_i$ of every node $i \in V$ converges.

$$X_i(t) = \frac{\sum_j A_{ij} X_j(t-1)}{k_i^{in}} \tag{2.12}$$

After convergence, let now be $n^+$ the number of nodes with a positive value and $n^-$ the number of nodes with a negative one. $\Delta A$ is now the normalised difference of these two population sizes (2.13). Moreover, let $gc^+$ be the average polarization value among the vertices with a positive value ($n^+$), and $gc^-$ the average among $n^-$. The pole distance $d$ is now defined as equation 2.14. With $d = 1$ indicating extreme polarization, and $d = 0$ when the opinions are not divergent.

$$\Delta A = \left| \frac{n^+ - n^-}{|V|} \right| \tag{2.13}$$

$$d = \frac{|gc^+ - gc^-|}{2} \tag{2.14}$$

The *polarization index* $\mu$ is then given by equation 2.15. Yielding values of $\mu = 1$ when the network is perfectly polarized, i.e. node values are centred near $\pm 1$. Whereas $\mu = 0$ would represent a network which is not polarized at all. In this case, opinions are centred around 0, or around one of both poles, which although indicating the presence of extreme opinions, is contradicting to the definition of polarization.

$$\mu = (1 - \Delta A)d \tag{2.15}$$

### 2.4.7 Random Walk Controversy (RWC)

This measure, originally defined by Garimella et al. [GMGM15] is based on random walks in a partitioned network to measure the inherent controversy. Similar to *polarization index* it is assumed that there are high-degree vertices in both partitions, i.e. sides of a discussion, having high influence inside the network, spreading their opinion among adjacent low-degree nodes.

Consider a graph $G$, partitioned into two disjoint sets $X$ and $Y$. The *random walk controversy (RWC)* measure is now defined as follows. Let there be two random walks– one ending in partition X and the other one ending in partition Y. The RWC measure is now the difference between the probabilities of the following two cases. Both random walks started in the same partition they ended in (i), and both random walks started in a different partition than they ended in (ii) (Equation 2.16). Where $P_{P1P2}$, with $P1, P2 \in \{X, Y\}$, is the conditional probability $P_{P1P2} = Pr[\text{start in partition P1}|\text{end in partition P2}]$.

$$RWC = P_{XX}P_{YY} - P_{YX}P_{XY} \tag{2.16}$$

The resulting score lies between zero and one. With low values are indicating that the probability of random walks crossing sides is the same as staying in the same partition, i.e. a lack of polarization in the network. Whereas values close to one are reached when the probability of random walks crossing sides is very low, a sign of high polarization.

## 2.5 Usage in literature

To get an overview of the occurrences of every measure in literature, we conducted an extensive literature review. For this, we started with the original paper for every measurement and went through every paper referencing it. Through this process, we aim to find all potential papers which use one of the measurements either directly in an experiment or build upon the theoretical base of it.

We limit our search to papers which are available in English and in full text to the general public (or to students of the TU Wien) and are listed as cited on Google Scholar. With this procedure, we ended up with almost 9000 papers. Out of which a vast majority (over 8000) cited the original *Modularity* paper [NG04]. As it was predetermined, that this will be the most used and most popular measurement, we will exclude it in our analysis. Furthermore, we excluded all papers where the original author of a measurement, is among the authors of a relevant citing paper. As this literature review was conducted on 21.01.2020, it will only include papers which met the conditions described above at this point of time.

The resulting papers of this review are given in Table 2.1. In addition to the usage of one of the measurements in a paper, we also recorded the usage of content-based methods in conjunction with the measurements. In the majority of cases, sentiment analysis was used in various forms, to improve the performance of classification models or polarization models.

| **Author** | **BCC** | **BC** | **EC** | **MBLB** | **PI** | **RWC** | **Content** |
|---|---|---|---|---|---|---|---|
| Al-Ayyoub et al. [AARJ+18] | | x | x | | | x | x |
| Atienza-Barthelemy et al. [ABMGLB19] | | | | x | | | |
| Badami et al. [BNSS17] | | | | x | x | | x |
| Chen at al. [CLDB18] | | | | | x | | |
| Chitra and Musco [CM19] | | | | | x | | |
| Darwish [Dar19] | | | x | | | x | |
| Garimella et al. [GMGM15] | | x | | x | | | x |
| Kumar et al. [KHLJ18] | | | | | | x | x |
| Morel [Mor16] | | | | x | | | x |
| Olivares et al. [OCLB19] | | | | x | | | |
| Ozer et al. [OYD17] | | | | | | x | |
| Primario et al. [PBI+17] | | | | x | | | x |
| Rabab'ah et al. [RAAJAK16] | | x | x | | | x | |
| Roth et al. [RMM20] | | | | | | x | |
| Rumshisky et al. [RGP+17] | | | | | | x | x |
| Sirrianni et al. [SLA18] | | | | x | | | |
| Nair et al. [NIS19] | | x | | | | | |
| **Citations (usages)** | 0 | 4 | 3 | 7 | 3 | 7 | 7 |
| **Citations (all)** | (156) | 88 | (156) | 95 | 25 | (156) | - |

Table 2.1: Usages of the described polarization measurements in literature (excl. *Modularity*)

We recorded the most usages of a single measurement for *RWC* and *Polarization Index (Morales et al.)* with a total of seven papers. Not a single occurrence of the *Betweenness Centrality Controversy* could be recorded on the other hand. Out of 17 papers in total, seven papers additionally used content-based methods.

Comparing citations overall as available on Google Scholar, *Modularity* is uncontested as the most cited measure (8298 citations). We put the number of 156 citations for the measures defined by Garimella et al. in parentheses, as all three were defined in the same paper, making it difficult to differentiate here. Additionally, Garimella et al. and Guerra et al. are more often cited for their theoretical concepts in the area of controversy and polarization than for the measurements they defined. In this context, the most citations are recorded for Polarization Index (MBLB, Morales et al.), whereas Polarization Index (PI, Matakos et al.) is cited the least often.

Among the referenced literature stated in Table 2.1, there is one single proposed measurement we did not include in our experiments and comparisons. Sirrianni et al. [SLA18] propose a polarization measure based on a polarization measure used in the field of economics originally proposed by Esteban and Ray [ER94]. Their model formulates polarization as the presence of clusters, in which the attributes of its members are very similar (intra-group identification), but different clusters have members with very dissimilar attributes (inter-group alienation).

Sirrianni et al. extend this definition in various ways, such as, among others, limiting the measurement to a range of $[0, 1]$ and the measurements must be normalised for the number of participants (i.e. the measurement is not affected by the size of the population). In their experiments they proof that their *modified Esteban and Ray measurement (MER)* is superior to *MBLB*, which was used as a comparison.

Although this measurement seems promising for further evaluation, its application lies beyond the scope of this thesis. We are limiting our studies to measurements used in very similar contexts. Sirrianni et al. however use their measurement to assess the polarization based on pre-calculated similarity scores between users. These scores are based on the data present in a specialised online argumentation tool. Based on the expressed opinion of every user towards certain topics, and a set of inference rules, the agreement level towards a topic is calculated for every user.

To apply *MER* in our experiments, we would also need to determine specific agreement levels for every node in a network. This would require an extension of the original definition of the measurement by extending it with a pre-calculation phase in which such agreement levels are calculated. As there are several possibilities to achieve this and the measurement strongly depends on this phase, it would result in a range of new measurements, one for every possibility. The formulation of new measurements, as well as the adoption of measurements in any significant ways, is beyond this thesis. Nevertheless, *MER* is definitively a subject for future research.

## 2.6 Content-based methods

The methods of social network analysis mainly rely on structural properties of a network as the subject to analysis. Another approach in the area of social media platforms, blogs or online forums, is using the content generated by users, i.e. the text of comments, posts or articles. The information extracted this way may be incorporated in different phases when analysing polarization.

Analysing the meaning and properties of the text of a comment or similar may help in partitioning the nodes in a network into separate communities or clusters – effectively improving the quality of the partition, leading to cleaner cuts, and less false-positives. Which in turn will positively affect the outcome in the end. Not all signs of endorsement are unambiguous. Analyzing the content leads to a process more similar to how humans perceive a discussion and its polarization.

Such methods, however, can even be used as an alternative to social network analysis. Either through applying sentiment analysis and combining single scores to a final statement about the polarization [CJM10, SPP16, SZLK13], or through analysing other properties of the content, among others differing bag of words, i.e. different sides of a discussion use different vocabularies to express their opinion [BKvdV17, CGGL17, GMGM15].

Overall content-based methods seem very promising, as they also come closer to human perception than pure structural ones. The effectiveness of such methods strongly depends on the quality and expressiveness of the content. Especially in the case of social media or online forums, where most texts are very concise and do not adhere to standard grammar rules, such methods have their limitations.

Applying these methods in the context of Austrian online news blogs would be even more difficult, as sentiment analysis works best, and is applied most often, on English texts. However, comments often include slang words, misspelt words or are written in local dialects. Also, the limited input would be a factor to consider, experiments conducted in literature considered far more data than in our case, where most discussions do not exceed more than 1.000 comments. Although the application of such methods in our context would nevertheless be very interesting, we want to limit ourselves to the methods of social network analysis and leave the others as a subject for future research.

## 2.7 Indicators

This thesis focuses on the analysis, comparison and application of polarization measurements. These are the only option for quantifying the strength of polarization in a social network. Sometimes, depending on the problem and setting at hand though, it might be enough to only identify controversial or polarizing topics without ultimately quantifying the polarization. In such cases polarization indicators are suitable to indicate if a network or discussion is polarized or not, without saying something about the exact strength of the polarization.

These indicators are often used in literature in the context of "automated controversy detection" problems. The main objective in these cases is solely the distinction of topics into ones which are controversially discussed, and the ones which are not. With the growing number of applications and experiments also grew the number of features or polarization indicators studied. Nevertheless, they can also be grouped into content-based ones, based on the text generated by users, and structural ones, based on the structure of network graphs.

Conover et al., while studying political polarization on *Twitter*, discovered that controversial discussions, when transformed into a network structure, are organised into highly segregated communities [CRF+11]. Though not only network-wide features can be a sign of polarization, but also very local ones such as so-called motifs, i.e. local patterns of user interactions in conversation graphs. Coletto et al. discovered that controversial threads create engagement among users not being directly connected in the social network and that more heated discussions induce shorter reply times [CGGL17].

Historically much relevant research is based on the data obtained from Wikipedia and is concerned with detecting controversial articles there. Dori-Hacohen and Allen show that structural properties of the networks, such as the number of editors (=size of the network) or the length of the editing page, are reliable indicators when identifying controversial topics. Proofing that there are often multiple ways to transform data into networks, and that they encode different, but very relevant semantics.

Content-based indicators can put a different focus on different features of user-generated content. The most common one is the sentiment or polarity of comments, where methods are used to predict if a phrase or comment is either more positively or negatively framed. Besides that, indicators may also focus on the words used (lexical features) or the general text structure (linguistic features). In the former case, the emphasis lies on the words used and their frequency, whereas in the latter, the emphasis is put onto the structure of sentences and phrases. They are often used together in experiments as part of large sets of features [RB12] where up to 150 features are used to identify controversial topics [KAHF14]. Although experiments also prove their value in isolation, such as experiments by Jang et al. comparing the language used in discussions with ones which are known to be polarized, to detect new ones [Jan19].

The most commonly applied content-based indicators are sentiment based ones. They were applied in numerous settings and experiments, ranging from Wikipedia articles, over political debates to discussion on *Twitter* and other social networks. Topics are considered controversial if there is a significant difference between the amounts of the two different sentiment polarities (e.g. positive-negative, or -1 - 1) [CJM10]. A novel measure for contradiction was introduced by Tsytsarau et al. [TPD11] based on the mean value and the variance of sentiments among different texts used later on in other experiments [PP10].

Mejova et al. combined lexical features with the emotions retrieved through sentiment analysis in their studies. They found that in the case of controversial discussions and topics, the use of negative affect and biased language is prevalent, but to the contrary of previous findings in literature, the use of strong emotion is tempered. Although they also note that the findings of sentiment analysis are not necessarily directly related to controversy. Some topics such as *guns*, *war*, *army* automatically involve negative sentiments, whereas others such as *health* involve positive ones [MZDC14].

Some research conducted in the past even combined content-based with structural features, such as Amin et al., following a matrix factorization approach to unveil polarization in social networks [AAY+17]. Whereas Beelen et al. combined a total of 20 features to detect controversy in online news media, concluding that lexicon-based indicators (offensive words as well as those indexing disagreement) act as solid predictors of controversy as well as negative emotions [BKvdV17].

In some cases, however, neither structural nor content-based indicators are used to detect if a discussion is controversial or not. Badami et al. for example, proof that sometimes completely different data than the network or comments can be used to identify polarized discussions. They use the histogram of product ratings, obtained from recommender systems. Following the rationale that specific distributions (e.g. bipolar ones) are reliable indicators for polarization and effectively outperform sentiment-based indicators with their method [BNSS17].

## 2.8 Approaches

Numerous publications examined the aspect of polarization in social networks, with one of the earliest dating back to Bienenstock et al. [BBO90] and Zachary [Zac77]. In the context of online social media, polarization has been studied since shortly after the origin of social media itself. One of the most common subjects to research is the political polarization in the U.S., first studied in the context of online media by Adamic and Glance [AG05] which showed that blogs were divided into two segregated groups, mimicking the political segregation into Democrats and Republics. Through community detection, they detected the opposing groups inherent in the data extracted from political blogs and visualized polarization that way.

Social networks are projected into graphs to analyse them efficiently. As with every projection, the emphasis is laid on a selection of features, whereas the non-selected features are omitted. Therefore for most social networks, there are different projections for the same graph. Because of this, a range of different ways to construct networks exists in literature. Most networks consist of users or persons, the nodes, which are connected through activities, indicated by edges [NG04]. Much research is based on data extracted from *Twitter*. Frequently the data is used to construct so-called retweet networks. In which an edge $A \rightarrow B$ indicates that user B retweets the content generated by user A, indicating that information has propagated from A to B [CGR+11]. Some social media platforms also offer the functionality of friendships or followers, which can also be incorporated into networks, in which an edge then indicates friendship [DGL13].

Although such inherent structure-inducing properties are used to construct network graphs, advancements in *Natural Language Processing* lead to a rising number of publications using the content generated by users in social media platforms to construct networks from it. Most approaches are based on the sentiment of articles, comments or posts generated by users, to derive polarization measures from it. The outcome are signed social networks, not only indicating the strength of activity between two nodes but also if the activity is positive or negative [AK08, BCM+09, HAJR12]. Even though a combination of Natural Language Processing with social network analysis seems promising; it is beyond the scope of this work, which primarily deals with measuring polarization in settings solely based on methods from social network analysis.

Polarization in networks is often analysed as a matter of community detection and graph partitioning. Partitioning a graph and extracting communities reveals insights about how polarized a network, therefore consisting of more than one community, is. To quantify the extent of polarization, *Modularity* is considered frequently as metric (see Section 2.4.1). *Modularity* is, up to a multiplicative constant, the number of edges falling within groups minus the number of edges in a network were edges are distributed randomly [NG04, New06]. Since introduced by Newman and Girvan in 2004, it has been used extensively in literature [CGR+11, CRF+11, BHMDW15, GAS+15, WMKHL15, ZKWM18].

*Modularity* is not only frequently used, but also frequently criticised in literature, leading to a number of other ways of measurement proposed. Guerra et al., for instance, proposed one of these different measurements in 2013. They are arguing, that *Modularity* does not respect real antagonism between groups, leading to wrong statements about the polarization of networks in some cases. They instead propose a metric based on the analysis of the boundary of a pair of (potentially polarized) communities [GJCK13].

A different approach was chosen by Garimella et al. in 2015. Their proposed *Random Walk Controversy (RWC)* (Section 2.4.7) measurement is based on the idea of the difference of the probabilities of the outcome of two random walks. Distinguishing between two cases, (i) that both random walks end in the same network partition they started in, and (ii) both random walks end in a different partition than they started in [GMGM15, GMGM18]. Matakos et al. proposed another approach, the *polarization index* (Section 2.4.5), capturing the tendency of opinions to concentrate in network communities [MTT17]. Although the number of different measurements is increasing, an extensive comparison of different measurements in literature has not been conducted yet.

# CHAPTER 3

# Experiment design

## 3.1 Measurements

Except for *Modularity*, there exists no publicly available implementation of the measurements described in section 2.4. Therefore we have to implement them following the definitions of the original authors. As Python has become one of the major programming languages for Data Science, many libraries have been published in the past. To utilise this technological foundation, our implementation will be completely done in Python. Among other libraries, we rely on the two popular graph processing libraries *NetworkX*[1] and *igraph*[2].

To verify the correctness of our implementation, we will compare the results of our implementation to the results of either the original authors or other usages of the same measure in literature. For this, we will use a set of datasets previously used in similar studies in literature (Table 3.1).

The first four datasets were originally created by Guerra et al. who kindly shared them with us: *Brazilian soccer supporters*, retweets extracted from *Twitter* mentioning two popular Brazilian soccer teams; *Gun control*, retweets mentioning gun control issues after the Sandy Hook Elementary School shooting; *New York City sports teams*, retweets about two New York City sports teams; *University friendships' network*, friendships established on *Facebook* by members of a Brazilian university; *Zachary's karate club* and *2004 U.S. political blogosphere*; web links between blogs discussing politics in the U.S. around the 2004 election, originally created by Adamic et al.[AG05].

Figure 3.1 shows the network representation of each of the six datasets from literature used in our experiments. The layout was produced in Gephi using the *ForceAtlas2* graph

[1] `https://networkx.github.io/`
[2] `https://igraph.org/python/`

| Dataset | $\lvert V \rvert$ | $\lvert E \rvert$ | largest CC |
|---|---|---|---|
| Brazilian soccer supporters | 27,415 | 94,043 | 75.12% |
| Gun control | 33,762 | 350,107 | 98.50% |
| New York City sports teams | 113,840 | 199,329 | 84.26% |
| University friendships' network | 303 | 4,449 | 92.74% |
| Zachary's karate club | 34 | 78 | 100% |
| Zachary's karate club [EK10] | 34 | 78 | 100% |
| 2004 U.S. political blogosphere | 1,490 | 16,718 | 82.01% |

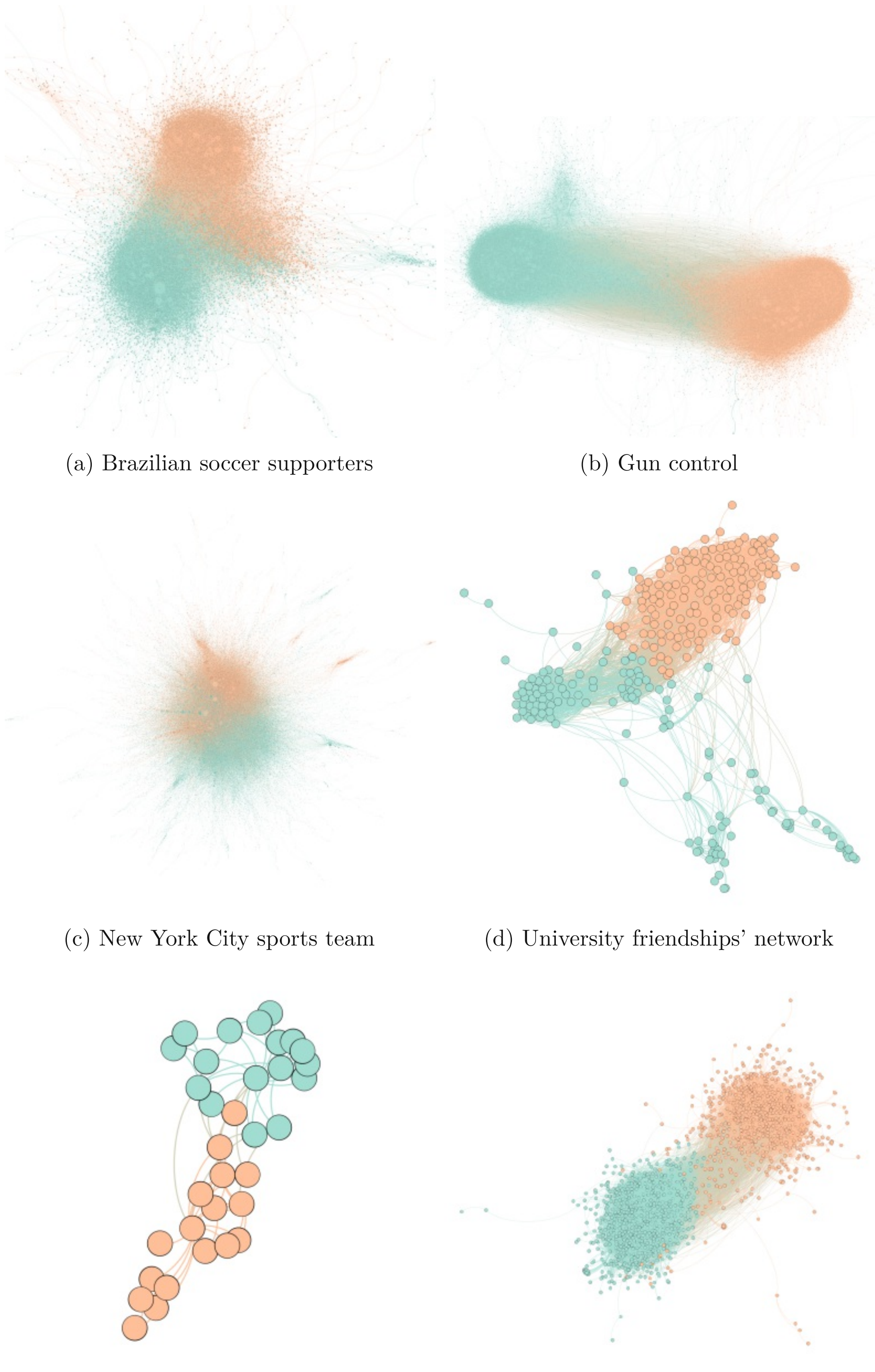
Table 3.1: Datasets from literature and their size (incl. largest connected component relative to the whole graph)

layout algorithm. The nodes are coloured according to the partitioning produced by METIS. The networks are not only varying in size but also the structure of their layout is very different. Whereas *Zachary's karate club* or *2004 U.S. Political Blogs* consist of two very distinct communities, the communities of networks like *New York City sports team* or *Brazilian soccer supporters* stick much closer together, i.e. there is no clear distinction between the communities.

*Zachary's karate club* is actually used twice in our experiments, but with different partitions. In the first case, we rely on the same automatic partitioning as in the other cases. In the second case, however, we use the partitioning originally provided by Easley and Kleinberg 2010 [EK10] and also used by Guerra et al. in their experiments. Though both settings only differ by one node, this could lead to significant differences later on, as the network only contains 34 nodes overall.

The low number of datasets, measurements and usages of them in literature reduces the accuracy and significance of any experiments, as the sample size is just too small. Therefore we include two additional datasets, which cover extreme cases in the context of polarization and should function as a sanity check of the measurements and our implementation. Both graphs are shown in Figure 3.2

The first one is a complete graph with 600 nodes where every node is connected to every other node. This graph should cover the case of a complete absence of polarization, as a complete graph cannot be split into two or more loosely coupled communities. The second graph consists of two smaller complete graphs, with 300 nodes each. Both complete graphs are connected by 15 edges, each connecting two randomly chosen nodes of each cluster. With this extreme case, we want to cover a setting where extreme polarization is present, as there are two almost wholly isolated communities with only a couple of connections to the other one.



(a) Brazilian soccer supporters

(b) Gun control

(c) New York City sports team

(d) University friendships' network

(e) Zachary's karate club

(f) 2004 U.S. Political Blogs

Figure 3.1: Networks of all datasets from literature used in this thesis.

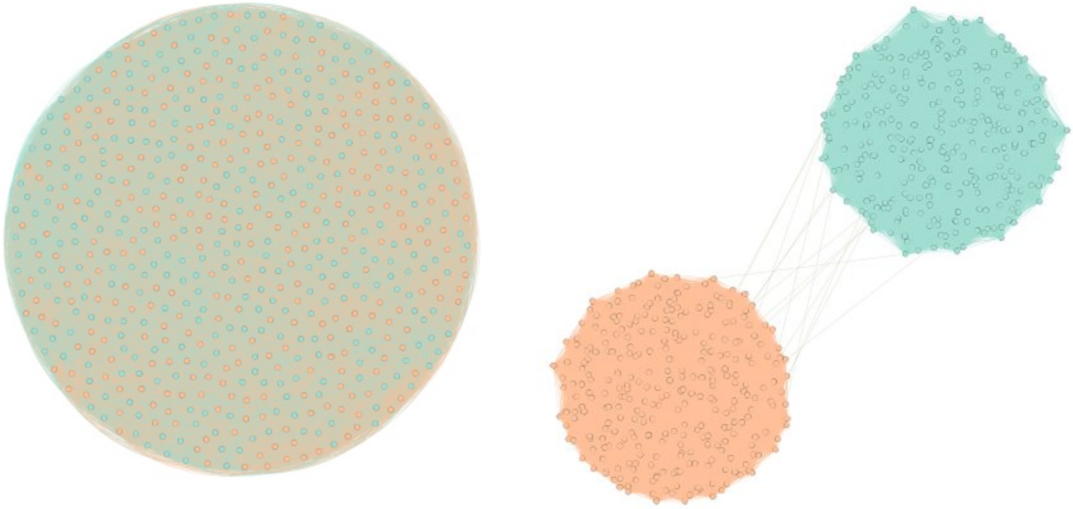


Figure 3.2: Synthetic graphs used in the evaluation phase.

## 3.2 The data

The data we use in our experiments was kindly provided by *derStandard.at* in the form of a database excerpt of 06.06.2019. It is the second most popular online news platform in Austria according to the *Österreichische Webanalyse (ÖWA)*, with more than 448.500 daily unique users [Der19]. Of all the data available, we only focus on the parts which are relevant to our experiments. An overview of these tables is given in the ER model shown in Figure 3.3.

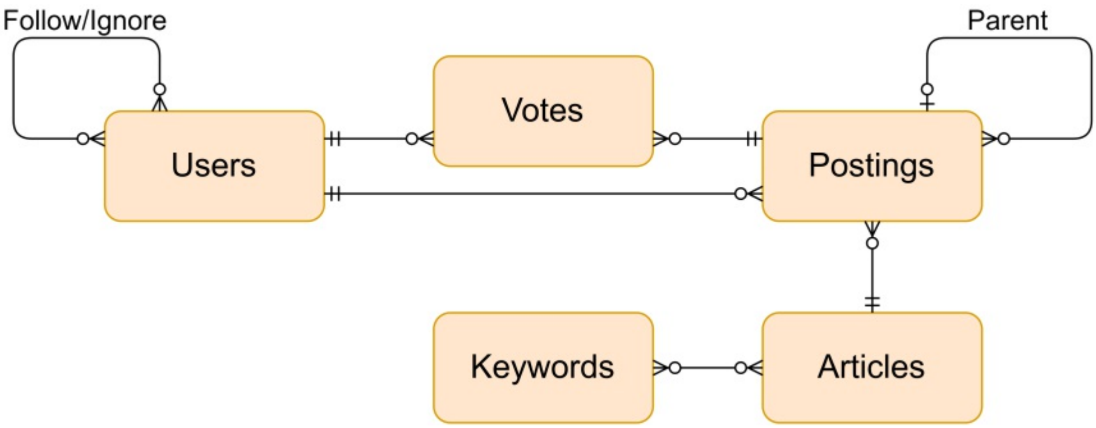



Figure 3.3: *derStandard.at* ER diagram


Figure 3.4: Visually edited Screenshot of the article with the most comments on *derStandard.at* (marginally altered structure and content for more conciseness)

Similar to other online news platforms, the content on *derStandard.at* is presented in the form of single articles, which are grouped in different ressorts and topics. Users, once registered and logged in, have the possibility to write postings or comments underneath every article or reply to comments written by other users. Moreover, users can express their opinions about certain postings by up or downvoting comments of other users.

Figure 3.4 shows a screenshot of the browser version of *derStandard.at*. As usual for newspapers the main content is the article text itself (which has been shortened to one paragraph for a more concise picture). The comment section is underneath the main content. The comments are presented in a flat tree structure as users may write their comments in reply to other users' comments. Furthermore, every comment can be liked or disliked. One red and one green bar indicate the respective count of each at the upper right-hand corner of every comment's box.

The most important table is the one containing data about all registered users, which allows us to establish all relationships we have in our experiments on a per-user basis. One such relationship is the *follow/ignore* relationship, which reflects the possibility on the platform, that users can follow or ignore other users. Following the interaction possibilities described above, dedicated tables are containing the votes and postings of all users, which are associated with articles (contained in the *Content* table). Additionally, articles may be associated with one or more keywords (e.g. Brexit or Obama), making it possible to group articles by particular topics.

Table 3.2 gives an overview about general statistics of the data. Users tend to vote more often than they write comments as it takes less effort than writing a couple of sentences to express one's opinion. The Pearson correlation coefficient of the number of postings and the number of votes per article yields a result of 80,49% – confirming our hypothesis, that both numbers are strongly correlated, which means that articles with a high number of comments also have a high number of votes and vice versa. The users themselves are distinguished into active (64.46%) and inactive ones (i.e. revoked their registration or similar), although it influences the data displayed on the platform itself in certain ways, we do not distinguish between them in our experiments.

| **Statistic** | **Value** |
|---|---:|
| Votes total | 318,611,187 |
| Postings total | 81,083,682 |
| Users total | 599,560 |
| Active registered users | 386,478 |
| Articles total | 1,689,912 |
| Follow relationships | |
| following | 25,624 |
| total | 37,815 |
| mean per user | 6.68 |
| 50% quantile | 2 |
| 75% quantile | 5 |
| max | 1598 |
| largest CC | 87.61% |
| Ignore relationships | |
| ignoring | 5,641 |
| total | 18,174 |
| mean per user | 7.55 |
| 50% quantile | 2 |
| 75% quantile | 5 |
| max | 296 |
| largest CC | 98.35% |

Table 3.2: Overview of the general statistics of the data

42,842 users follow or ignore another member of the community; this feature is therefore only used by a minority of users. There are also more following relationships than ignoring ones. The statistics above differentiate between ignoring/following and *total*, with the first counting users who actively follow or ignore another user. When constructing a network for each relationship, with the nodes representing users follow/ignore someone or being followed/ignored, the total number of nodes in this networks is given by *total*.

The ignore relationship shows a more substantial imbalance in this case than the following relationship. Less than a third of all users participating in this network actively ignore other users. The degree distribution follows a power-law distribution in both cases, with the majority of nodes having a degree of two or less, and vast extremes on the upper end. Here the following network shows larger extremes, with a maximum degree of 1598 compared to the maximum degree of 296 in the ignore network. Both networks consist of multiple isolated communities (connected components) which are dominated by a single largest one whose size relative to the network size is given by *largest CC*.

The experiments conducted in this thesis are not restricted to single articles only. We also want to analyse the polarization of whole topics, as this may cover different and long-term dynamics. Therefore it is necessary to analyse the comments and votes of more than one article belonging to the same topic. There are two ways how the content and articles can be organised into groups. On the one hand, they are organised hierarchically. Articles belong to so-called *ressorts* or departments, which are grouped again into broader categories, which then belong to a single channel (e.g. Web, International or Sports). Figure 3.5 shows an example for the channel sports, which consists of separate departments for soccer, tennis etc. These, in turn, cover several topics such as the national soccer team of Austria, or in the case of tennis, the French Open.

The fact that a group of articles belong to the same channel or the same department does not guarantee that they also cover the exact same topic. The most common department is *Austria* with 43.242 articles, followed by *IT-Business* with 32.842. It is difficult to imagine that all of these over 40 thousand articles are concerned with the same topic. Of course, this divergence is exaggerated when going the hierarchy up until channels, with *Economy* and *International*, both comprising over 220.000 articles.

However, articles do not only belong to a department or channel, but they are also associated with specific keywords. These keywords are based on the content of the article and are much more fine-grained than a hierarchical grouping. This should make them more suitable for grouping related articles together for analysis. Of all articles, a majority (89.09%) is classified, i.e. they are associated with at least one keyword (Table 3.3).

At first glance, the number of unique keywords seems astonishingly high, with almost as many unique keywords as articles. At a closer look, the table contains many duplicates. Recep Erdoğan, the current president of the Republic of Turkey, alone is mentioned more than 20 times with varying spellings. However, spelling alone is not the only reason for this high number; many keywords are also used in different phrases or broader contexts.

The name Obama (almost exclusively associated with Barack Hussein Obama II, 44th

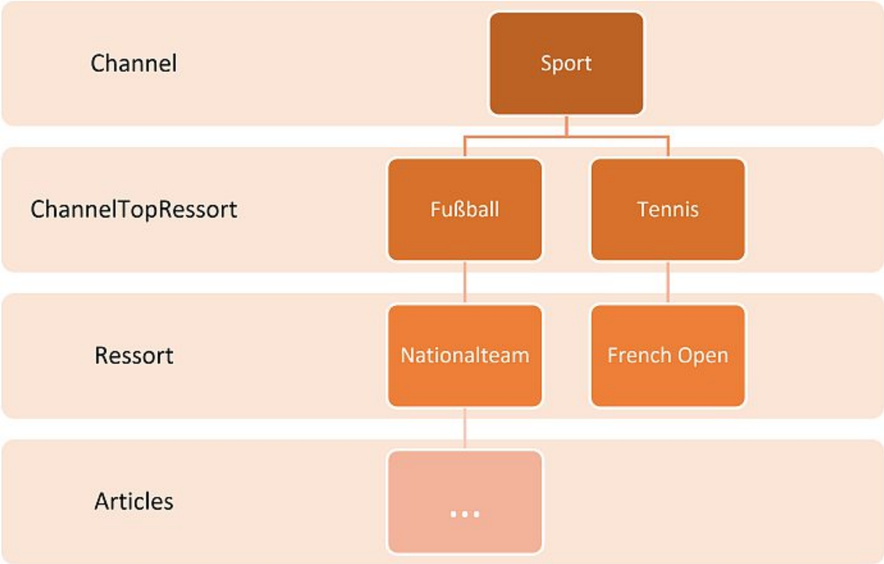



Figure 3.5: Hierarchy of content on *derStandard.at*

president of the United States) for instance, is used more than 60 times, either alone, with different spellings or contained in larger text phrases. This high entropy has to be accounted for in respective experiments, as the selection of one specific keyword exclusively may lead to related articles not being selected (false negatives).

| Statistic | Value |
|---|---|
| Articles total | 1,689,912 |
| Keywords total | 1,034,529 |
| Classified articles total | 1,505,540 |
| Classified articles (%) | 89.09% |
| Avg. keywords/article | 9 |
| Most frequently used keywords | |
| Wien | 405,090 |
| Österreich | 224,845 |
| USA | 135,361 |
| Eu | 113,047 |
| Deutschland | 110,518 |

Table 3.3: Keywords: General statistics

## 3.3 Data selection

Data selection is the first and most crucial step in our workflow, as the data selected in this stage, is the basis for all the following steps. Therefore we pursue different strategies to select relevant content, i.e. a particular topic, keyword, article, or similar. The following enumeration gives an overview of the different possible strategies:

1. Exhaustive analysis
2. Popularity
   a) Number of votes
   b) Number of comments
   c) Keywords
3. Human judgement
   a) Article metadata
   b) Keywords

The first and most straightforward strategy to implement would be the exhaustive analysis, i.e. anlaysing all content available. Only considering the potential outcome, this would be the most desirable, as it would result in a complete overview. However, today's social networks are too large for this strategy to be suitable. In our case, this would mean the analysis of over a million articles, which in turn is rather small in comparison to the content generated by billions of users on *Facebook* and similar. An exhaustive analysis, therefore, is only applicable if the size of the data is small enough, or the computational power available is high enough to complete the computation in reasonable time.

As an exhaustive strategy is not applicable, the available data must be prioritised somehow. In general, we are interested in the most popular content, which is of the most public interest and therefore has the highest negative potential if strongly polarized. There are several ways to identify popular articles or topics with the data at hand. Either by the number of votes, the number of comments or through associated keywords. The first two associate popularity with either the total number of votes or the total number of comments. Keywords open up additional ways, as these, in turn, can be associated with an average number of comments or votes per article, or a total number of articles in general.

While the first two groups of strategies are completely automatable, the third strategy follows the rationale that human judgement might be able to improve the pre-selection of potentially polarized topics. Educated guesses may be possible based on the article metadata, such as the title or the content. When considering the most popular articles (measured by the number of comments) for instance, there are many political ones in the top ten, and also an article about the *Charlie Hebdo shooting*. Although these articles are all popular, the political articles are much more likely to be polarized than the latter one, at least from a human point of view as the majority of comments underneath this article are condolence messages or express sheer bewilderment and are not part of a controversial discussion.

On the level of topics or keywords, this pre-selection might have an even greater impact. Table 3.3 references the most frequently assigned keywords, with all five being cities or countries. However, we do not expect a single one of them to be highly polarized, because on one side they are associated with too many articles, and on the other side there is hardly any reason for a city to be associated solely with strongly polarized articles.

When considering the keywords ordered by the average number of comments per article (Table 3.4), there are also ones which may be associated with polarized topics, and some which may not. “Amaq”, a news agency closely linked to the IS, for instance, is not expected to be any polarized, as we do not expect to find any sympathy for it in an Austrian online news forum. In contrast, “Halal” may be strongly polarizing, as we expect the users to be split into two sides here, supporters and opponents. The data selection process would strongly benefit from basic human judgement in these cases.

| **Keyword** | **Avg.** | **Art.** | **Description** |
|---|---|---|---|
| *Nestroyplatz* | 3322 | 5 | Place in Vienna (knife attacks 2018) |
| *Amaq* | 2961 | 6 | News agency linked to IS |
| *Wolfgang Albers* | 2730 | 8 | Cologne police president (2015-16 New Year’s Eve sexual assaults) |
| *Gleichbehandlungs-anwaltschaft* | 2096 | 10 | Discrimination of women (esp. allegations against Peter Pilz) |
| *Fatma Betül Sayan* | 1974 | 5 | Turkish minister (controversy about presidential election in Rotterdam) |
| *Halal* | 1931 | 9 | Controversies about Halal Food (esp. SPAR) |
| *Jude Ben Gurion* | 1906 | 8 | Controversy about national socialistic song-books - Udo Landbauer (FPÖ) |
| *Rouen* | 1896 | 5 | Place near the 2016 Normandy church attack |
| *Tobias Plate* | 1865 | 5 | German Interior Ministry spokesman (European migrant crisis, 2016 Ansbach bombing) |
| *Wolfgang Preiszler* | 1842 | 6 | FPÖ politician, policeman (BVT affair) |

Table 3.4: Keywords with highest average number of comments and the number of associated articles (considering only articles with more than 50 postings and keywords with at least 5 articles)

## 3.4 Network design

The data structure, as described in Section 3.2, allows us to extract a variety of different networks. These networks are different in the kind and scale of data they are created from. Table 3.5 gives an overview of these different possibilities.

The two most obvious options are based on the postings and votes underneath every article. Additionally, the data may also be aggregated on the level of keywords, to analyse

| Data | Article | Keywords | Global |
|---|---|---|---|
| Postings | x | x | - |
| Votes | x | x | - |
| Hybrid/Blended | x | x | - |
| Follow | - | - | x |
| Ignore | - | - | x |
| Follow-Ignore | - | - | x |

Table 3.5: Possible network structures

all postings or votes of one specific topic, to analyse its polarization. Besides analysing both, postings and votes, in isolation also a hybrid approach may be considered. When using both of them simultaneously to form relationships between users, both values may be simply summed up, resulting in higher edge weights, or they are combined according to a specific ratio. This ratio should resemble the value of the endorsement users expressed with their actions.

In general, a single posting would be considered more valuable in this context than a single vote, as it merely has a much higher inhibition threshold to express. On the other side, the meaning of a vote is unambiguous, with an affirmative vote expressing agreement, and a negative vote expressing disagreement. To take these factors into account it would be necessary to conduct different experiments assessing varying ratios of comments and votes, e.g. using a 1:1 ratio, i.e. summing up votes and postings, and once using a 1:2 ratio, effectively making postings two times more expressive than votes.

Although votes or even combining votes and postings might be valuable options worth considering, we leave them as a subject to further research. We follow the assumption that postings exhibit much more information about a community than votes, as voters are only secondary participants of a discussions, they do not add anything to the discussion itself, they do not add new postings nor add comments to existing ones.

On the dataset level, the follow and ignore relationships can be used to create networks. On the one hand, through using a specific type of relation in isolation, i.e. follow or ignore. On the other hand, through creating a signed network, with ignoring relations contributing negative scores, leading to potential negative edge weights.

All of the procedures described above potentially lead to disconnected graphs. A graph which is not connected contains at least two vertices which are not connected by a path, i.e. there is no set of edges connecting these vertices. Therefore the graph as a whole contains several isolated communities. In the context of polarization, the presence of isolated communities may be interpreted in two ways. Either it is interpreted as a sign of strong polarization, to such an extent that there are no relationships anymore between two communities, or as irrelevant in the context of polarization, as the absence of relations between these communities may be of any reason.

At this point, we follow the reasoning of Guerra et al. that assessing polarization between two isolated communities is not possible. It may also be that these groups do not know each other at all, which would make it impossible to reason about the polarization between these groups. It should also be noted that the measurement proposed by Guerra et al. also cannot be computed without a boundary ($B = \emptyset$) [GJCK13, p. 220].

This makes it necessary to reduce every network created in our experiments to the largest (weakly) connected component of the network. Which is also the reason we included the size of the largest connected component (CC) for all datasets used in this thesis. For all the external datasets used, the largest CC makes up over 75% of the whole network. Underneath this threshold, it may be necessary to also consider the second largest CC. If a network is divided into two large connected components, every component should be analysed in isolation.

## 3.5 Workflow

Our objective is the creation of a workflow to measure polarization, which is as automated as possible, ideally without any human interaction needed at all. The steps of our workflow, which are inspired by findings in literature ([Gar18]) and well-known processes ([Fay96]), are shown in Figure 3.6 and have been briefly described in previous sections.

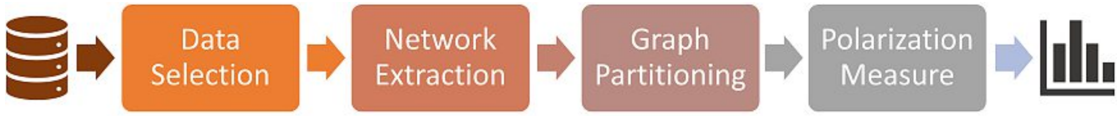



Figure 3.6: Workflow for the calculation of polarization scores

The first step of our workflow is concerned with data selection, i.e. the selection of suitable and eligible data from a much larger database. Depending on the setting, different strategies are possible in this case, varying greatly in the amount of human interaction needed (Section 3.3). The data selected in this step is then transformed into a network or graph structure, where different data may be used to form relationships (edges) between single users (vertices).

Before it is possible to apply one of the various measurements described above, the graph must be partitioned into two separate communities. Each of these communities represents one side of a discussion or a controversial topic. As this is a well-studied problem, many algorithms are applicable, though we will use the algorithm *METIS*, as it has been proven viable in previous experiments in literature.

The two resulting partitions will then be fed into the measurements described above to obtain a quantitative assessment of the polarization of the data selected in the first step. The results from this stage, however, are not universally comparable, as the possible value range is different from measurement to measurement. This means only results obtained from the same measurement are directly comparable to each other.

CHAPTER 4

# Evaluation of measurements

This chapter presents and discusses our findings from the experiments we conducted with the measurements identified in literature. In the first section, we discuss the assessment of different polarization measures in conjunction with datasets previously used in literature. We evaluate their performance in terms of differentiation of unpolarized from polarized networks and explanatory power, and terms of run-time performance. Furthermore, we evaluate the usage of UMAP for measuring polarization and compare the results we obtained for every measurement with the results from the original studies.

## 4.1 Experiment design

We evaluated the different measurements described in this thesis using a set of datasets previously used in literature, as explained in Section 3.1. This, on one side, ensures that we cover a range of different network sizes, structures and strengths of polarization, and on the other side allows us to compare our results to the scores calculated in previous experiments. Which also allows us to verify our implementation of the polarization measures, as significant divergences might be caused by implementation errors, leading to very different results.

We calculated the polarization score for every combination of measurement and dataset exactly 20 times, to mitigate any random influences on the experiments. The results obtained from these 20 runs are shown in Table 4.1. For every combination, it contains the average polarization score obtained and the respective standard deviation in the first line for every dataset. The second line contains the average time needed in seconds for a single computation. It must be noted that Polarization Index by Matakos et al. ($PI$) was calculated with a small adoption, which we explain in detail further below.

In total, we evaluate three different versions of the *Embedding Controversy* measure. *EC*, the first one is based on the original definition and uses ForceAtlas2 to create a two-dimensional layout of the network. The other two versions are based on Darwish et al. [Dar19] and use *UMAP* instead of ForceAtlas2. As it is not mentioned which parameters were used in their experiments, and as we found mainly two different parameter settings online, we originally evaluated a total of four different parameter settings – the two best of which we included in our results. One with a neighbourhood of 30 (*ECN*) and one with the target metric set to *correlation* and a neighbourhood of 30 (*ECC*). A more detailed description of the *UMAP* evaluation can be found below.

| Dataset | BCC | BC | EC | ECC | ECN | MBLB | Modul. | PI* | RWC |
|---|---|---|---|---|---|---|---|---|---|
| Soccer | 0.75/0.03 | 0.11/0 | 0.31/0 | 0.21/0.02 | 0.12/0.03 | 0.61/0 | 0.41/0 | 0.34/0 | 0.53/0 |
| | 182 | 0 | 212 | 4700 | 1818 | 1609 | 0 | 46 | 769 |
| Comp. Graph | 0/0 | 0/0 | 0/0 | 0.06/0.01 | 0.05/0.01 | 0.01/0 | 0/0 | 0.0/0 | -0.07/0 |
| | 58 | 0 | 3 | 19 | 19 | 55 | 0 | 59 | 4 |
| Conn. Graphs | 0.85/0.11 | 0.50/0 | 0.94/0 | 0.96/0.01 | 0.96/0.01 | 0.99/0 | 0.50/0 | 0.83/0 | 0.98/0 |
| | 14 | 0 | 2 | 13 | 14 | 67 | 0 | 35 | 3 |
| Gun Control | 0.84/0.01 | 0.25/0 | 0.66/0 | 0.53/0.01 | 0.48/0.03 | 0.73/0 | 0.48/0 | 0.40/0 | 0.86/0 |
| | 1314 | 0 | 440 | 14920 | 5229 | 2712 | 1 | 132 | 1278 |
| Karate | 0.36/0 | 0.19/0 | 0.45/0.11 | 0.57/0.03 | 0.55/0.02 | 0.70/0 | 0.37/0 | 0.36/0 | 0.17/0.01 |
| | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| Karate [EK10] | 0.35/0 | 0.19/0 | 0.44/0.10 | 0.57/0.01 | 0.52/0.03 | 0.70/0 | 0.36/0 | 0.36/0 | 0.17/0.01 |
| | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| NY Teams | 0.74/0.02 | 0.10/0 | 0.04/0.01 | - | - | 0.58/0 | 0.39/0 | 0.54/0 | 0.41/0 |
| | 3230 | 0 | 2286 | - | - | 7828 | 0.8 | 118 | 3407 |
| Pol. Blogs | 0.66/0.01 | 0.17/0 | 0.64/0.01 | 0.76/0.02 | 0.27/0.02 | 0.54/0 | 0.42/0 | 0.12/0 | 0.47/0 |
| | 1 | 0 | 3 | 12 | 12 | 33 | 0 | 8 | 3 |
| Univ. Friends | 0.77/0.01 | 0.15/0 | 0.44/0.10 | 0.52/0.04 | 0.20/0.01 | 0.15/0 | 0.34/0 | 0.14/0 | 0.13/0 |
| | 0 | 0 | 0 | 2 | 1 | 16 | 0 | 2 | 0 |

Table 4.1: Results for datasets from literature. Every cell contains the mean value and the standard deviation separated by "/". The first row for every dataset contains the polarization scores, whereas the second one contains the time needed per calculation.

## 4.2 Polarization score comparison

Figure 4.1 illustrates the information of the table above. The translucent band around each line shows the standard deviation for a specific measure-dataset combination. Both, the table above and the respective illustration in Figure 4.1 contain missing values for the *UMAP* variants of Embedding Controversy *(ECC, ECN)* in the case of the *NY Teams* dataset. This is due to the fact, that the implementation provided by the original authors consumes that much memory in the case of larger datasets (in this case around 100,000 nodes), that we could not even compute it on machines with 128GB RAM. Therefore we had to omit these two specific cases in our evaluation.

As we expected, the lowest score for all measurements was achieved for the *Complete Graph* dataset, which acts as a synthetic example for the complete absence of polarization. The opposite extreme of perfect polarization (*Conn. Graphs*) showed that using the right stopping criterion in our implementation is crucial. Initially, *MBLB* failed this sanity check, yielding higher polarization scores for almost every other dataset. Although this was a contradiction to the formal definition of the measurement, that polarization is highest if two communities of the same size have opposite opinions.

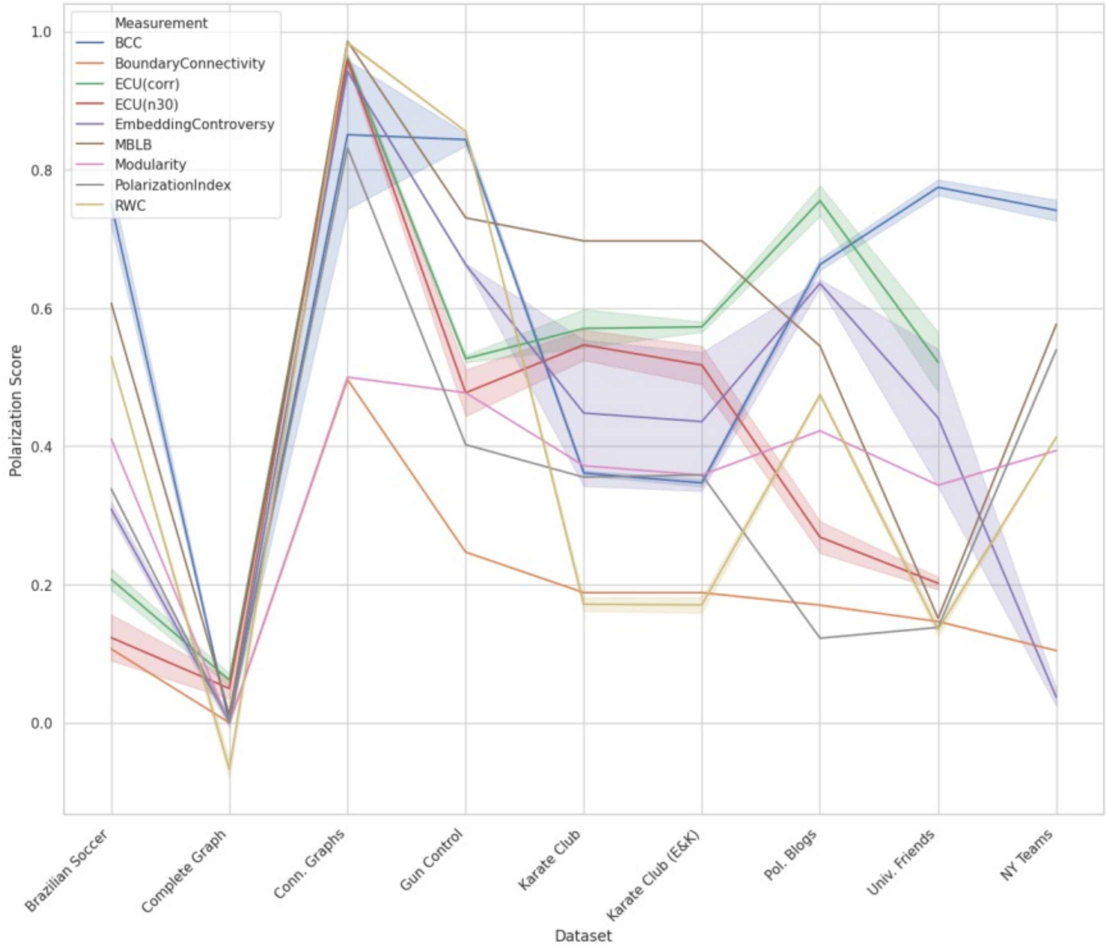



Figure 4.1: Polarization scores for combinations of measurements and datasets from literature.

In the case of two complete graphs, connected to each other through a couple of edges, the calculation does not converge in a linear way or at least decreases monotonically. In the opinion generation process for *MBLB*, the nodes which connect both communities are chosen as *elite*-nodes, as they have a higher degree than the rest of the nodes (as they have one incoming edge more than every other node in the respective complete graph). Therefore the polarization of every other node in the network is only influenced by nodes with similar polarization (as all connections to the other community are cut off by the *elite*-nodes not propagating this opposing opinion). As a consequence, the opinion generation process yields the highest possible polarization in the network. Our original implementation, however, aborted the calculation at a score of 0.65 as the network did not converge, i.e. the number of nodes still changing their polarization was constant from the start. Making the threshold more sensible and extending the iterations needed for early stopping then resulted in a more plausible score of 0.99.

For our evaluation, we adopted the definition of the *Polarization Index (PI)* from Matakos et al. [MTT17]. With the original definition, the index produces very small values. In the original paper, the highest value obtained was 0.107. This would make a direct graphical comparison difficult. Therefore we calculate the opinion vector in the same way as defined by the authors. However, we do not square the opinion vector before normalising it. This leads to the index producing results in a range from zero to one without changing the general outcome of the measure, i.e. the measurement still distinguishes unpolarized from polarized topics in the same way, although utilising a different value range.

It should be noted though, that the measurements are not comparable, meaning that a score of 0.5 for *BC* has a different meaning than a score of 0.5 for *RWC*. While some measurements have slightly different value ranges, some even may result in negative values; others tend to produce more extreme values. Therefore the results are only comparable among different datasets but holding the polarization measure fixed except for *Modularity*. As already described above, *Modularity* scores are only comparable among different networks of roughly the same size. It is therefore not advisable to directly compare the score for NY Teams with the score for Zachary's karate club for example.

We also use correlation coefficients to analyse the similarities between the different measurements and datasets. These are shown in the heat-map in Figure 4.2. All coefficients are positive and in a range from 0.34 to 0.94. The highest correlation scores are found between *BC*, *EC* (and its variants *ECC* and *ECN*) and *MBLB*, which is also confirmed by the visual representation of the results in Figure 4.1. Whereas *BCC* is the measurement with the lowest absolute and average correlation coefficient.

Overall the average correlation coefficients between the measurements point towards a high similarity between the various measurements. We excluded *BCC* in this stage of our analysis, due to the low correlation with the other measures and other reasons we will explain further down below. As shown in Table 4.2, the mean correlation coefficients are between 0.71 and 0.82. The mean is calculated from all pairwise calculated correlation coefficients for that measurement with all other ones, i.e. without considering the correlation of a measurement with itself. Note, that this leads to small differences in numbers between Figure 4.2 and Table 4.2.

| | **BC** | **EC** | **ECC** | **ECN** | **MBLB** | **Modularity** | **PI** | **RWC** |
|---|---|---|---|---|---|---|---|---|
| **mean** | 0.82 | 0.74 | 0.75 | 0.79 | 0.77 | 0.72 | 0.73 | 0.71 |
| **std** | 0.08 | 0.15 | 0.12 | 0.13 | 0.07 | 0.07 | 0.14 | 0.08 |
| **min** | 0.69 | 0.49 | 0.58 | 0.61 | 0.67 | 0.61 | 0.49 | 0.58 |
| **25%** | 0.78 | 0.69 | 0.67 | 0.71 | 0.72 | 0.68 | 0.66 | 0.66 |
| **50%** | 0.80 | 0.71 | 0.73 | 0.82 | 0.78 | 0.71 | 0.74 | 0.74 |
| **75%** | 0.88 | 0.85 | 0.84 | 0.87 | 0.83 | 0.76 | 0.82 | 0.77 |
| **max** | 0.94 | 0.93 | 0.93 | 0.94 | 0.84 | 0.82 | 0.90 | 0.78 |

Table 4.2: Correlation Coefficients - Statistics

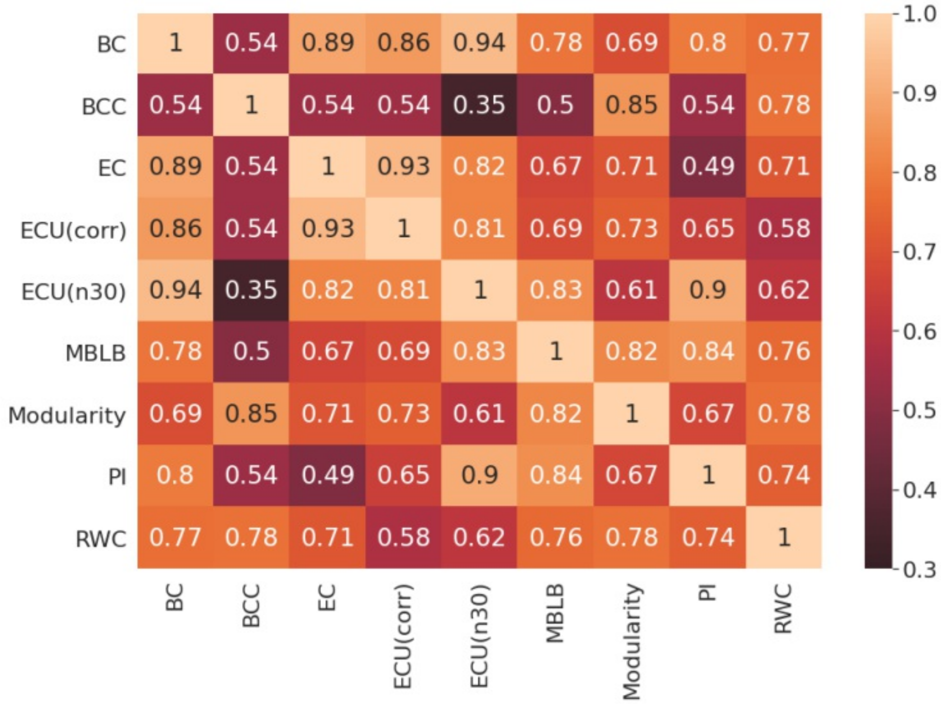



Figure 4.2: Correlation heat-map

The lowest coefficients are found between *EC* and *PI*, followed by *ECC* and *RWC*. *RWC* is also the measurement with the lowest average correlation coefficient. However, even the lowest correlation coefficient, 0.49, means that they are moderately correlated to each other. So there is not a single pair of measurements, which would be only weakly or even negatively correlated to each other. There is de facto no polarization measure which produces contradicting results. The highest similarity was found between *BC* and *ECN*, followed by *EC* and *ECC*.

*BC* is, in general, the measurement with the highest average correlation coefficients, which can be interpreted as *BC* having the highest similarity with all the other measurements. In contrast to our initial presumptions, the measurements are very similar to each other, without clear patterns or contexts in which some measurements fail, while others shine. All measurements except BCC seem to deliver similar results for the different datasets. Therefore all measurements seem viable in analysing and quantifying polarization in social networks. Moreover, out of all these, *BC* seems to be the safest option, being some kind of common denominator.

The reason for *BCC* not being compliant with the other measures does not lie in the reasoning behind its definition. Using betweenness centrality to quantify the polarization of a network can be considered a reasonable approach. The assumption that the cut between two communities in polarized networks has a much higher betweenness centrality than the other parts of the graph is reasonable too. The problem is the second part of its mathematical definition. After sampling the betweenness centrality using kernel density estimation to compute the probability density function, the Kullback-Leibler divergence is used to quantify the difference between both distributions, which is then used to calculate the final score $BCC = 1 - e^{-d_{KL}}$ (see Section 2.4.2).

| **Metric** | **Conn. Graphs** | **Brazilian Soccer** |
|---|---|---|
| Avg. bc (all) | 0.00002 | 0.00004 |
| Avg. bc (cut) | 0.03339 | 0.00009 |
| Std. bc (all) | 0.00043 | 0.00013 |
| Std. bc (cut) | 0.00000 | 0.00035 |
| Avg. Entropy | 1.99777 | 1.39817 |
| Avg. Score | 0.82561 | 0.75174 |

Table 4.3: Comparison of *Conn. Graphs* and *Brazilian Soccer* for *BCC*

For our explanation, we take the two datasets *Conn. Graphs* and *Brazilian Soccer*. For the latter *BCC* reports a score of 0.75, much higher than any other measure. Table 4.3 contains the average values and standard deviation of the betweenness centrality for all edges of the graph and for all edges in the cut. Followed by the average entropy for both datasets and the calculated polarization score.

The difference between the centralities in the cut, in comparison to the centralities of all edges, is much higher for *Conn. Graphs* than for *Brazilian Soccer*. In the first case, it is over 1,600 times higher and only two times as high in the second. Despite this significant difference, the final entropy and polarization scores are lying much closer together. Because of this *BCC* fails to differentiate between almost completely polarized networks and networks which are only little to moderately polarized.

With a mean correlation coefficient of 0.73, *PI* is slightly below the average, performing better than we assumed initially. The *Polarization Index* is the length of the calculated opinion vector under the $L^2$ norm $||z||$, normalised by the number of nodes in the network (see Section 2.4.5). The relationship between the distribution of opinions and the resulting score is not linear though. To illustrate this, we repeatedly create an opinion vector, with one half of the vector containing the respective opinion value, while the other half contains the negative opinion value, representing two different communities with opposing opinions. Performing this experiment for opinion values from zero to one yields the scores seen in Figure 4.3 - showing an exponential instead of a linear relationship.

The score is quickly dropping with every decrement of the opinion value while at the same time, the slope is much flatter for low opinion values. A rather high opinion value of 0.6, for instance only yields a score of 0.36, which would mean the network is only slightly polarized. Whereas an opinion value of 0.85 yields a score of 0.7225 - more than twice as high. This means that already moderately high scores, when compared to other measures' value range, indicate strong polarization in the case of *PI*.

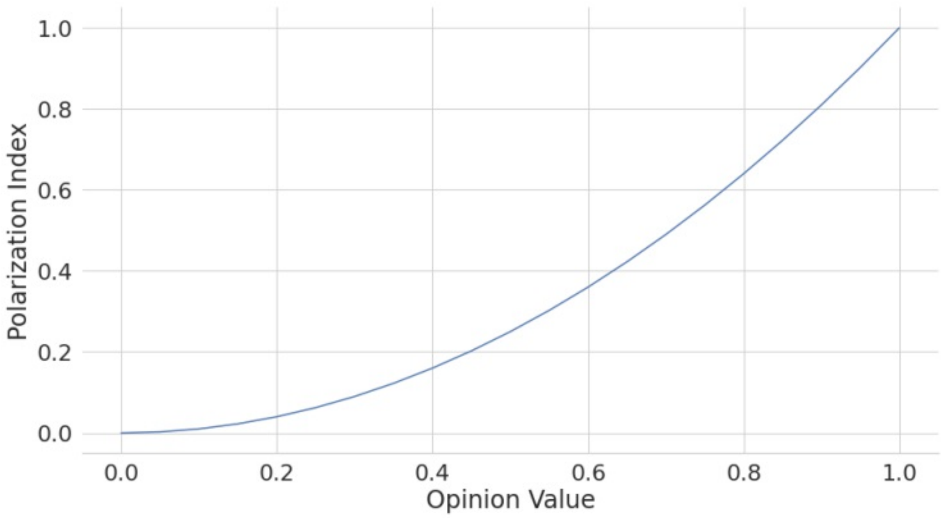



Figure 4.3: Relationship of opinion values and polarization index

## 4.3 Standard deviation

The standard deviation of the results produced by a polarization measure can be used to assess the reliability of it. Ideally, a measure produces the exact same polarization score with every calculation, given the same dataset. However, this is not the case for measures based on random walks or other varying factors such as layout algorithms, which produce a slightly different layout with every new run.

The reliability of a measure is vital in the sense that in the case of a high standard deviation, one must calculate the polarization score several times to get the average score. Otherwise, the score might be significantly higher or lower than the average, rendering one single result utterly useless. For exact polarization measures, however, the score only has to be calculated once. This, in turn, influences the measurement selection, as multiple runs of an already slow measurement may be insufficient for some use cases like near real-time calculations for applications such as interactive user interfaces.

As shown in Figure 4.4 most of the combinations always yield the same result with every run, i.e. the standard deviation is 0, or produce only minimal variations. Overall there are only four combinations with a standard deviation of 0.05 or higher. *BCC* and the various *Embedding Controversy* variants *EC*, *ECC* and *ECN* exhibit the largest standard deviations overall. With the largest ones for *EC* with 0.11 for Karate Club and 0.10 for Karate Club [EK10] and University Friendships' Network and *BCC* for *Conn. Graphs*. *EC*, therefore, seems to become more unreliable the smaller the network, with deviations of 10% and more into both directions. The high standard deviation for *BCC* is caused by the sampling of betweenness centrality values using kernel density estimation, as already described above.

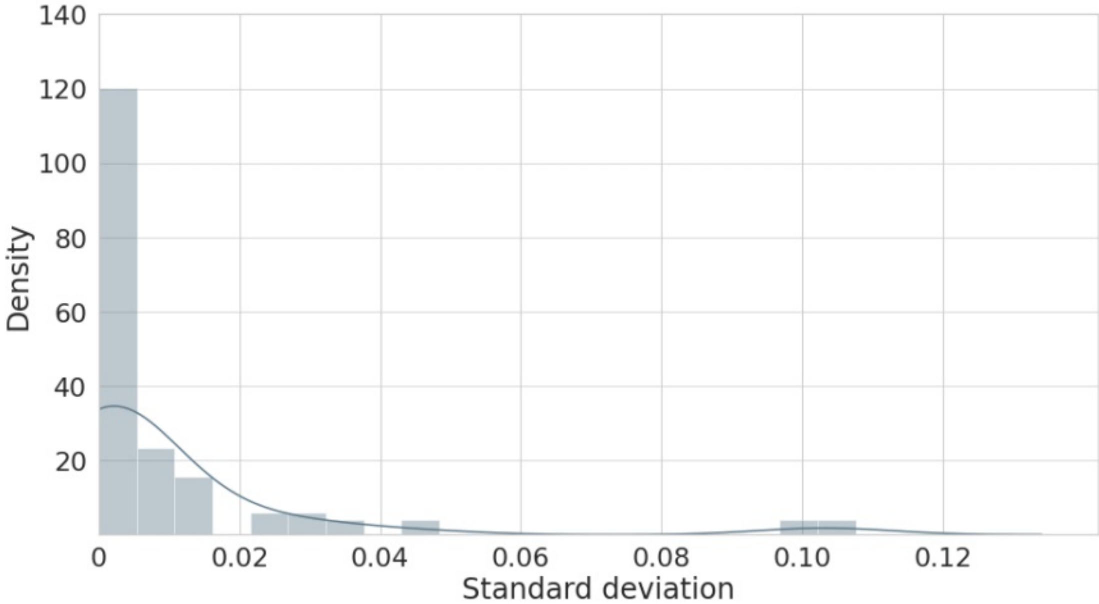



Figure 4.4: Distribution of standard deviation across all combinations

There is one other measure we recorded small deviations for, *Random Walk Controversy (RWC)*. However, similar to the *EC* variants and most of the cases for *BCC*, the standard deviation is minimal, with values smaller than 0.05. *RWC*, similar to *Embedding Controversy (EC)*, tends to become unreliable on very small networks. However, it is more reliable in terms of average standard deviation than any of the *EC* variants.

We highly recommend to conduct several runs and average the result for the more unreliable measurements. Although even for the various *EC* variants a standard deviation of 0.05 and more usually does not drastically influence the general statement about the polarization of a specific dataset. But it could, which is precisely the case for the Karate Club dataset, which due to the high standard deviation might be classified as highly polarized or as rather unpolarized, depending on if the calculated score was exceptionally high or low.

## 4.4 Evaluation of UMAP

Darwish et al. were the first using *UMAP* as graph layout algorithm for *Embedding Controversy*, arguing that it is more aggressive than Force Atlas 2 or similar force-directed layout algorithms, which they confirmed with their results showing consistently higher polarization scores when using *UMAP* than with FA2. However, they do not state their parameter settings used in their experiments. Which is quite unfortunate, as *UMAP* has in total 23 changeable parameters leaving much space for deviations due to wrong or different configurations [Dar19].

For our experiments, we use the library created by the original authors of *UMAP*[1]. In general, we keep the default settings, which is consistent with experiments found in the official documentation and similar experiments found online[2,3,4]. Accordingly, we focus on two parameters that we will try with different settings, *n_neighbours* and *metric*.

The *n_neighbours* parameter controls how *UMAP* balances local versus global structure in the data by constraining the size of the neighbourhood the algorithm will look at. For low values, *UMAP* will concentrate on very local structures, while it tries to concentrate on the big picture for large neighbourhoods. For the analysis of polarized networks, we focus more on the big picture, than on local phenomena, which potentially may be to the detriment of the big picture. We assume that a larger neighbourhood leads to a better understanding of the network structure and the inherent polarization. Therefore we assume that choosing a larger value (30) than the default (15) will improve results.

The second parameter we focus on, *metric*, controls how distance is computed in the ambient space of the input data. By default, *UMAP* uses the Euclidean distance or Euclidean metric. In total *UMAP* supports 22 different metrics for very different use cases. Out of them, *correlation* is most often used in the context of graph and network analysis. Therefore we compare this to the default setting.

| **Dataset** | **ECU(corr,n15)** | **ECU(corr,n30)** | **ECU(eucl,n15)** | **ECU(eucl,n30)** |
|---|---|---|---|---|
| Soccer | 0.21/0.01 | 0.21/0.02 | 0.09/0.01 | 0.12/0.03 |
| | 2890 | 4700 | 1701 | 1818 |
| Comp. Graph | 0.07/0.01 | 0.06/0.01 | 0.07/0.01 | 0.05/0.01 |
| | 13 | 19 | 13 | 19 |
| Conn. Graphs | 0.94/0.01 | 0.96/0.01 | 0.96/0 | 0.96/0.01 |
| | 7 | 13 | 8 | 14 |
| Gun Control | 0.53/0.03 | 0.53/0.01 | 0.36/0 | 0.48/0.03 |
| | 8374 | 14920 | 4879 | 5229 |
| Karate | 0.72/0.07 | 0.57/0.03 | 0.69/0.01 | 0.55/0.02 |
| | 1 | 1 | 0 | 0 |
| Karate [EK10] | 0.79/0.03 | 0.57/0.01 | 0.64/0.01 | 0.52/0.03 |
| | 1 | 1 | 0 | 0 |
| Pol. Blogs | 0.66/0.01 | 0.76/0.02 | 0.35/0 | 0.27/0.02 |
| | 11 | 12 | 10 | 12 |
| Univ. Friends | 0.44 / 0.06 | 0.52/0.04 | 0.25/0.04 | 0.20/0.01 |
| | 1 | 2 | 1 | 1 |

Table 4.4: Results for datasets from literature for all 4 *UMAP* parameter settings.

We compared these four different settings as described above by calculating the polarization scores with *EC* for every dataset. The results are given in Table 4.4. All four measures lead to different polarization values for every dataset, except for the two

[1]Implementation in Python by the original authors: `https://umap-learn.readthedocs.io`
[2]`github.com/annaaniol/evolNET`
[3]`github.com/lmcinnes/umap`
[4]`pair-code.github.io/understanding-umap`

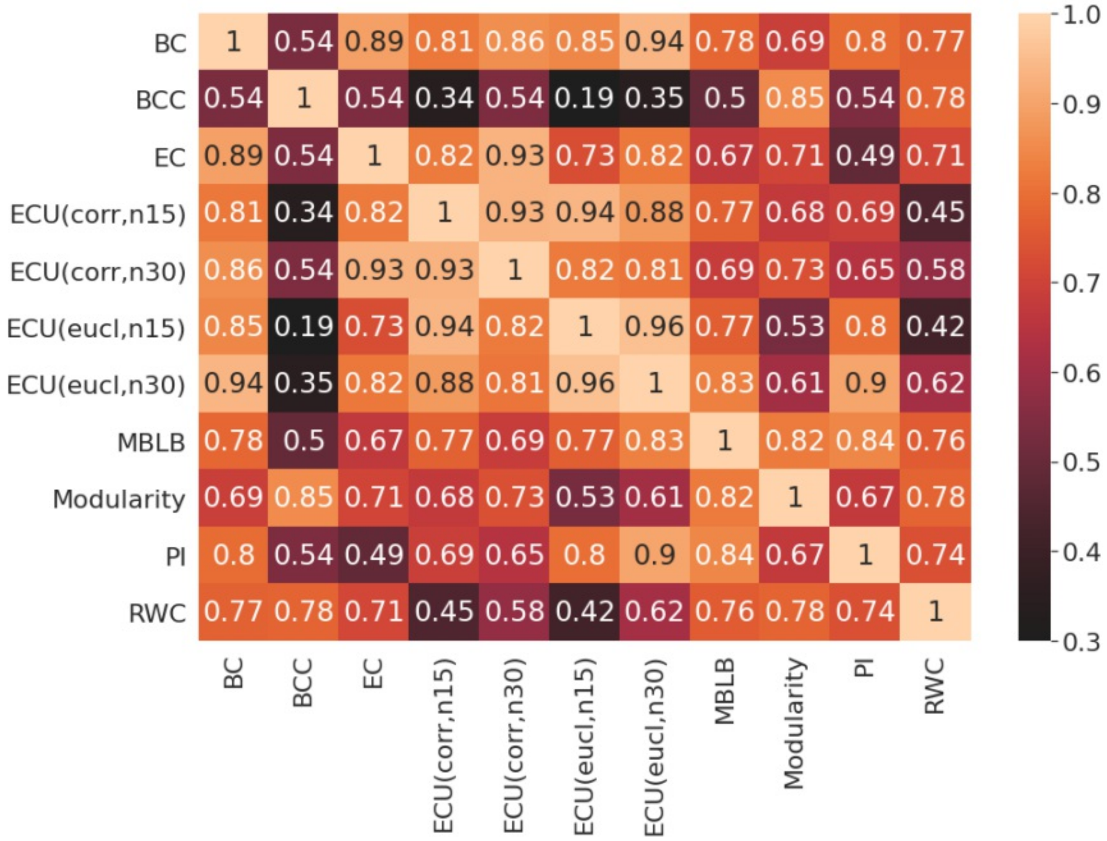



Figure 4.5: Correlation heat-map including all parameter settings for *UMAP*

synthetic cases. Here the differences are negligible. A larger neighbourhood also seems counterproductive in the case of tiny networks such as Karate Club. A neighbourhood of 30 means that *UMAP* has to consider the whole graph, which is in contrast to the presence of two distinct communities with 17 nodes each. In terms of run time, it seems that using Euclidean distance is much faster than using correlation, although a larger neighbourhood also increases run time in both cases.

Following the approach to compare different polarization measures as described previously, we use correlation coefficients to determine which polarization measures are able to quantify polarization correctly. Note that in this case, picking the measurements yielding the highest scores would be detrimental. Our objective is not to find a measurement producing the highest polarization scores for every dataset, but the most appropriate ones. In this case, this would mean that the scores should correlate with the findings from other measurements.

All correlations coefficients, including the four different *UMAP* parameter settings, are shown in the heat-map in Figure 4.5. *ECU(eucl,n30)* has the highest correlation coefficients, and also the highest average correlation coefficient (0.67), calculated based on the correlation coefficients from Figure 4.5, which is also the third-highest overall after *BC* (0.74) and MBLB (0.69).

We observed the second-highest average correlation coefficient for *ECU(corr,n30)* with 0.63. Although much smaller than the other *UMAP* variant, it exhibits a higher similarity to *BCC*, *EC* and *Modularity*. It seems therefore, that the neighbourhood is more important for the polarization analysis of a network than the metric. This confirms our initial assumption that the big picture and structure as a whole is more important than local structures.

The drawbacks of using *UMAP* instead of FA2 for *Embedding Controversy* are excessive memory consumption and the long run time. We were not able to calculate the scores for the NY teams dataset consisting of around 100.000 nodes as even machines with 128GB RAM were too small. This problem might be fixed in the near future as upcoming releases of the library will come with settings to reduce memory consumption ($low\_memory = true$). Although we needed several hours to calculate the polarization score for large graphs, this is not directly relevant to our experiments, as we are only dealing with networks which consist of only a couple of thousand nodes.

## 4.5 Comparison to previous studies

Several studies have conducted experiments with some of the measurements and datasets we use in this thesis. This allows us to compare our results to the ones obtained in previous experiments. A close match of our results with the ones from literature would mean that our implementation at least produces the same result as previous ones. Not all combinations have already been studied in previous experiments. The cells in Table 4.5 therefore only contain a value if the respective combination was analysed in a previous study.

| | Garimella | | | | | Guerra | | Matakos |
|---|---|---|---|---|---|---|---|---|
| **Dataset** | **BCC** | **BC** | **EC** | **MBLB** | **RWC** | **BC** | **Modul.** | **PI** |
| Soccer | 0.09 | -0.06 | -0.37 | -0.14 | -0.14 | -0.09 | 0.02 | - |
| Gun Control | 0.17 | 0.26 | 0.11 | -0.08 | 0.16 | - | - | - |
| Karate Club | -0.28 | 0.02 | -0.06 | 0.59 | 0.06 | - | - | - |
| Karate Club [EK10] | - | - | - | - | - | 0.02 | 0.01 | 0.014 |
| NY Teams | 0.51 | 0.09 | -0.13 | 0.39 | 0.07 | 0.10 | 0.24 | - |
| Pol. Blogs | 0.13 | -0.01 | 0.15 | 0.09 | 0.05 | -0.01 | 0.00 | -0.018 |
| Univ. Friends | 0.52 | 0.14 | 0.06 | -0.12 | -0.22 | 0.39 | 0.10 | - |
| **MAD** | 0.28 | 0.10 | 0.15 | 0.24 | 0.12 | 0.12 | 0.07 | 0.02 |

Table 4.5: Difference of calculated polarization scores to ones stated in original papers

Unfortunately, the results we got from our experiments are sometimes drastically different from the ones recorded in previous studies. The differences recorded have a *mean absolute deviation (MAD)* of 0.16 and a standard deviation of 0.21 (calculated based on the results shown in table 4.5). Considering a basic polarization scale from 0 to 1, these are unsatisfyingly high values. However, there are several reasons for the deviations of our values from the ones achieved previously.

*BCC* shows the largest *MAD* of 0.28 and two of the three largest deviations overall. We already discussed the fundamental problems of *BCC* above, that make it hard to reproduce the results of the original paper. Furthermore, the implementation found on GitHub[5] is incomplete and non-functioning. When excluding the results of *BCC* from the measures above the standard deviation is reduced to 0.18 and the *MAD* to 0.13.

Garimella et al. pointed out, that both *MBLB* and *RWC* have problems with small graphs, indicating that both are not working as expected or even delivering wrong results for small graphs such as Karate Club. We can only confirm this for *RWC*, resulting in a score far below the average. For *MBLB* however, we cannot, as it behaves in accordance with the other measurements. The deviation of our results for *MBLB* and the original paper is therefore partly justified.

Both *MBLB* and *RWC*, are strongly influenced by the choice of an initial set of special nodes. For *RWC* these nodes represent a specific partition, effectively stopping every random walk arriving at such a node. In the case of *MBLB*, such nodes also represent a partition by having a fixed extreme opinion $\pm 1$, which cannot be changed during the opinion initialisation phase of the measurement. The choice and especially the amount of the nodes chosen is crucial. The more nodes are selected, the higher the potential polarization in the end. As this means, that more random walks end in the same partition in the case of *RWC*, or the expressed opinions will be more uniform inside the same partition in the case of *MBLB*.

This is in the end also a drawback of using either *MBLB* or *RWC*, as they are influenced and dependent on the amount and kind of nodes selected. The resulting polarization score can therefore be manipulated. Selecting fewer nodes than would be ideal, results in lower polarization scores and selecting more nodes results in higher polarization scores. Although the "ideal" number of nodes is usually unknown, yet it also depends on the dataset. Making it hard to justify the number of nodes chosen.

The percentage of nodes chosen is unfortunately unknown for many of the experiments conducted in literature. Moreover, in the rare cases it is known, the statements are contradictory. In [GMGM15, p.14] Garimella et al. cites Morales et al. [MBLB15], selecting the top-5% highest-degree nodes, although Morales et al. explicitly state their choice only once in the original paper and only for one specific dataset (0.02% [MBLB15, p.5]). Therefore it is not possible for us to reproduce any results for *MBLB* or *RWC*. Moreover, we were forced to come up with our own node-selection procedure.

Many of the social networks we use in our experiments are scale-free networks. A scale-free network is a network whose degree distribution follows a power law, i.e. most of the nodes have very small degrees, while a small minority have many more connections [BA99]. Our procedure for selecting an appropriate number of nodes should therefore be

[5] `https://github.com/gvrkiran/controversy-detection`

based on this fact, yet should also select a number small enough to not automatically lead to high polarization scores for large networks.

Choosing a simple function was among our top priorities. The function selects 0.02% of the nodes until reaching a threshold at 2000 nodes. From here only 0.01% of all additional nodes are chosen, until reaching a second threshold at 10,000 nodes. From here, only 0.005% of all additional nodes are chosen. The outcome is the percentage of the highest-degree nodes that should be selected used in conjunction with the ceiling function to guarantee that at least one node is selected. This procedure is based on the node percentages chosen in literature (see above), although even 0.02% turned out to be too much for larger networks, which automatically led to high polarization score values. Therefore an even smaller fraction of the nodes is chosen in the case of large networks.

Listing 4.1: Determination of percentage.

```
def get_node_percentage(number_of_nodes):
    if number_of_nodes >= 10000:
        percent_n = 2000 * 0.02 +
                    8000 * 0.01 +
                    (number_of_nodes - 10000) * 0.005
        return percent_n / number_of_nodes
    elif number_of_nodes >= 2000:
        percent_n = 2000 * 0.02 +
                    (number_of_nodes - 2000) * 0.01
        return percent_n / number_of_nodes
    else:
        return 0.02
```

Another reason why we are not able to accurately reproduce the results of previous studies is the unknown partitioning of the datasets. Many papers only mention the rationale behind the partitioning used but do not publish the datasets and the partitioning used. Another problem is using *METIS* for partitioning, as it does not produce the exact same partitions with every run, but with small variations. This may also explain the difference of our results to the ones obtained by Garimella et al. and Guerra et al. for *BC*. We registered almost the same variations to both authors, though it is entirely off for the Gun Control dataset for Garimella et al.

The unknown partitioning also explains the deviations for *Modularity* which is among the most stable measurements. As the partitioning is only known for the datasets Karate Club and Political Blogs, we got different results for the other datasets here. Although this also does not explain the 0.01 difference for Karate Club, as the partitioning is known, the dataset is very small, and we are using a standard library here. We assume the deviation is caused by rounding errors.

Besides these points, there are some factors causing deviations we are unable to explain. One of these are the deviations for *BC*. The implementation of *BC* is directly based on the implementation provided by Garimella et al. and our results show consistent deviations to both experiments, the ones from Garimella et al. and Guerra et al., except for the *Univ. Friends* dataset. We assume that the unknown partitioning of the datasets is again the root cause here.

## 4.6 Runtime comparison

To mitigate the effect of any random variations of the run-time, we executed every combination a total of 20 times. The experiments were conducted on dedicated VM instances (n2-standard-8 - 8 vCPUs, 32 GB memory) on the Google Cloud Platform. On every machine used, a single Python program was started calculating the scores for all combinations, using 8 processes in parallel to utilise the eight cores available in an optimal way.

However, the results obtained (Figure 4.6) are not representative or directly comparable because of several factors. The biggest of which lies in the implementation details of the libraries used and the Python ecosystem itself. Libraries which are implemented directly in C and are only called from Python tend to be magnitudes faster than others. An example of this is the implementation of *Modularity*. While *networkx* is faster than self-implementing it, *iGraph* again is magnitudes faster than *networkx*. Furthermore, we did hardly any parameter tuning, e.g. *RWC* is calculated 10,000 times, regardless of the fact that small networks converge faster than larger ones. Therefore it is not advisable to compare the run-time across different measurements. This would require implementing all measurements under similar conditions, with similar libraries followed by dedicated optimizations.

Almost all of our implementations may be further optimised by utilising caching and optimisation strategies such as early stopping, which was partly implemented for *MBLB*. Also, the number of iterations for some measurements is only loosely coupled to the network size, e.g. 500 iterations for Karate Club may be far too much and only increase run time unnecessarily. Therefore we do not advise to solely rely on the run-time metrics obtained in our experiments to choose a measurement but only identify major trends. There may be much room for improvement in that direction, though this is beyond the scope of this thesis.

It should also be noted that for some of the algorithms, a more efficient version was proposed in literature. For *RWC* a variant was published by the original authors, which is defined as "random walk with restart". If certain conditions are met during the computation, the random walk restarts at a random location, making the computation more efficient [GMGM15]. Darwish et al. improved the performance of *EC* by repeatedly sampling a subset of the dataset and calculating the average of the results. This approximation of the overall score allowed for a faster calculation [Dar19].

Although we cannot draw conclusions from comparing the run-times across different measurements, we can to identify major trends. The more nodes and edges a graph has, the more time is needed to calculate the various polarization measures. For most of

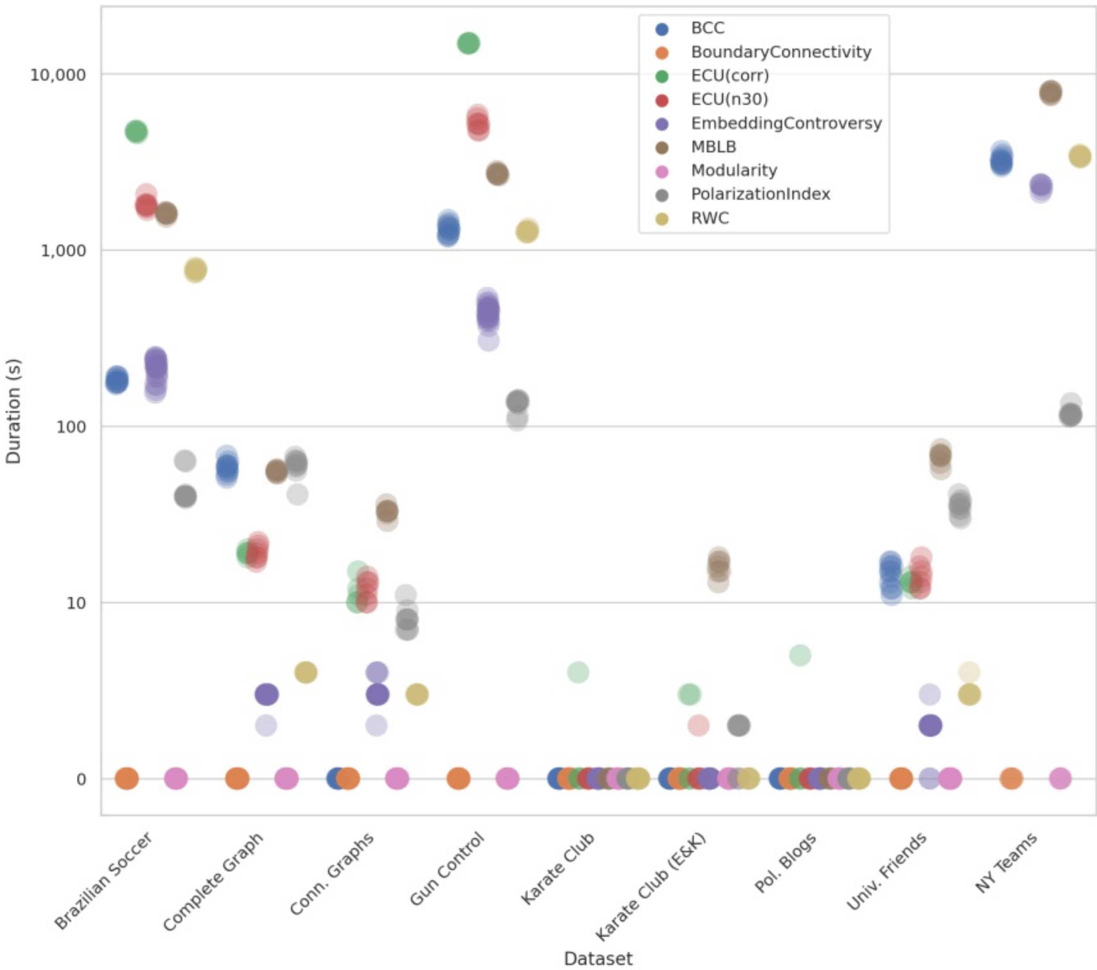



Figure 4.6: Strip plot: Runtime in seconds for all measurement-dataset combinations

the measurements, the run-time disproportionately increases with the number of nodes and edges in the network. Especially the run-time for *EC* increases with the number of nodes in a network, as the pair-wise distances for all pairs of nodes must be calculated. Nevertheless, this also holds for measurements relying on random walks, as they not only increase in number but also take longer till they reach the measurement-specific stopping criteria in larger graphs (e.g. reaching a high-degree node which belongs to the other community).

As the run-time depends on the size of the graph, this has to be considered when choosing a measurement. For small networks with a couple of hundred nodes, similar to *Univ. Friends*, this is less critical than for large networks with 100,000 nodes and more. As the datasets we extracted have only a couple of hundred or thousand nodes, we do not consider run-time a strongly limiting factor in the selection process for suitable polarization measurements. We therefore also haven't invested much time into optimising our implementations in regard to run-time.

CHAPTER 5

# Networks

In this chapter, we use the data from *derStandard.at* to extract and evaluate different kinds of networks from it. We use the postings and their replies to form a network to calculate the respective polarization scores - one for every polarization measurement. Then we aggregate the data by taking all related articles and construct a combined network per topic, to then analyze polarization on a higher level of data. Finally, we calculate the polarization for the follow and ignore relationships, i.e. users following or ignoring other users.

## 5.1 Comments/Postings

As the total number of postings is a good proxy for popularity and public interest, we first take a look at the most popular articles, i.e. the articles with the highest number of postings. These 5 articles are shown in Table 5.1. All five articles belong to two topics: politics and terrorism. Although the first two articles have over 12,000 comments, the number quickly drops below 10,000, showing that only a very small number of articles are associated with a very high number of postings. Also the largest connected component for every network makes up the majority of nodes in every case, i.e. the ratio of the size of the largest connected component to the total network size is over 0.95 in every case.

Although these articles are the ones of the most interest, they are slightly polarized at most. Therefore the number of postings, the popularity and controversy associated with a certain article is not directly linked to polarization. Besides the first and fourth article, there are not even two opposing sides, as the majority of users shares the same opinion (e.g. condolences in case of the third and fifth article). And even in the case of the political articles, which are usually very controverse, the majority of users shares the same opinion. The reason for this is partly due to *Der Standard* being a rather liberal newspaper, attracting people mostly sharing the same liberal point of views.

| Postings | \|CC\| | ratio | BC | EC | ECC | ECN | MBLB | Modul. | PI | RWC |
|---|---|---|---|---|---|---|---|---|---|---|
| Vorwürfe wegen sexueller Belästigung: Peter Pilz tritt zurück | | | | | | | | | | |
| 12244 | 2495 | 0.98 | 0.05 | 0.15 | 0.09 | 0.00 | 0.17 | 0.32 | 0.31 | 0.14 |
| Strache soll Staatsaufträge für Wahlkampfspenden in Aussicht gestellt haben | | | | | | | | | | |
| 12042 | 2414 | 0.97 | 0.07 | 0.09 | 0.04 | 0.02 | 0.33 | 0.32 | 0.33 | 0.12 |
| Messerattacken in Wien: 67-Jähriger nicht mehr in Lebensgefahr | | | | | | | | | | |
| 9326 | 1942 | 0.99 | 0.02 | 0.15 | 0.03 | 0.03 | 0.27 | 0.30 | 0.24 | 0.08 |
| Knalleffekt: SP-Bundesgeschäftsführer Niedermühlbichler tritt zurück | | | | | | | | | | |
| 8748 | 1569 | 0.99 | -0.01 | 0.02 | 0.02 | 0.01 | 0.29 | 0.27 | 0.25 | 0.04 |
| Geiselnahme in Kirche in Nordfrankreich: Ein Attentäter trug Fußfessel | | | | | | | | | | |
| 8746 | 2033 | 0.99 | 0.03 | 0.22 | 0.15 | 0.03 | 0.27 | 0.31 | 0.27 | 0.07 |

Table 5.1: Results - articles with the most postings

To gather more empirical data, we execute the polarization measurement workflow for the 5,000 articles with the most postings. This allows us to draw conclusions from a large dataset and makes it possible to identify trends and correlations in the data. One of the more apparent findings is, that the size of the retrieved component, i.e. connected users, is directly related to the total number of postings underneath the same article. The respective correlation coefficient is 0.94. Although this might not seem obvious when considering the results in Table 5.1, as here these two indicators are not directly proportional, as here a larger number of postings does not necessarily mean that more users are involved in the discussion.

As a next step, we took the five articles with the highest achieved polarization score for every measurement, of which the results are shown in Table 5.2. The majority of article titles is clearly controverse, with many political topics. Some measurements produce similar results, with some articles being in the top five of three or more different measurements.

The articles shown in Table 5.2 cover a broad range of different topics, from politics over social media and abortion to migration. All these topics are well-known for being controversial and polarizing. As these articles are the ones with the highest polarization scores out of all articles, they are concerned with exceptionally prominent political controversies. The job change of the former federal spokeswoman of the Austrian Green Party to the gambling company Novomatic (8) or the challenge of the election in Lower Austria by the Green Party (10) or a controversy about far-right politicians condemning Nazi crimes (20) are only some examples.

| | Title | BC | EC | ECC | ECN | MBLB | Modul. | PI | RWC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 26-Jährige ruinierte sich finanziell um Instagram-Star zu werden | 0.21 | 0.21 | 0.24 | 0.06 | 0.58 | 0.47 | 0.92 | 0.73 |
| 2 | Antiregierungsprotest diesmal unter neuen Bedingungen | 0.18 | 0.41 | 0.41 | 0.03 | 0.50 | 0.43 | 0.68 | 0.35 |
| 3 | Bud Spencer gestorben | 0.18 | 0.23 | 0.16 | 0.03 | 0.80 | 0.48 | 0.91 | 0.67 |
| 4 | Dönmez mag jetzt Kickl und seine Reden | 0.27 | 0.25 | 0.25 | 0.07 | 0.89 | 0.48 | 0.93 | 0.32 |
| 5 | Das Gymnasium in Zeiten der Bildungspanik | 0.22 | 0.16 | 0.32 | 0.39 | 0.54 | 0.44 | 0.74 | 0.36 |
| 6 | Dicke Luft in München: Merkel und zwei laute Amerikaner | 0.19 | 0.33 | 0.42 | 0.16 | 0.57 | 0.44 | 0.69 | 0.50 |
| 7 | Ein mutmaßlicher Islamist vorläufig festgenommen | 0.12 | 0.18 | 0.18 | 0.30 | 0.31 | 0.35 | 0.43 | -0.01 |
| 8 | Eva Glawischnig: Eine Ex-Grüne findet zur moralischen Flexibilität | 0.21 | 0.23 | 0.30 | 0.07 | 0.64 | 0.48 | 0.88 | 0.53 |
| 9 | Flüchtlinge am Westbahnhof: Alle wollen nur nach Alemania | 0.17 | 0.45 | 0.46 | 0.15 | 0.55 | 0.43 | 0.65 | 0.33 |
| 10 | Grüne prüfen Wahlanfechtung in Niederösterreich | 0.30 | 0.18 | 0.24 | 0.13 | 0.59 | 0.48 | 0.89 | 0.61 |
| 11 | Heiliger Abend sonntags: Die meisten Geschäfte bleiben geschlossen | 0.19 | 0.26 | 0.41 | 0.22 | 0.63 | 0.44 | 0.68 | 0.45 |
| 12 | Kanzler Kern für Ausbildungspflicht bis 25 | 0.17 | 0.45 | 0.31 | 0.05 | 0.56 | 0.46 | 0.82 | 0.68 |
| 13 | Männer und die Abtreibungsfrage | 0.13 | 0.42 | 0.44 | 0.27 | 0.47 | 0.35 | 0.46 | 0.16 |
| 14 | Nach FPÖ-Diffamierung des ORF: Wrabetz will sich wehren | 0.26 | 0.20 | 0.20 | 0.10 | 0.65 | 0.47 | 0.87 | 0.63 |
| 15 | Nach ruhiger Nacht Nickelsdorf erwartet lebhaften Sonntag | 0.16 | 0.48 | 0.37 | 0.14 | 0.46 | 0.39 | 0.45 | 0.27 |
| 16 | Pfeffersprays vom Team Stronach: Aktion ging nach hinten los | 0.27 | 0.22 | 0.02 | 0.13 | 0.40 | 0.44 | 0.81 | -0.34 |
| 17 | Präsidentenwahl: Alle Gemeindeergebnisse auf einen Blick | 0.23 | 0.47 | 0.37 | 0.14 | 0.73 | 0.47 | 0.83 | 0.64 |
| 18 | Schreckliche Musik: US-Nutzer lernen Helene Fischer kennen | 0.24 | 0.19 | 0.28 | 0.29 | 0.67 | 0.47 | 0.85 | 0.49 |
| 19 | Sechs Todesopfer bei Terrorangriff in London | 0.19 | 0.50 | 0.40 | 0.12 | 0.55 | 0.44 | 0.65 | 0.42 |
| 20 | Shitstorm gegen Strache, nachdem er NS-Verbrechen verurteilte | 0.25 | 0.26 | 0.31 | 0.05 | 0.82 | 0.48 | 0.91 | 0.68 |
| 21 | Spannungen zwischen Kroatien und Serbien nach Hitler-Vergleich | 0.06 | 0.08 | 0.03 | 0.29 | 0.22 | 0.22 | 0.27 | -0.26 |
| 22 | Spindelegger wird Direktor von ukrainischer Modernisierungsagentur | 0.30 | 0.12 | 0.06 | 0.03 | 0.64 | 0.46 | 0.82 | 0.13 |
| 23 | Klubchef Schieder: "Keil zwischen Lopatka und die FPÖ getrieben" | 0.20 | 0.21 | 0.14 | 0.05 | 0.68 | 0.48 | 0.92 | 0.60 |
| 24 | Steiermark: Sieben Masern-Fälle seit Jahresbeginn | 0.08 | 0.13 | 0.20 | 0.33 | 0.48 | 0.27 | 0.36 | -0.12 |
| 25 | Von Nordkorea freigelassener US-Student Warmbier gestorben | 0.23 | 0.51 | 0.40 | 0.25 | 0.71 | 0.45 | 0.77 | 0.47 |
| 26 | Vorurteile gegen Arbeitslose: Auf der faulen Haut liegen | 0.20 | 0.47 | 0.32 | 0.07 | 0.60 | 0.44 | 0.72 | 0.23 |
| 27 | Was hinter Kickls Jackenwahl steckt | 0.26 | 0.18 | 0.16 | 0.12 | 0.82 | 0.48 | 0.90 | 0.67 |
| 28 | Weihnachten ruinieren mit nur vier Worten oder weniger | 0.23 | 0.18 | 0.06 | 0.00 | 0.80 | 0.46 | 0.84 | 0.19 |

Table 5.2: Top 5 networks for every measure with respective measure highlighted in orange

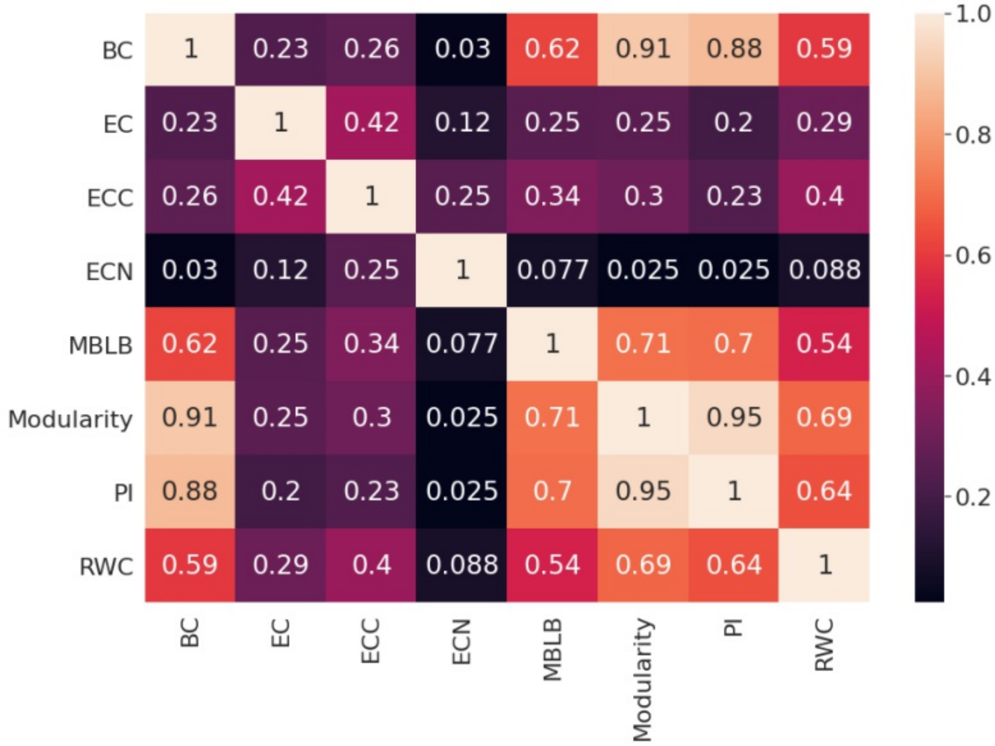



Figure 5.1: Correlation heat-map of 5000 postings networks

Of this list of the top 28 articles, 16 are related to politics and political controversies, four are related to migration and immigration (7, 9, 15, 19) and eight articles are related to various other topics or false-positives. Many articles about terrorism, Islamism or Islamic-related topics are discussed quite intensively, where the discussions often boil down to the two sides of migration opposers and supporters. Articles like article 11, about which shops are open if Christmas is on a Sunday, show how such discussions quickly tend to get to extreme opposites. In this case, discussions quickly reached the very fundamental discussion point of capitalism versus communism basically, where one side fights for shops being open 365 days a year, whereas the opposing users argue against that. Much like described by the theory of *group polarization*, the tendency for a group to make decisions and express opinions that are more extreme than the initial ones of every individual, these discussions converge to opposite extreme opinions [Sun99].

Several false-positives, such as the article about Bud Spencer's death (3) emphasize the need for sanity checks when analyzing polarization. Although the article and the respective discussions are not controversial, except for some discussions about his films' quality, some measurements consider it strongly polarized. In this case, communication structures in the discussions happen to be of a form that mimics cases where polarization is present. However, such false-positive cases can only be detected by human intervention or by additional means, checking if the content is also controversial.

The calculated correlation coefficients based on 5,000 articles are shown in Figure 5.1. Compared to the scores obtained for the datasets from literature (Figure 4.2) it shows an entirely different picture. In contrast to the correlation coefficients obtained previously, *BC* is not correlated with *EC* and its variants anymore, but with all other measurements and vice versa. *ECN* shows the largest difference overall, although strongly correlated with other measurements in the previous setting, it is not correlated to any other measurement anymore.

*BC*, *MBLB*, *PI* and *RWC* are based on the same rationale, that polarization can be measured by analyzing the antagonism inside the network, and that antagonism between two sides is more important than communication structures inside of these communities as would be the case for *EC* and its variants. Therefore it is very plausible that these two groups of measurements produce very different results.

Though the correlation coefficient for *Modularity* is astonishingly high, as it even has the highest correlation coefficients, correlating almost perfectly with *BC* and *PI*. According to the general statistics given by Table 5.3, almost every article would be considered polarized according to *Modularity*, as 75% of all articles have a score higher than 0.37, which would be already considered moderately to highly polarized. Therefore, *Modularity* should only be used as a polarization measurement with great caution.

Overall the vast majority of discussions we analysed are far from being strongly polarized, even for very controversial topics such as politics, veganism or migration. According to the calculated quantiles of the polarization scores (Table 5.3), at least 25% of the articles are not polarized at all or only slightly at most, whereas the next 50% of the articles can also only be considered only slightly polarized. Only a small minority of the articles is moderately to highly polarized, with none of them reaching such high polarization scores as we achieved for the datasets from literature.

| | **BC** | **EC** | **ECC** | **ECN** | **MBLB** | **Modularity** | **PI** | **RWC** |
|---|---|---|---|---|---|---|---|---|
| mean | 0.14 | 0.17 | 0.17 | 0.05 | 0.41 | 0.39 | 0.55 | 0.17 |
| std | 0.04 | 0.11 | 0.09 | 0.07 | 0.11 | 0.04 | 0.12 | 0.13 |
| min | -0.12 | -0.06 | -0.10 | -0.45 | 0.10 | 0.06 | 0.11 | -0.36 |
| 25% | 0.12 | 0.08 | 0.10 | 0.01 | 0.34 | 0.37 | 0.47 | 0.08 |
| 50% | 0.15 | 0.16 | 0.16 | 0.04 | 0.41 | 0.39 | 0.55 | 0.16 |
| 75% | 0.17 | 0.25 | 0.23 | 0.08 | 0.49 | 0.42 | 0.64 | 0.25 |
| max | 0.30 | 0.51 | 0.47 | 0.34 | 0.89 | 0.48 | 0.93 | 0.73 |

Table 5.3: General statistics of 5000 postings networks

In general, the analysis of polarization in complex networks is no straightforward task. It is impossible for humans to create a network in their head, just from analyzing the data such as postings and judge if the resulting network is polarized or not. Therefore we have to rely on polarization measurements such as the ones described in this thesis. In the case of the data present in this case, an online news forum of an Austrian liberal newspaper, the task is even more complicated when compared to experiments from literature relying on re-tweeting networks or hashtags on *Facebook*.

In the case of *Twitter* and other social networks people separate themselves into groups using hashtags, which has been proven to be very efficient and to produce clearly separated communities multiple times. One current example would be the usage of "#JeSuisMila" and "#JeNeSuisPasMila" on *Twitter* after a French teen girl criticised the Islam on Instagram[1]. Afterwards, people clearly express their opinion on *Twitter* by marking their post with either one of these two hashtags. In the case of an online news forum, or when working with user-generated text data, things are much more difficult.

People do not exclusively and clearly express their opinion in such forums. Many comments are noise, containing only short phrases without any context making it hard to detect the opinion of the user, or comments are used to perform votes (using +1 to express agreement for instance) or conduct small surveys. Furthermore, people tend to broaden the scope of a discussion, switching the topic or discuss something completely unrelated. This often leads to not only one large discussion about one single topic being present, but there are multiple small more or less related discussions about a range of sometimes completely different topics. All these points make it much harder to (1) determine the opinion of a user and (2) detect opposing communities to analyze the polarization of a network. This is also the reason we strongly propose hybrid workflows as subject to further research, consisting of Natural Language Processing methods and measurements from social network analysis, to mitigate the vagueness of user-generated content.

Another critical difference between online social networks and other platforms like online news forums is the platform itself. Polarization and its adverse effects are caused by the presence of isolated communities, each with its own opinion without communicating with the other community, leading to these communities and their opinions increasingly drifting apart. On platforms such as *Facebook*, users primarily see content which is previously filtered by recommendation algorithms, creating an endless feed of content which is assumed to interest the user the most. This, in the end, leads to the nowadays famous filter bubbles, where users are imprisoned in their own bubble where they only have access to information adherent to their opinions and believes. This is where polarization is most present and leads to the worst effects.

An online news forum is entirely different. Every user has complete access to all information of the platform, i.e. all articles, topics, postings and comments. Content might be ranked and presented according to general interest and popularity, but mostly regardless of the personal opinions of single users. Furthermore, discussions and communication in general mainly are used as a tool to express disagreement and express one's own opinion, reinforcing more or less constructive discussions. Which is completely different to the behaviour in online social networks, where content is primarily shared with like-minded people and friends, and also almost exclusively consumed only from like-minded ones.

[1] https://www.bbc.com/news/world-europe-51369960

That is the reason many articles and discussions at *derStandard.at* are controverse, dealing with very controverse topics, like religion, politics and sports. But although they are controverse, they are not polarized most of the time, as everyone can communicate with every other one, cluster with different opinions are not isolated, but discuss their point of view quite intensely with others.

## 5.2 Topics

An online news article is only concerned with one very particular event, frequently omitting the bigger picture related to it. A set of related articles, however, may catch this bigger picture, or in our case, the polarization of a whole topic and not one single article. One article alone about migration may not be suitable to capture the overall controversy and polarization around the topic of migration, immigrants and related geopolitical dynamics. All articles about migration or immigrants, published in one year, on the other hand, may be more than enough to grasp opinion dynamics.

As already described above, articles at *derStandard.at* may be grouped using keywords. As almost every article is associated with a set of related keywords, one keyword is associated with all articles concerned with one particular topic. However, there are over 1.5 million classified articles, associated with over 1 million individual keywords, with every article being associated with nine keywords at average - making the task of selecting relevant and suitable topics non-trivial.

To limit the number of potential keywords, we defined the following restrictions. We omit all articles prior 2015 to only focus on current topics, articles with a small number of postings, due to their low importance and only focus on keywords associated with a minimum and a maximum number of articles to filter out too narrow or too broad topics. Although these numbers and thresholds were not determined empirically, they ensure that data unsuitable for analysis is filtered out:

- articles must be published not before 2015-01-01
- articles must have more than 400 postings
- keywords must be associated with at least two articles fulfilling the first two requirements
- keywords must be associated with at most 50 articles fulfilling the first two requirements

Applying these filtering restrictions still yields a list of several thousands of keywords. Therefore, we sorted the list by the average number of postings of the associated articles, having the more controversial and important topics at the top. It must be noted though, that this is not entirely accurate but a very efficient proxy in this case. To further

| | | |
|---|---|---|
| Abtreibungsverbot | Aiman Mazyek | Aistersheim |
| Allahu Akbar | Alternativmedizin | Amaq |
| Ampelpärchen | Amtsgeheimnis | Antifa |
| Antigesichtsverhüllungsgesetz | Arbeitsverweigerung | Asylobergrenze |
| Badeschiff | Beziehungstat | Biken |
| Brice Robin | Burkaverbot | Carlotta Sami |
| Christl Sedlar | Christoph Wiederkehr | Christoph Zielinski |
| Costa Concordia | Deutschförderklasse | Diesel |
| Dijon | Doppelstaatsbürger | Egypt Air |
| Einreisestopp | Evolutionstheorie | Fatma Betül Sayan |
| Fleischproduktion | Flüchtlingskosten | Flüchtlingsstreit |
| G4S | Gerald Fleischmann | Gleichbehandlungsanwaltschaft |
| Glyphosat | Glücksspielkonzern | Halal |
| Handyvideo | Hasskampagne | Herbert Langthaler |
| Herwig Götschober | Idi Amin | Infektionskrankheit |
| Integrationsbericht | Integrationstopf | Integrationsvereinbarung |
| Islamkindergarten | Jude Ben Gurion | Keuchhusten |
| Kindergartenstudie | Kopftuch | Koran |
| Laudamotion | Liederbuch | Lobautunnel |
| Machtmissbrauch | Mario Thaler | Massenquartier |
| Migrationskrise | Notstandsverordnung | Palmers Textil |
| Peter Goldgruber | Philippa Beck | Radverkehr |
| Rainer Wendt | Rauchergesetz | Regierungsklausur |
| Religionsunterricht | Rettungsschiff | Rudolf Taschner |
| Satirezeitschrift | Schmarrn | Schulnote |
| Siemens Healthcare | Silvesterübergriff | Sinkflug |
| Sozialkürzung | Spielfeld | Stefan Steiner |
| Steuergeheimnis | Stundenlohn | Störaktion |
| Synagoge | Tobias Plate | Umgangssprache |
| VW E-Golf | Veganer | Vergewaltigungsversuch |
| Verhetzungsparagraf | Wahldebakel | Wahlkartenprognose |
| Walter Nowotny | Weltuntergang | Wilhelminenspital |
| Wolfgang Albers | Wolfgang Preiszler | Zentralrat der Muslime |

Table 5.4: List of topics

condense this list to 100 relevant keywords, we skimmed through the ordered list and picked ones which probably may be strongly polarized. The final selection as given in Table 5.4 is therefore biased, as it is based on our assumptions.

The selected keywords cover a broad range of different topics. Many keywords are associated with Muslims or radical Islamism, e.g. *Allahu Akbar*, *Islamkinderkarten* (kindergartens operated by then-suspected Islamistic institutions), *Halal* or *Koran*. Migration-related keywords frequently occur too, like *Einreisestopp* (migration stop), *Integrationbericht* (a governmental report about integration-related statistics) or *Migrationskrise* (migration crisis). Another frequently appearing topic are political controversies, with keywords like *Glücksspielkonzern* (gambling company; various controversies around the company Novomatic) or *Schulnote* (school grade; public discussion about whether to grade primary school children). Keywords associated with right-wing or left-wing extremism are also prevalent, like *Antifa*, *Jude Ben Gurion* (controversy about National Socialist songs found with politicians), or *Philippa Beck* (wife of the right-wing politician H.C. Strache). The list is then completed with various other controversies and public discussions such as *Lobautunnel* (a controversial construction project near Vienna) or known-to-be polarizing topics like Veganism (*Veganer*), smoking (*Nichtrauchergesetz*) or tax-laws (*Steuergeheimnis*).

Although the selection of topics may be biased, there is a far more important problem which effectively prohibits the usage of keywords or topics for the analysis of polarization, at least for *derStandard.at*. The postings underneath every article form one unique, isolated network. Before it is now possible to analyse the network structure of a whole topic, consisting of several articles, these isolated networks are joined together forming a new, larger network. The networks are connected through users who have written postings underneath more than one article.

The strength of polarization depends on the network structure. However, when joining different networks together, the links between these networks naturally are much less dense, than in every single network itself. In other words, the resulting network might not be cohesive, but consists of loosely coupled, cohesive sub-networks. This already makes any analysis of this network biased, as this composed network has e.g. a much higher *Modularity* than any of the sub-networks.

Figure 5.2 illustrates this fact. Every network (red, green and blue) has a distinct network structure, which might or might not be polarized. When now creating a new network by connecting them by intersecting posters, the resulting network now consists of several loosely coupled components. Due to this, such a network might be considered polarized, although not one of the networks it is composed of, might be polarized.

Therefore the number of intersecting nodes between two networks, in our case users who have written postings underneath both articles, is critical. The smaller the ratio of intersecting to total nodes, the more biased the analysis becomes. In the worst case, two large networks with thousands of nodes might be connected by only a single edge. This leads to the introduction of one additional selection criteria that the ratio of intersecting nodes between all networks must be high enough to yield a cohesive new network in the end.

After a thorough analysis of the data and networks available, it became apparent that the ratios we achieved for any of the topics we selected, was not nearly high enough to empirically define any threshold satisfying this requirement. We primarily focused on two metrics: (total) intersecting nodes, i.e. the nodes in the intersection of all networks, and the average pairwise intersecting nodes, i.e. the average of calculating the intersecting nodes for every pair of networks.

The distributions of both ratios are shown in Figure 5.3. The ratio of average pairwise intersecting nodes to total nodes, in turn, is also minimal, with most being between 0

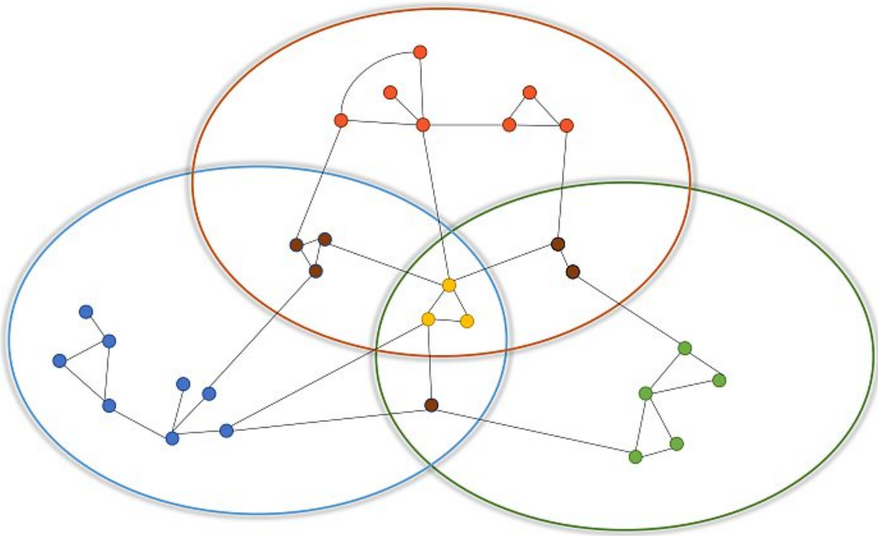


Figure 5.2: Network structure of several smaller connected networks

and 0.10. The ratio of average pairwise intersecting nodes to total nodes in turn is also very small, with mostly being between 0 and 0.10. This means only a small minority of users attends discussions in more than one article of the same topic. Therefore any network constructed for one of the topics selected above would be biased and therefore do not qualify for further analysis.

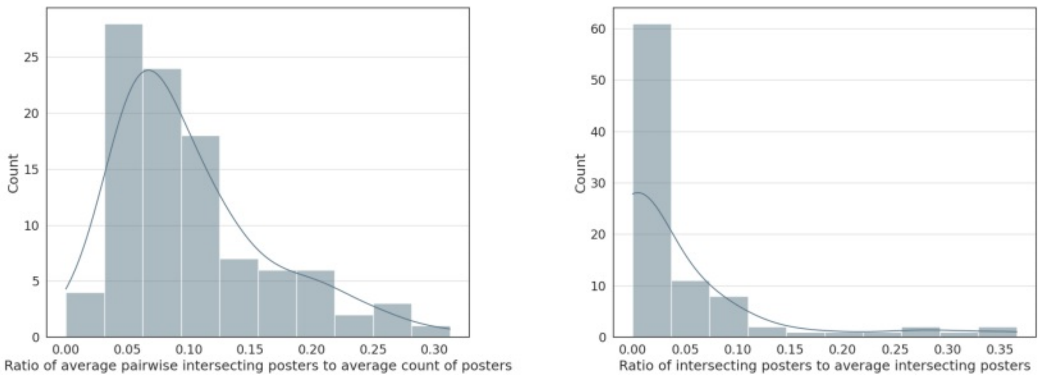



Figure 5.3: Ratios of posters, pairwise average and total intersecting posters

The larger a topic, i.e. the more articles are associated with it, the more important it can be considered. Unfortunately, the more articles a topic is associated with, the smaller the ratio of intersecting nodes gets (Figure 5.4). The highest ratios and number of intersecting nodes were recorded for topics with less than five articles. Therefore the analysis of broader, more general topics is even more difficult than for small, very particular ones.

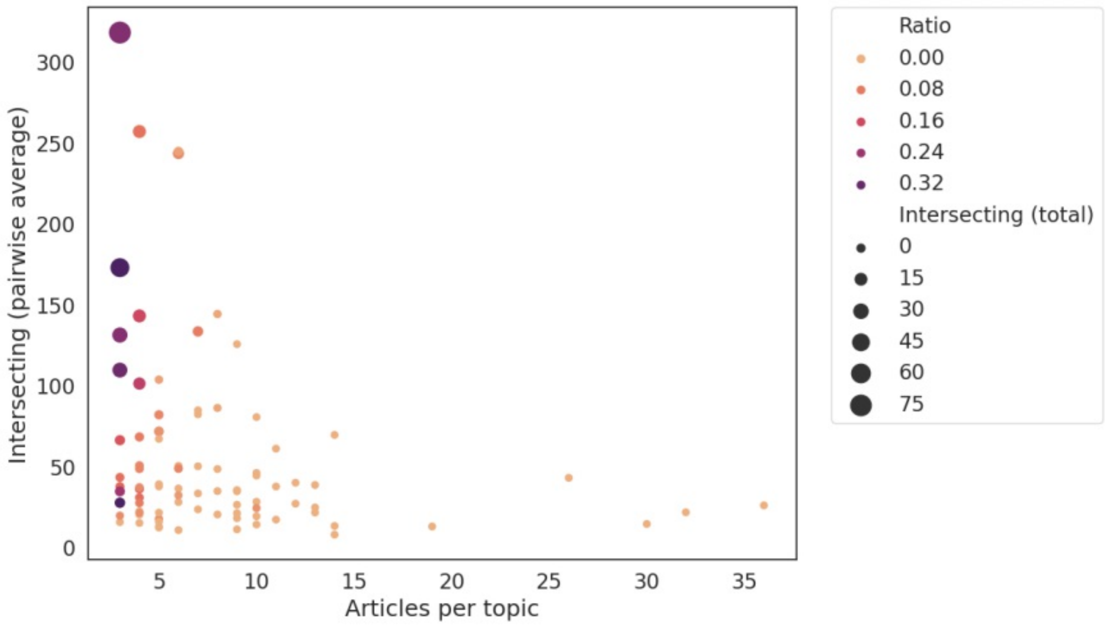



Figure 5.4: Distribution of intersecting nodes

Although the measuring of the polarization on the level of topics with the workflow used in this thesis is not possible for the data present at *derStandard.at*, it neither implies that the analysis might not be possible at all, nor does it imply that it not might be possible for other social networks or contexts. Different methodologies and workflows might be more suitable in such a case. It might also be possible that networks for e.g. different hashtags on *Twitter*, or *Facebook*, are more overlapping, making the analysis of composed networks less difficult. However, it must be noted, that the problem of intersecting nodes and the imposed network modularity on the composed network holds for every setting and therefore must be considered regardless of the concrete context.

## 5.3 Follow and ignore (Community Connection)

Users on *derStandard.at* have the possibility to follow other users by marking them accordingly (called "MitposterInnen")[2,3]. Comments in the comments section written from these followed users are then marked in blue to be found more easily. The mechanism is comparable to friends on *Facebook* or followers on *Twitter*, although the content the user sees is not filtered accordingly. Other platforms use this information to tailor the content a user sees to the preferences of friends or followers. In the case of *derStandard.at*

[2] https://www.derstandard.at/story/1350260404136
[3] https://www.derstandard.at/story/2000061269436

the content of followed users is just coloured in blue, but the overall experience and content stays the same.

If a user violates the forum's guidelines or is generally misbehaving, they can be reported and may be banned according to the violation. Though, if two users just can not stand each other, it is also possible to ignore each other. However, this now alters the content the users sees on the platform, in contrast to following other users. Now every comment of this ignored user is hidden from all comment sections.

Both relationships, following and ignoring, are strong signs of endorsement. Only if a user has a keen interest in the opinion and comments of another user, they will follow each other. Ignoring another user on the other side is a very strong negative sign of endorsement as users ignore each other usually only if they strongly diverge in their opinions – for rude behaviour the respective user would be banned otherwise, so the only disagreement must be the main reason for ignoring other users.

Based on these relationships, we extracted three networks from the database. For this, the *Community_Connection* table was filtered for the respective relationship, extracted and remodelled into an (un-)directed graph:

- Follow - containing all users who follow another user
- Ignore - containing all users who ignore another user
- Follow-Ignore - containing all users who follow or ignore another user

The *Follow-Ignore* network is a signed network with weighted edges. Every follow-relationship between users is indicated by a positive edge weight of +1, whereas every ignore relationship is indicated by a negative edge weight of -1. The general statistics of these networks are given in Table 5.5. As there are much more follow relationships, the Follow network is therefore almost twice the size of the Ignore network. In each of the cases, there exists only one single dominant connected component, with the second-largest connected component being smaller than 0.05% of the complete network.

| Network | Nodes | edges | Largest CC | 2nd Largest CC |
|---|---|---|---|---|
| Follow | 33128 | 126455 | 87.61% | 0.04% |
| Ignore | 17875 | 68598 | 98.35% | 0.02% |
| Follow-Ignore | 38949 | 194710 | 90.91% | 0.03% |

Table 5.5: Community Connection: General statistics

It is essential to mention that these networks contain all users, regardless if they are active, deleted or unregistered users. Furthermore, the networks were cleared from so-called self-loops. A loop, or self-loop, is an edge that connects a vertex to itself. Somehow the Follow network contained a total of 6,577 self-loops, which semantically means these users follow themselves, which makes little sense. However, it is easily possible to the day of writing this thesis, that users can follow themselves on the platform by clicking on “Follow this user”. Ironically a click on “Ignore this user” is ignored by the platform instead of allowing a user to ignore him or herself.

For the visual analysis of the Follow and Ignore networks, we used Gephi and the ForceAtlas2 algorithm to produce a graph layout for each of the networks. The layouts are shown in Figure 5.5. Ironically the visual representation of the Ignore network is shaped like a heart. Besides that, neither the layouts produced for the Follow network on the left nor the Ignore network on the right exhibit any signs of strong polarization. When there is no clear community structure, ForceAtlas2 cuts the graph in two halves by drawing a line through the middle of it, producing visual representations similar to the ones for the community connection networks.

Figure 5.5: Graph layouts of the Follow network on the left and the Ignore network on the right.

In addition to analysing each network in isolation, we also created two layouts for the Follow-Ignore network, shown in Figure 5.6. In the case of the left layout, the edges are coloured according to the colour of the source vertex. In the case of the right, edges are coloured in green if the weight is positive (i.e. if a user follows another one), and red if the weight is negative (i.e. if a user ignores another one). Evidently, neither of the networks shows a distinct community structure and therefore, signs of polarization. The thick lines on the left are a side-effect of Gephi and ForceAtlas2 which produces square shapes for large graphs due to the maximum layout size. Because of this, the majority of nodes are arranged in a square shape with the most nodes being positioned in the corners. This leads to the thick lines connecting the four corners creating the illusion of polarised communities.

The numerical analysis confirms the results of the visual analysis. Table 5.6 depicts

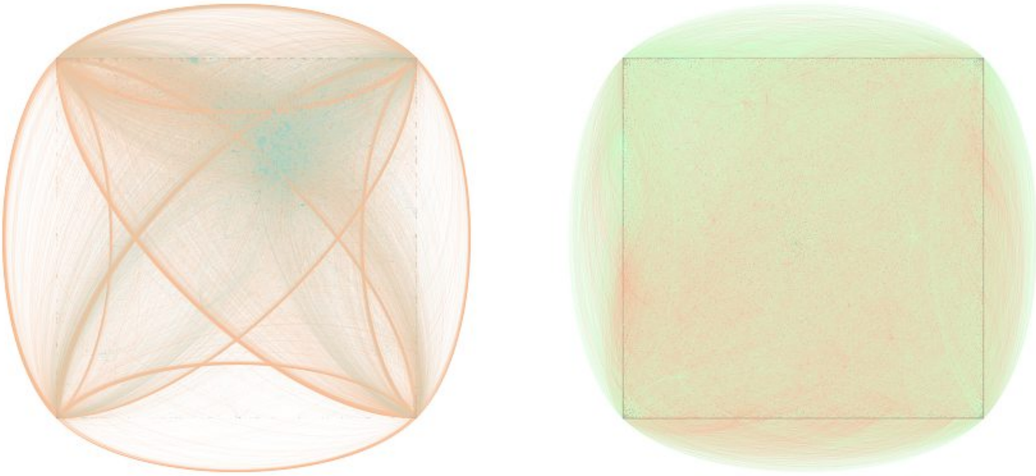


Figure 5.6: Graph layouts of the Follow-Ignore network, with edges coloured according to the colour of the source node on the left, and edges coloured according to their weight on the right.

| Network | BC | EC | ECC | ECN | MBLB | Modul. | PI | RWC |
|---|---|---|---|---|---|---|---|---|
| Follow | 0.02/0 | 0.08/0.05 | 0.01/0 | -0.01/0.01 | 0.27/0 | 0.32/0 | 0.30/0 | 0.25/0 |
| Ignore | -0.01/0 | 0.09/0.04 | 0.04/0.02 | 0.04/0.01 | 0.41/0 | 0.31/0 | 0.24/0 | 0.18/0 |
| Follow-Ignore | -0.04/0 | -0.04/0 | 0.02/0 | 0.03/0.01 | 0.23/0 | 0.31/0 | 0.23/0 | 0.22/0 |

Table 5.6: Results for the community connection networks. Every cell contains the mean value and the standard deviation separated by "/".

the results obtained by calculating every measure-network combination ten times also to calculate the standard deviation. We did not use *BCC* due to the reasons already explained above. All polarization measures show that all three networks are unpolarized, except *Modularity*. This again underlines the importance of using several measures to increase accuracy when measuring the polarization of social networks. Using more than one measure reduces the (rare) risk of outliers.

We suspect the resolution limit of Modularity for being the reason for the unreasonable high polarization scores for all three networks, although all other measures produce very different results. For some network-measurement combinations, the score is even lower than for the synthetic extreme case *Comp. Graph*. Both *BC* and two *EC* variants yield negative values which indicate an even lower polarization than for the extreme case where both achieved a score of 0, and 0.01 respectively. In the case of *BC*, a negative value indicates not only the absence of polarization but that members of one community are even more likely to connect to members of the other community than to members of the own one.

One reason for these networks to not showing any signs of polarization is the scope of analysis being too large. The follow and ignore relationships exist on a global, inter-topic level. However, polarization is analysed based on opinions and topics. An example: Take two users who are on opposing sides of political discussions as the one is more conservative, whereas the other one is social-liberal, the typical left-right political spectrum. Although they would maybe ignore each other due to heated discussions about politics, they might also be fans of the same football club, debating side-by-side against football fans of rival teams. Alternatively, take two users who are against the current measurements taken to prevent the spread of the COVID-19 virus and promote this opinion actively in discussions, but on the other side fight each other on opposing sides in discussions about veganism.

The data and relationships are not able to capture these opinion dynamics. The functionality of following and ignoring other users, similar to *Twitter* and *Facebook*, is therefore not suitable to analyse polarization. The mismatch of the level of data and level of analysis prevents the analysis of polarization in this case. If it would be possible to ignore or follow different people for different topics, this would probably create very robust indicators for polarization.

We therefore also refrained from our initial intuition of creating different networks for different topics. For this, we would have created networks for specific topics containing all users who attended at least one discussion about the topics (e.g. politics or football), and their follow and ignore relationships. The relationship between two users might be influenced by entirely different reasons, outside of the topic to be analysed. Two users ignoring each other, when analysing the polarization of users discussing football, might ignore each other due to their opposing political views. This would make any analysis based on these restricted views very questionable.

CHAPTER 6

# Conclusion

Although the polarization of social networks and society, in general, is still growing in importance, it has only been studied mostly in two settings so far, namely *Twitter* and *Facebook*. Though not in the context of online news media, which are also playing an important role in spreading information and opinions today. To fill this research gap, we identified different ways to measure polarization in social networks.

Through an extensive literature review, we identified the main workflows and methodologies on how to identify and measure polarization in networks. We categorised them into polarization measurements, content-based methods like NLP and NLU and polarization indicators, which are only producing a general impression if a network is polarized or not, without making any statement about the strength of the polarization. As there has not been conducted any comparison of the range of measurements, we conducted an extensive evaluation to answer our second research question (RQ2). For this, we gathered datasets previously used in literature, and created synthetic datasets and evaluated the measurements in terms of run-time performance, quality of separating non-polarized networks from polarized ones and correctness of the score produced.

Although all measurements were positively correlated to each other in terms of the polarization scores, some produced more valuable results than others. *BCC* failed to produce similar results to other measurements due to flaws in its mathematical definition. Besides *BCC*, *RWC* was the one with the lowest average correlation coefficient. Although all measurements might be viable in analysing and quantifying polarization in social networks, *BC* seems to be the common denominator, making it a sound choice for measuring polarization, also due to its advantages in terms of reproducibility and run-time performance. *EC* and its *UMAP*-variants produce scores with the highest standard deviation, making it necessary to compute multiple scores to reduce variability, which makes their run-time performance even worse, making them the slowest measurements overall, followed by *MBLB* and *RWC*.

Based on this comprehensive evaluation, we described and evaluated various ways to extract networks and measure polarization in an online news forum to answer our first research question (RQ1). As there are different kinds of data such as articles, topics, user relationships, postings and votes, we analysed and described each of them thoroughly. We showed that popular and controversial content must not necessarily be polarized. After calculating the polarization scores for 5,000 articles, we also showed that the measurements produce far more diverging results than for the previous evaluation phase. *BC*, *MBLB*, *PI* and *RWC* produce quite similar results, as they all focus on antagonism to detect polarization, whereas *EC* and its variants are unable to correctly measure polarization. On the other hand, *Modularity* produces scores correlated to *BC* and others but qualifies almost every network as strongly polarized, making it not viable for quantifying polarization.

Furthermore, we evaluated the possibility of analysing polarization on the level of whole topics by aggregating similar articles. However, combining different networks into a new one without enough overlapping nodes imposes a modular structure onto the new network, effectively leading to a biased analysis of polarization. We also analysed the relationship between users, as they can ignore or follow each other, but were unable to detect any polarization in these networks.

The workflow presented in this thesis has been proven to be an appropriate and valuable way to analyse and quantify polarization. However, not all data is suitable for these means. The comments underneath one single article led to promising results, whereas user-relationship data was simply too broad as they do not belong to one topic or discussion. Furthermore, topics themselves would be an appropriate way to construct networks to measure polarization, but only if the resulting composed networks consist of enough intersecting nodes.

We could prove that most of the evaluated measurements are appropriate means to measure polarization, namely *BC*, *MBLB*, *PI* and *RWC*. Out of this group, we argue that *BC* would be the best choice, as it offers best run-time performance, zero variance and is able to detect polarization correctly. Although we strongly advise using a combination of several measurements to reduce the rate of outliers and false-positives to improve overall results.

During the analysis of the postings-networks, various points came up, strengthen our assumption that the analysis of polarization in social networks can be significantly improved by combining content-based methods and the methodology from social network analysis. Extracting the opinion out of the usually short and unstructured text of comments and postings is a challenging task, which could greatly benefit from applying content-based methods. Making it a very promising subject for further research. However, in general, we were relieved to conclude that polarization is less common in online news forums than in social networks, where filtering and recommendation algorithms accelerate the adversarial effects of a polarized and segregated society.

CHAPTER 7

# Limitations and future research

This chapter judges the limitations of our work from a technical and methodological point of view. Limitations whose fix would be beyond the scope of this thesis. During the work on this thesis, various ideas for future research arose that would be interesting to pursue, but outside the scope of this thesis. These are summarised in this chapter.

The data our research is based on is limited to *derStandard.at*, the online portal of the largest liberal newspaper in Austria *Der Standard*. Therefore the content on this platform is effectively biased. Not only the content itself is biased towards liberal point of views, but also users and their comments are more liberal than the populations' average. This becomes evident when comparing it to the largest newspaper in Austria, the "Kronen Zeitung" and its online portal "krone.at". The content there, its users and the comments written often express completely opposing and conservative opinions. This effectively limits the analysis of polarization as the content is already focused towards a specific point of view. Classical discussions from literature such as abortion or gun control would not occur in these biased and isolated platforms.

The analysis of polarization for specific topics across different platforms and mediums might be a relevant future research direction. Especially news platforms and social media groups tend to become even more polarized as time goes by, making it more and more challenging to analyse polarization in isolation. As tools and methods of social network analysis alone are not enough to analyse isolated, disconnected communities the combination with content-based methods is necessary to identify major opinion trends and point of views for a specific topic across different platforms by analysing the meaning of user comments.

As our research is based on an online news platform, it is subject to further research if our findings also apply to other platforms and mediums. Communication structures on other platforms might be different, and findings might vary. Most of the relevant research is based on *Twitter*. Therefore our findings must also be validated with data from maybe not only *Twitter* but also *Facebook* and other social networks.

The polarization analysis of articles was only conducted by the authors who read through the comments of an article and judged the strength of polarization. We justify this approach with the rationale that everyone must be able to judge if a discussion is polarized or not. This introduces various biases of which we are aware but believe are not of significant importance to our research and findings. Future empirical experiments with a significant number of people might be used to confirm this assumption and our findings in general.

The concepts used and described in this thesis are limited to methods producing a quantifiable assessment of the polarization in a network. Besides these polarization measurements exist a wide range of methods solely indicating if a discussion is polarized or not, but nothing about the strength of the polarization. These indicators are often content-based, although there are some graph-based ones too.

The possibility to quantify polarization in networks can be used in future research to cross-validate polarization indicators. This would also make it possible to analyse the sensitivity of these indicators, i.e. polarization score at which a network is considered polarized when using the indicator. Furthermore, a combination of measurements and indicators would further enhance the analysis workflow, as indicators in average need far less computation time than measurements and could, therefore, be used as a pre-filtering step in the workflow to quantify polarization only for polarized networks.

The *modified Esteban and Ray measurement (MER)* defined by Sirrianni et al. seems very promising and performed better than *MBLB* in experiments. However, it was originally designed to measure polarization based on similarity scores between users, which were retrieved, in their experiments, from an online argumentation tool. Therefore before analysing this measurement in a setting similar to the one of this thesis, these similarity scores must be calculated somehow.

The range of possibilities and methods able to do this makes this a non-arbitrary task. The similarity between users might be calculated, for example, based on similar topics users commented on in the past or based on the users who liked or disliked the comments written by the same user. This could be further enhanced with external data to generate specific user profiles as it is state of the art in social networks and other platforms, where behavioural and demographic data is used to find clusters of similar users. Every single one of these possibilities encodes different semantics effectively creating a new MER variant which would need to be considered separately, definitely being a subject to potential further research.

The most promising subject for future research is the incorporation of content-based methods into social network analysis based workflows. The number of relevant papers about hybrid approaches, combining content-based methods and social network analysis, is still manageable with our non-exhaustive literature review only yielding one result ([AA17]). We believe the combination of graph-based methods and polarization measurements from the field of SNA and content-based methods could advance the research about the polarization of social networks significantly. Hybrid polarization analysis workflows would make it possible for automated pipelines to understand discussions from a human point of view to quantify polarization then.

Methods such as sentiment analysis, natural language processing (NLP) or natural language understanding (NLU) make it possible to automatically infer the meaning and sentiment of texts and comments from users. This would allow us to unambiguously assign users to specific sides of a discussion by automatically understanding what they are writing about, including pitfalls such as irony or dialects or noise. When solely relying on SNA methods, the text-inherit semantics are entirely ignored, throwing essential and relevant data out of the window. Although, the application of NLP for text preprocessing and filtering would not only be used as a preprocessing step before the network partitioning step. Such a clear assignment to one of the sides of a discussion made through sentiment analysis and NLU would more or less replace partitioning algorithms needed in SNA based workflows altogether. Polarization measurements would then finally be used in such a hybrid workflow to quantify the polarization in these network graphs to get exact, quantifiable results.

# List of Figures

# List of Tables

# Bibliography


[AA17] A. Alamsyah and F. Adityawarman. Hybrid sentiment and network analysis of social opinion polarization. In *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, pages 1–6, May 2017.

[AARJ+18] Mahmoud Al-Ayyoub, Abdullateef Rabab'ah, Yaser Jararweh, Mohammed N. Al-Kabi, and Brij B. Gupta. Studying the controversy in online crowds' interactions. *Applied Soft Computing*, 66:557–563, May 2018.

[AAY+17] M. T. A. Amin, C. Aggarwal, S. Yao, T. Abdelzaher, and L. Kaplan. Unveiling polarization in social networks: A matrix factorization approach. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

[ABMGLB19] Julia Atienza-Barthelemy, Samuel Martin-Gutierrez, Juan C. Losada, and Rosa M. Benito. Relationship between ideology and language in the Catalan independence context. *Scientific Reports*, 9(1):1–13, November 2019.

[AG05] Lada A. Adamic and Natalie Glance. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, pages 36–43, New York, NY, USA, 2005. ACM. event-place: Chicago, Illinois.

[AK08] Michelle Annett and Grzegorz Kondrak. A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs. In Sabine Bergler, editor, *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pages 25–35. Springer Berlin Heidelberg, 2008.

[AKR14] John Adams, Hafiz Khan, and Robert Raeside. *Research Methods for Graduate Business and Social Science Students*. SAGE publications India, January 2014.

[BA99] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999.

[Bas14] Sujit K. Basak. A Comparative Analysis of Zotero and Mendeley Reference Management Software. *International Journal of Psychological and Behavioral Sciences*, 8(11):1969–1972, October 2014.

[BBO90] Elisa Jayne Bienenstock, Phillip Bonacich, and Melvin Oliver. The effect of network density and homogeneity on attitude polarization. *Social Networks*, 12(2):153–172, June 1990.

[BCM$^+$09] A. Bermingham, M. Conway, L. McInerney, N. O'Hare, and A. F. Smeaton. Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation. In *2009 International Conference on Advances in Social Network Analysis and Mining*, pages 231–236, July 2009.

[BF16] Alessandro Bessi and Emilio Ferrara. Social Bots Distort the 2016 US Presidential Election Online Discussion. SSRN Scholarly Paper ID 2982233, Social Science Research Network, Rochester, NY, November 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. arXiv: 0803.0476.

[BHMDW15] Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. Content and Network Dynamics Behind Egyptian Political Polarization on Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 700–711, New York, NY, USA, 2015. ACM. event-place: Vancouver, BC, Canada.

[BKvdV17] Kaspar Beelen, Evangelos Kanoulas, and Bob van de Velde. Detecting Controversies in Online News Media. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 1069–1072, New York, NY, USA, 2017. ACM. event-place: Shinjuku, Tokyo, Japan.

[BNSS17] M. Badami, O. Nasraoui, W. Sun, and P. Shafto. Detecting polarization in ratings: An automated pipeline and a preliminary quantification on several benchmark data sets. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 2682–2690, December 2017.

[CGGL17] Mauro Coletto, Kiran Garimella, Aristides Gionis, and Claudio Lucchese. A Motif-Based Approach for Identifying Controversy. In *Eleventh International AAAI Conference on Web and Social Media*, May 2017.

[CGR+11] M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the Political Alignment of Twitter Users. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 192–199, October 2011.

[CJM10] Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng. Identifying Controversial Issues and Their Sub-topics in News Articles | SpringerLink. In *Proceedings of the Pacific-Asia Workshop on Intelligence and Security Informatics*, pages 140–153. Springer, 2010.

[CLDB18] Xi Chen, Jefrey Lijffijt, and Tijl De Bie. Quantifying and Minimizing Risk of Conflict in Social Networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*, pages 1197–1205, London, United Kingdom, 2018. ACM Press.

[CM19] Uthsav Chitra and Christopher Musco. Understanding Filter Bubbles and Polarization in Social Networks. *arXiv:1906.08772 [physics]*, June 2019. arXiv: 1906.08772.

[CRC08] Patricia Cronin, Frances Ryan, and Michael Coughlan. Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, 17(1):38–43, January 2008.

[CRF+11] Michael D. Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. Political Polarization on Twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*, July 2011.

[Dar19] Kareem Darwish. Quantifying Polarization on Twitter: The Kavanaugh Nomination. In Ingmar Weber, Kareem M. Darwish, Claudia Wagner, Emilio Zagheni, Laura Nelson, Samin Aref, and Fabian Flöck, editors, *Social Informatics*, Lecture Notes in Computer Science, pages 188–201, Cham, 2019. Springer International Publishing.

[Der19] DerStandard. Online-Reichweite: Immer mehr Menschen lesen den STANDARD. https://www.derstandard.at/story/2000109328049/online-reichweite-immer-mehr-menschen-lesen-den-standard, October 2019. Accessed: 2020-01-15.

[DGL13] Pranav Dandekar, Ashish Goel, and David T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, April 2013.

[EK10] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, New York, September 2010.

[ER94] Joan-María Esteban and Debraj Ray. On the Measurement of Polarization. *Econometrica*, 62(4):819–851, 1994.

[Fay96] Usama Fayyad. From Data Mining to Knowledge Discovery in Databases. *AI magazine*, 17(3):18, 1996.

[FB07] Santo Fortunato and Marc Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, January 2007. arXiv: physics/0607100.

[FJ90] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206, January 1990.

[Gar18] Kiran Garimella. *Polarization on Social Media*. PhD thesis, Aalto University, 2018.

[GAS+15] David Garcia, Adiya Abisheva, Simon Schweighofer, Uwe Serdült, and Frank Schweitzer. Ideological and Temporal Components of Network Polarization in Online Political Participatory Media: Ideological and Temporal Components of Network. *Policy & Internet*, 7(1):46–79, March 2015.

[GDFMGM17] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing Controversy by Connecting Opposing Views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 81–90, New York, NY, USA, 2017. ACM. event-place: Cambridge, United Kingdom.

[GJCK13] Pedro Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. A Measure of Polarization on Social Media Networks Based on Community Boundaries. In *Seventh International AAAI Conference on Weblogs and Social Media*, June 2013.

[GMGM15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying Controversy in Social Media. *arXiv:1507.05224 [cs]*, July 2015. arXiv: 1507.05224.

[GMGM17] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. The Ebb and Flow of Controversial Debates on Social Media. In *Eleventh International AAAI Conference on Web and Social Media*, May 2017.

[GMGM18] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying Controversy on Social Media. *Trans. Soc. Comput.*, 1(1):3:1–3:27, January 2018.

[Gra77] Mark S. Granovetter. The Strength of Weak Ties11This paper originated in discussions with Harrison White, to whom I am indebted for many suggestions and ideas. Earlier drafts were read by Ivan Chase, James Davis, William Michelson, Nancy Lee, Peter Rossi, Charles Tilly, and an anonymous referee; their criticisms resulted in significant improvements. In Samuel Leinhardt, editor, *Social Networks*, pages 347–367. Academic Press, January 1977.

[GTT13] Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. Opinion Maximization in Social Networks. *SIAM*, pages 387–395, January 2013.

[HAJR12] Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. Extracting Signed Social Networks from Text. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, TextGraphs-7 '12, pages 6–14, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. event-place: Jeju, Republic of Korea.

[Jan19] Myungha Jang. *Probabilistic models for identifying and explaining controversy*. PhD thesis, University of Massachusetts Amherst, 2019.

[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, October 2014.

[KAHF14] Manfred Klenner, Michael Amsler, Nora Hollenstein, and Gertrud Faas. Verb Polarity Frames: a New Resource and its Application in Target-specific Polarity Classification. *KONVENS*, pages 106–115, 2014.

[KHLJ18] Srijan Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Community Interaction and Conflict on the Web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 933–943, Lyon, France, 2018. ACM Press.

[Kit04] Barbara Kitchenham. Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.*, 33, August 2004.

[KK98] George. Karypis and Vipin. Kumar. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. *SIAM Journal on Scientific Computing*, 20(1):359–392, January 1998.

[KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[Kou18] Ioannis Kouvatis. The role of polarization and diversity in social networks. 2018.

[KPBB+09] Barbara Kitchenham, O. Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1):7–15, January 2009.

[LLM10] Jure Leskovec, Kevin J. Lang, and Michael W. Mahoney. Empirical Comparison of Algorithms for Network Community Detection. *arXiv:1004.3539 [physics]*, April 2010. arXiv: 1004.3539.

[MBLB15] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114, March 2015.

[MHM18] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*, December 2018. arXiv: 1802.03426.

[MMT18] Cameron Musco, Christopher Musco, and Charalampos E. Tsourakakis. Minimizing Polarization and Disagreement in Social Networks. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 369–378, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee. event-place: Lyon, France.

[Mor16] Victor Morel. *Generating co-evolutionary polarized opinion networks*. PhD thesis, Université de Lyon, Lyon, France, 2016.

[MTT17] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, 31(5):1480–1505, September 2017.

[MZDC14] Yelena Mejova, Amy X. Zhang, Nicholas Diakopoulos, and Carlos Castillo. Controversy and Sentiment in Online News. *arXiv:1409.8152 [cs]*, September 2014. arXiv: 1409.8152.

[New06] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, June 2006.

[NG04] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, February 2004.

[NIS19] Sreeja Nair, Adriana Iamnitchi, and John Skvoretz. Promoting social conventions across polarized networks: an empirical study. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 349–352, Vancouver British Columbia Canada, August 2019. ACM.

[Noa09] Andreas Noack. Modularity clustering is force-directed layout. *Physical Review E*, 79(2):026102, February 2009.

[OCLB19] G. Olivares, J. P. Cárdenas, J. C. Losada, and J. Borondo. Opinion Polarization during a Dichotomous Electoral Process, 2019.

[OYD17] Mert Ozer, Mehmet Yigit Yildirim, and Hasan Davulcu. Measuring the Polarization Effects of Bot Accounts in the U.S. Gun Control Debate on Social Media. page 10, 2017.

[Pau13] Marco Pautasso. Ten Simple Rules for Writing a Literature Review. *PLOS Computational Biology*, 9(7):e1003149, July 2013.

[PBI+17] Simonetta Primario, Dario Borrelli, Luca Iandoli, Giuseppe Zollo, and Carlo Lipizzi. Measuring Polarization in Twitter Enabled in Online Political Conversation: The Case of 2016 US Presidential Election. In *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 607–613, August 2017. ISSN: null.

[PP10] Ana-Maria Popescu and Marco Pennacchiotti. Detecting controversial events from twitter. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*, page 1873, Toronto, ON, Canada, 2010. ACM Press.

[RAAJAK16] Abdullateef Rabab'ah, Mahmoud Al-Ayyoub, Yaser Jararweh, and Mohammed N. Al-Kabi. Measuring the controversy level of Arabic trending topics on Twitter. In *2016 7th International Conference on Information and Communication Systems (ICICS)*, pages 121–126, April 2016. ISSN: null.

[RB12] Hoda Sepehri Rad and Denilson Barbosa. Identifying Controversial Articles in Wikipedia: A Comparative Study. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 7:1–7:10, New York, NY, USA, 2012. ACM. event-place: Linz, Austria.

[RGP+17] Anna Rumshisky, Mikhail Gronas, Peter Potash, Mikhail Dubov, Alexey Romanov, Saurabh Kulshreshtha, and Alex Gribov. Combining Network and Language Indicators for Tracking Conflict Intensity. In *International Conference on Social Informatics*, pages 391–404. Springer, September 2017.

[RMM20] Camille Roth, Antoine Mazières, and Telmo Menezes. Tubes & Bubbles – Topological confinement of YouTube recommendations. *arXiv:2001.05324 [nlin]*, January 2020. arXiv: 2001.05324.

[SLA18] Joseph Sirrianni, Xiaoqing Frank" Liu, and Douglas Adams. Quantitative Modeling of Polarization in Online Intelligent Argumentation and Deliberation for Capturing Collective Intelligence. In *2018 IEEE International Conference on Cognitive Computing (ICCC)*, pages 57–64, July 2018. ISSN: null.

[SPP16] Kangwon Seo, Rong Pan, and Aleksey Panasyuk. Detecting Communities by Sentiment Analysis of Controversial Topics. In Kevin S. Xu, David Reitter, Dongwon Lee, and Nathaniel Osgood, editors, *Social, Cultural, and Behavioral Modeling*, volume 9708, pages 206–215. Springer International Publishing, Cham, 2016.

[Sun99] Cass R. Sunstein. The Law of Group Polarization. SSRN Scholarly Paper ID 199668, Social Science Research Network, Rochester, NY, December 1999.

[Sun02] Cass R. Sunstein. The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

[SZLK13] L. M. Smith, L. Zhu, K. Lerman, and Z. Kozareva. The Role of Social Media in the Discussion of Controversial Topics. In *2013 International Conference on Social Computing*, pages 236–243, September 2013.

[TPD11] N. Tsytsarau, T. Palpanas, and K. Denecke. Scalable detection of sentiment-based contradictions. In *DiversiWeb*, volume 1, pages 9–16, 2011.

[WH00] Rüdiger Wirth and Jochen Hipp. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, pages 29–39, 2000.

[WMKHL15] Hywel T. P. Williams, James R. McMurray, Tim Kurz, and F. Hugo Lambert. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32:126–138, May 2015.

[WPF+11] Andrew Scott Waugh, Liuyi Pei, James H. Fowler, Peter J. Mucha, and Mason A. Porter. Party Polarization in Congress: A Network Science Approach. *arXiv:0907.3509 [physics]*, July 2011. arXiv: 0907.3509.

[Zac77] Wayne W. Zachary. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 33(4):452–473, December 1977.

[ZFT+08] Yan Zhang, A. J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler, and Peter J. Mucha. Community structure in Congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712, March 2008.

[ZKWM18] Hengmin Zhu, Yuehan Kong, Jing Wei, and Jing Ma. Effect of users' opinion evolution on information diffusion in online social networks. *Physica A: Statistical Mechanics and its Applications*, 492:2034–2045, February 2018.