Informatics

Measuring Controversy in online discussions

BACHELORARBEIT

zur Erlangung des akademischen Grades

Bachelor of Science

im Rahmen des Studiums

Software & Information Engineering

eingereicht von

Ivan Andreev
Matrikelnummer 12025981

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Dipl.-Ing. Dr.techn. Stefan Woltran
Mitwirkung: Univ.Ass.in Mag.a rer.nat. Dr.in techn. Julia Neidhardt

Wien, 28. Juni 2024

Ivan Andreev	Stefan Woltran

TU WIEN Informatics

Measuring Controversy in online discussions

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Software & Information Engineering

by

Ivan Andreev
Registration Number 12025981

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Stefan Woltran
Assistance: Univ.Ass.in Mag.a rer.nat. Dr.in techn. Julia Neidhardt

Vienna, June 28, 2024

Ivan Andreev	Stefan Woltran

Erklärung zur Verfassung der Arbeit

Ivan Andreev

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, 28. Juni 2024

Ivan Andreev

Danksagung

Ich danke meinen Betreuen Prof. Stefan Woltran und Dr. Julia Neidhard für die Betreuung dieser Arbeit und für ihre Expertise. Prof. Woltran hat während des Seminars „Wissenschaftliches Arbeiten“ mein Interesse für das Gebiet der Künstlichen Intelligenz geweckt. Dr. Neidhard hat mir mit guten Ratschlägen und Ideen sehr bei der Themenwahl geholfen. Ich danke meinen beiden Betreuern herzlich für ihre Freundlichkeit und die investierte Zeit.

Darüber hinaus möchte ich meiner Mutter und meiner Oma für ihre ermutigende Unterstützung während meines Bachelorstudiums danken.

Zum Abschluss möchte ich meiner Freundin Darina von Herzen danken, die mich durch mein gesamtes Studium hindurch unterstützt hat und mir immer wieder Mut gemacht hat, niemals aufzugeben.

Acknowledgements

I am grateful to Prof. Stefan Woltran and Dr. Julia Neidhard for their guidance and expertise in supervising this work. Prof. Woltran ignited my passion for artificial intelligence during the "Scientific Work" seminar. Dr. Neidhard provided invaluable advice and ideas that greatly influenced my choice of topic. I sincerely thank my two supervisors for their kindness and the time they invested.

I also extend my heartfelt thanks to my mother and grandmother for their unwavering support during my bachelor's study.

Lastly, I would like to sincerely thank my girlfriend Darina, who has supported me throughout my entire study and has consistently encouraged me never to give up.

Kurzfassung

Im heutigen digitalen Zeitalter ist die Analyse und Verbesserung des Online-Diskurses von entscheidender Bedeutung. Polarisierung ist in der Tat ein bedeutendes Problem in der heutigen Gesellschaft, das insbesondere in Online-Diskussionen verstärkt wird. Die Fähigkeit, Polarisierung in diesen Kontexten genau zu messen, ist entscheidend, um gesellschaftliche Dynamiken zu verstehen und die damit verbundenen Herausforderungen effektiv anzugehen. In dieser Arbeit schlagen wir eine Methode vor, die sowohl auf der Netzwerkanalyse als auch auf inhaltsbasierten Methoden basiert. Mit unserer Methode lässt sich der Grad der Kontroverse erfolgreich berechnen und auf einer Skala zwischen 0 und 1 bewerten. Der Hauptbeitrag dieser Arbeit ist die Integration zweier primärer Methoden zur Messung der Polarisierung: inhaltsbasierte Methoden und Netzwerkanalyse. Durch die Kombination dieser Ansätze bietet die Studie einen umfassenden Rahmen zur Bewertung der Polarisierung in Online-Diskussionen. Darüber hinaus verbessert die Studie das Verständnis des Online-Diskurses im Kontext von *derStandard.at*, einer österreichischen Online-Zeitung. Durch die Anwendung und Verfeinerung von Messtechniken innerhalb dieser spezifischen Plattform bietet die Arbeit wertvolle Einblicke in die Manifestation und Entwicklung von Polarisierung in digitalen Nachrichtenmedienumgebungen.

Abstract

In today's digital era, analyzing and improving online discourse is crucial. Polarisation is indeed a significant issue in contemporary society, particularly amplified within online discussions. The ability to accurately measure polarisation in these contexts is crucial for understanding societal dynamics and addressing associated challenges effectively. In this thesis, we propose a method based on both Social Network Analysis and content-based methods. Our method succeeds in computing the level of controversy, scoring it between 0 and 1. The main contribution of this thesis is the integration of two primary methods for measuring polarisation: content-based methods and Social Network Analysis. By combining these approaches, the study offers a comprehensive framework for assessing polarisation in online discussions. Furthermore, the study advances the comprehension of online discourse within the context of *derStandard.at*, an influential Austrian online newspaper. By applying and refining measurement techniques within this specific platform, the thesis provides valuable insights into how polarisation manifests and evolves within digital news media environments.

Contents

CHAPTER 1

Introduction

In the era of digital technologies and online communication, social networks have become a very powerful tool for expressing the positions and opinions of individuals. These platforms serve as influential venues for public discussions, where a myriad of views and opinions converge. Ultimately, we would like to understand how users perceive the world through the lens of social media. In our study, however, we concentrate on the still complex task of determining the degree of controversy and polarisation of a given discussion. This analysis allows us to identify topics that spark debate and cause public opinion to split into distinct factions.

Controversy holds significant social and political importance. The presence of controversy is often associated with harassment and hate speech. Strong differences in the opinions of different communities often lead to attacks from one community to another, such as trolling and harassment.While social media disputes filled with hate speech are undoubtedly harmful, the absence of disagreement can also lead to greater polarisation. This is highlighted in various studies such as [MMT18] that point out the dangers of echo chambers, where individuals engage only in discussions that reinforce their existing opinions. Often this creates extreme polarisation and radicalization of individuals, as they have neither opportunity nor willingness to understand the opinions of the other group. Therefore, intensified online discussions are in themselves a positive thing. As shown here [KZN^{+}21], measuring controversy provides the basis for improving consumers "news diet".

1.1 Motivation and Problem Definition

Measuring levels of controversy is a well-explored topic, with ongoing advancements in methods, primarily categorized into graph-based and content-based approaches. While much research focuses on Twitter, few methods directly translate to discussions in online news media. Therefore, our goal is to enhance understanding specifically within online news media discussions, particularly on *derStandard.at*.

Since the master's thesis [Rie21] exclusively evaluates graph-based methods and identifies content analysis as a promising area for future research, we aim to delve into this possibility. Leveraging advanced technologies such as BERT and FastText could significantly enhance our ability to extract and understand the real meaning of user posts.

In this thesis, we successfully integrated both primary methods to achieve tangible results. Our implemented approach effectively captures the levels of polarisation within discussions on *derStandard.at*.

1.2 Methodology

Our method is built upon the framework proposed in [OdZDGFB20]. We have tailored and optimized it to suit the specific domain of *derStandard.at*, ensuring its effectiveness through necessary adjustments and configurations. We utilize a four-stage pipeline in our methodology, which includes graph building, graph partitioning, graph embedding, and computation of the controversy score.

1.2.1 Graph Building

We use a standard approach by creating a graph that represents the relationships between different users. We propose four methods to build this graph. In our experiments, we utilize only one of these methods, as we have determined it to be the most effective in clearly capturing the different communities within the discussion. The methods we offer are:

1. Postings Graph
2. Votes Graph
3. Content-based Graph
4. Hybrid Votes and Postings Graph

1.2.2 Graph Partitioning

To identify different communities within a graph, we employ the Louvain algorithm. This method is widely used for community detection due to its efficiency and ability to uncover hierarchical structures. We discuss the algorithm's parameters in detail to achieve optimal and representative results.

1.2.3 Graph Embedding

In this stage of our pipeline, we embed user posts into vectors for the purpose of computing a controversy score. We employ FastText for this embedding process due to its ability to capture the semantic and syntactic information from text effectively.

1.2.4 Computation of the Controversy Score

In this section, we utilize the outcomes derived from the preceding stages to calculate the level of controversy, ensuring it is represented as a positive number. This process involves several steps, each building on the previously established results to produce a final metric that quantifies the degree of controversy within the community discussions.

CHAPTER 2

Related Work

This chapter aims to showcase and introduce some of the developments and methodologies in the field of Discourse Analysis. In addition, it will explore key terms such as controversy, polarisation, and disagreement, delineating the nuanced differences between these concepts.
The field of discourse analysis has long been devoted to developing methodologies for efficiently analyzing discussions in social media, aiming to achieve maximum accuracy. Controversy in social networks represents a phenomenon with significant social and political ramifications. Numerous studies have concentrated on the examination of controversy and polarisation within social media contexts [GDFMGM18, OdZDGFB20, Rie21].

2.1 Discourse analysis

Discourse analysis refers to the examination of spoken or written texts that encompass more than one sentence, taking into account their social context [cam]. According to [Gil00], discourse analysis encompasses a variety of approaches to studying texts, extending beyond the sentence boundary. Its aim is to reveal socio-psychological characteristics of individuals rather than merely analyzing text structure. Discourse analysis provides a comprehensive approach to understanding the complexities of communication by exploring not only the linguistic features of texts but also their social, cultural, and psychological dimensions. This methodology can be employed to assess the quality of discussions, identify different groups of people and opinions, analyze the emotions of participants, and determine the controversial nature of a given topic.

2.2 Controversy

The Cambridge Dictionary [cam] defines "controversy" as "a lot of disagreement or argument about something, usually because it affects or is important to many people."

In [MZDC14] they analyze how controversial a given news article is by analyzing the text of the article itself, rather than a discussion of it. Controversy is defined as something that causes widely differing, opposing opinions. They analyze the article as plain text, looking for certain words that speak to the level of controversy. So they identify strong correlation between controversial issues and the use of negative affect and biased language. One of the controversy detection methods namely the content-based method can be noticed here.

Several articles focus on case studies centered around enduring major events. However, in [GDFMGM18], the authors endeavor to identify and quantify controversy surrounding any topic discussed in social media, including those that are short-lived. Their methodology involves a graph-based three-stage pipeline for quantifying controversy. The paper suggests different ways of creating the graph, as well as different methods of measuring the level of controversy. They consider the content-based method to be insufficiently reliable and for this reason mainly use the second main method to measure the level of controversy, namely the Social-network analysis.

A very interesting combination between the two main methods of measuring the level of controversy is proposed by [OdZDGFB20]. In addition to creating a graph based on various connections between users, they also use the capabilities of Artificial Intelligence and Machine Learning to perform content analysis, thus combining the two approaches. They designed a NLP-based pipeline to measure controversy. In general their approach is not domain-, language-, geography- or size-dependent.

In general, controversy is something that relates to a particular topic or article. Therefore, we can try to measure the likelihood that an article or just a text in the online space will cause discussion and different opinions. It is not entirely correct to define a discussion as controversial. The discussion may contain polarisation or simply multiple differing opinions. The similarities and differences between these terms are discussed later in the paper.

2.3 Polarisation

Polarisation, as defined in the Cambridge Dictionary [cam], refers to "the act of dividing something, especially something that contains different people or opinions, into two completely opposing groups."

Polarisation can be seen as a consequence of a controversial topic. In fact, all of the articles mentioned above examine the degree of separation between two opposing groups of people and opinions, that is, polarisation. But the polarisation itself is not something bound to social media. Polarisation can be seen in public discourse, media representation, voting patterns, and social interactions, among other arenas.

In public discourse, polarisation manifests through the stark contrasts in perspectives and ideologies presented by different individuals or groups. These divergent viewpoints often lead to heated debates and entrenched positions, further exacerbating societal divisions.

The polarisation index, which, given a network and the opinions of the individuals in the network, quantifies the polarisation observed in the network, was defined in [MTT17]. They also consider the problem of reducing polarisation in the network by convincing individuals.

[DGL12] discusses in detail how polarisation is formed and suggests another way to measure it. Furthermore, the relationship between community polarisation and homophily is meticulously analyzed. Homophily, the tendency of individuals to associate and bond with similar others, plays a significant role in the dynamics of polarisation.

2.4 Group polarisation

In social psychology, group polarisation refers to the tendency for a group to make decisions that are more extreme than the initial inclination of its members. This phenomenon occurs because group discussions often amplify the prevailing attitudes of the members, leading to a shift towards more extreme positions. Consequently, individuals within the group reinforce each other's viewpoints, further solidifying and intensifying the group's overall stance.

2.5 Polarising or controversial

Something is controversial if it causes some sort of discussion or disagreement, while polarisation goes beyond that. To talk about polarisation we need strong division of the members in two or more contrasting groups. As we noted above, polarisation can be seen as a consequence of a controversial topic.It is important to mention that this is not necessarily the case, that is, a controversial topic may or may not lead to polarisation. It is also not completely excluded that a topic that is not controversial in itself can lead to polarisation.

2.6 Disagreement

The Cambridge Dictionary [cam] defines "disagreement" as "an argument or a situation in which people do not have the same opinion". Disagreement is a concept more closely tied to specific discussions rather than to general topics. According to [MMT18], minimizing disagreement can actually lead to greater polarisation. This phenomenon occurs because as users connect with others who share similar mindsets, they form two distinct clusters with strong and extreme opinions. Consequently, these groups become more polarised, and the level of disagreement within each group decreases. Thus, there is a trade-off between disagreement and polarisation: reducing disagreement tends to foster greater polarisation between the two groups, resulting in more extreme and less diverse viewpoints.

2.7 Echo Chambers

Echo chambers are environments where individuals encounter only opinions and beliefs similar to their own, without having to consider alternative perspectives. Echo chambers can create misinformation and to distort a person's perspective. Because of this a person may have difficulties considering opposing viewpoints. This insulation can foster the reinforcement of existing beliefs, exacerbating polarisation and hindering critical thinking. This phenomenon can contribute to the spread of misinformation and the distortion of reality, as individuals are less likely to encounter diverse sources of information or engage critically with conflicting viewpoints.The term 'echo chambers' is frequently mentioned in the literature [OdZDGFB20, GDFMGM18, MMT18].

2.8 Methodologies for quantifying controversy

Measuring the levels of controversy and polarisation in a social network is a complex task, so there are various methods by which it can be done. One of them is based on the connections between users, likes, dislikes, replies, follows, etc. The other is based on the content of the text or comments, trying to understand the sentiment of the user and the meaning of the written text.

2.8.1 Social Network Analysis

Social Network Analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. Quantifying the level of polarisation involves three main steps: constructing the network (graph), assigning sides, and performing quantification. To quantify the level of controversy, emphasis is placed solely on the structural aspects of the graph, devoid of any analysis pertaining to the textual content or its semantic significance. This approach underscores the importance of examining the network's configuration and relational patterns in understanding the dynamics of controversy within social contexts. [GDFMGM18] suggests different ways to create the graph, as well as several different metrics to measure the degree of controversy in a given discussion.

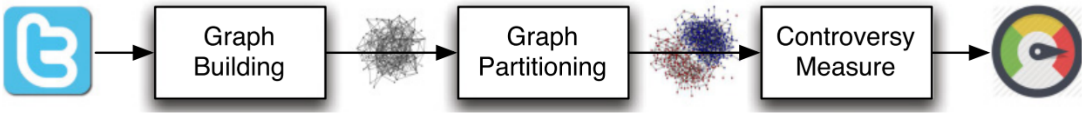


Figure 2.1: Block diagram of the pipeline for computing controversy scores.[GDFMGM18]

2.8.2 Content-based methods

The main methods of analyzing and measuring the level of polarisation and the degree of controversy of a given article or discussion are based on analyzing graphs. This to some extent distances the analysis from the real meaning invested in the various discussions. Another approach to the problem is content analysis. According to [Rie21], using content-based methods can lead to a better partitioning of the graph, as well as bring measurements closer to how people perceive polarisation and controversy.

Content-based methods for quantifying controversy prioritize the analysis of textual content and its semantic meaning rather than solely focusing on the structural aspects of the graph. These approaches delve into the substance of the information exchanged within the network, assessing the level of disagreement or contention based on the language used, the topics discussed, and the sentiment expressed.

2.8.3 Combination of Social Network Analysis and Content-based Methods

Combining social network analysis with content-based methods offers a promising approach to calculating the level of controversy. This approach captures both the structural features of the network and the meaning of the text in users' posts.

In [OdZDGFB20], the authors demonstrate an effective approach for combining the two methods. They offer a 4-phase pipeline that includes both graph building and analyzing the content using LLMs. They mainly concentrate on twitter. Their method can be used for different languages. Their approach not only proves that it can capture the controversy, but also improves the accuracy and speed of previous methods.

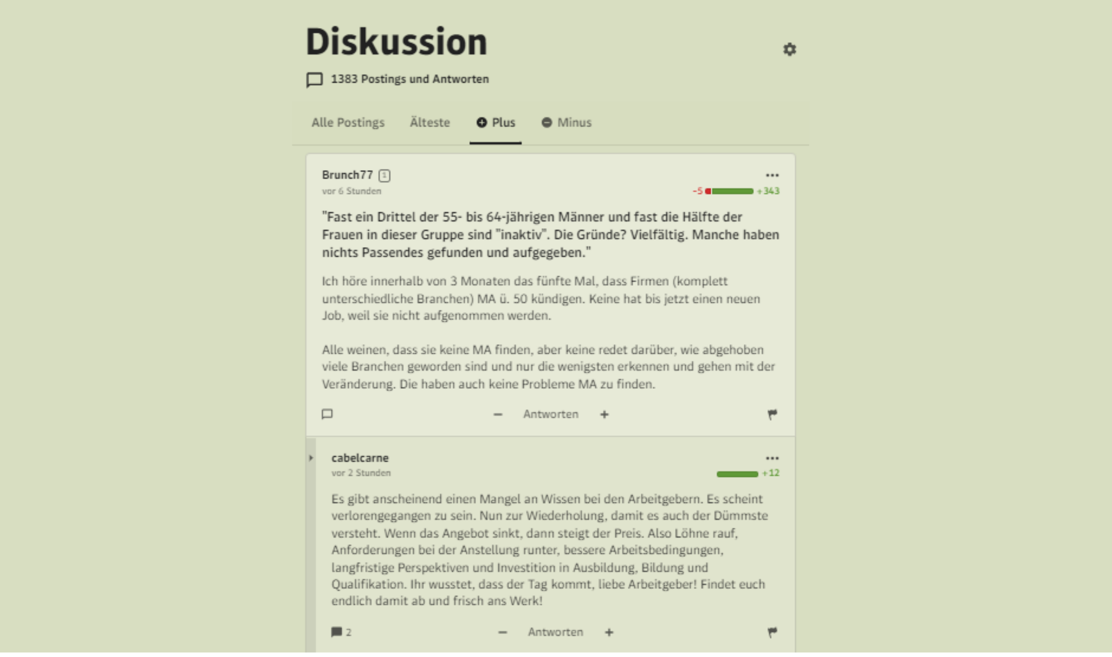
Integrating these two approaches captures different aspects of a given discussion, thereby enhancing the accuracy of the results.

CHAPTER 3

Methodology

3.1 Data Selection

The selection of data to be included in the analysis and measurements is key to the outcome.



Brunch77
vor 6 Stunden

"Fast ein Drittel der 55- bis 64-jährigen Männer und fast die Hälfte der Frauen in dieser Gruppe sind "inaktiv". Die Gründe? Vielfältig. Manche haben nichts Passendes gefunden und aufgegeben."

Ich höre innerhalb von 3 Monaten das fünfte Mal, dass Firmen (komplett unterschiedliche Branchen) MA ü. 50 kündigen. Keine hat bis jetzt einen neuen Job, weil sie nicht aufgenommen werden.

Alle weinen, dass sie keine MA finden, aber keine redet darüber, wie abgehoben viele Branchen geworden sind und nur die wenigsten erkennen und gehen mit der Veränderung. Die haben auch keine Probleme MA zu finden.

cabelcarne
vor 2 Stunden

Es gibt anscheinend einen Mangel an Wissen bei den Arbeitgebern. Es scheint verlorengegangen zu sein. Nun zur Wiederholung, damit es auch der Dümmste versteht. Wenn das Angebot sinkt, dann steigt der Preis. Also Löhne rauf, Anforderungen bei der Anstellung runter, bessere Arbeitsbedingungen, langfristige Perspektiven und Investition in Ausbildung, Bildung und Qualifikation. Ihr wusstet, dass der Tag kommt, liebe Arbeitgeber! Findet euch endlich damit ab und frisch ans Werk!

Figure 3.1: Discussion under an article on *derStandard.at*

The data we use is provided by the Austrian newspaper Standard. It is one of the most popular online media in Austria. As in other online news platforms, *derStandard.at* has various publications that are sorted by topic. Users can write comments (posts) under each news item. Of interest to our research is that users have the ability to reply to other users' comments, allowing discussions to be created. Another interesting option that can be useful for measuring the level of controversy is voting on other users' posts. Votes can be positive or negative. Understandably, the positive vote is clearly recognized as approval, while the negative vote is recognized as negation. A clearer idea about the platform structure and user capabilities can be gained from Figure 3.1

Each article is associated with various keywords that can be utilized to form topics encompassing multiple articles. This approach facilitates the analysis of a larger volume of information, offering both advantages and certain limitations. The positive thing is that this way we can look at the topics above, getting a lot more information and activity from the users. On the other hand, the accuracy and reliability of the results is lost. Because the fact that certain posts have common keywords does not guarantee that the topics discussed in them and in the discussions related to them are similar.

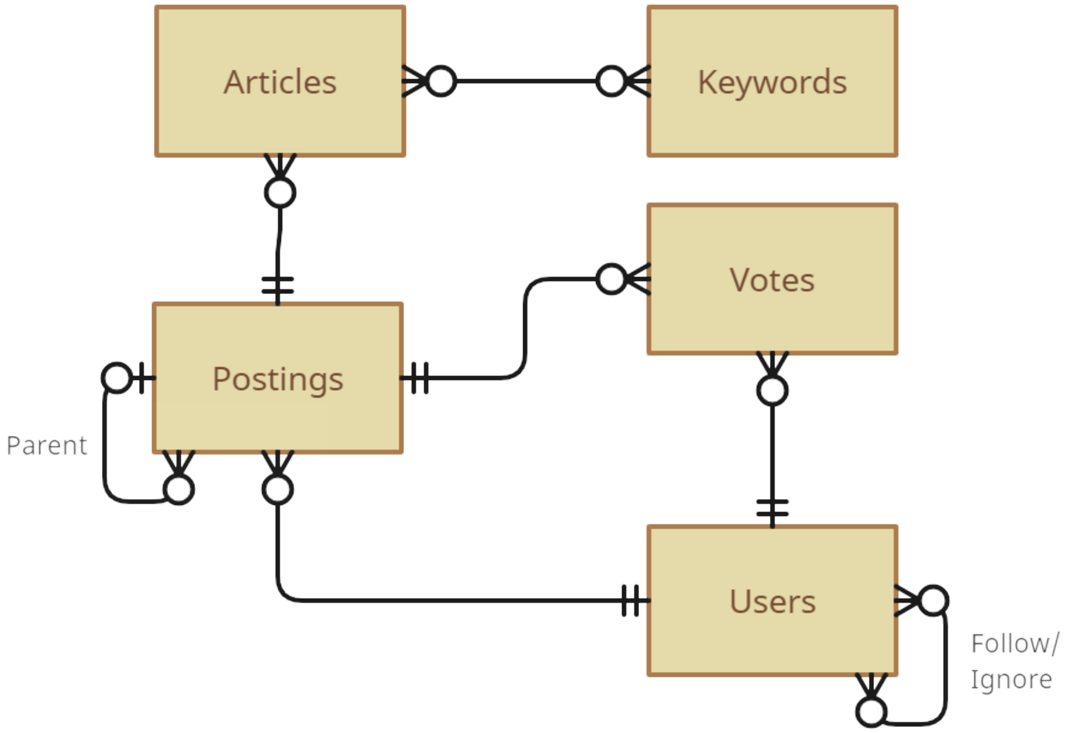


Figure 3.2: Entity-Relationship (ER) Diagram representing a subset of the *derStandard.at* database.

Perhaps one of the most important data in our research is the posts. Commenting on a news article shows engagement. Controversy cannot be inferred simply from the presence of comments. However, if we analyzed the sentiment of the comments, that is, using content-based methods, it would probably yield results. Just such a combination between Social Network Analysis and content based methods has been proposed as a very promising future work by [Rie21].

Analyzing user votes and integrating them with posts is a viable and frequently discussed option. While user votes may not represent the same level of engagement as posts, they provide a clearer and more definitive measure of user sentiment, given their binary nature—being either positive or negative. Furthermore, the volume of votes significantly exceeds the number of posts, by a factor of approximately four, as evidenced by the data presented in Table 3.1. By leveraging both types of data, we can achieve a more comprehensive analysis of the level of controversy and engagement within the discussions. This dual approach can help mitigate the limitations of each data type, enhancing the overall robustness of our findings.

Statistic	**Value**
Classified Articles	1,968,318
Users	858,926
ActiveUsers	450,895
TotalPostings	119,884,043
Mean Postings Per Article	104
Votes	484,347,541
Posts/Votes	0.247516572
Votes/Posts	4.040133523

Table 3.1: General Statistics 2021

The analysis reveals that a substantial segment of users demonstrates limited activity levels within the system. However, it is pertinent to note that these activity patterns do not impact our research objectives. Consequently, while variations in user activity exist, they do not substantially alter the course of our research inquiry. Therefore We are incorporating data from all users, regardless of their activity levels, into our analysis.

3.2 Topic Definition

An article focuses primarily on a specific event, honestly not providing insight into the overall big picture. A chance to capture the big picture and specifically the level of polarisation in an entire topic is to analyze groups of articles on the same topic. It is entirely possible that just one article on Covid-19 vaccines for example and the discussion below it is not a representative sample and cannot capture the real degree of polarisation.

As can be seen in the ER diagram in Figure 3.2 in *derStandard.at* each article is associated with keywords, precisely using these keywords would allow us to create groups of related articles and view them as a whole. A detailed analysis of this topic definition possibility can be read in the master's thesis [Rie21].

The idea is to create a graph based on the connections between users. Using more than one article to create such a graph understandably leads to the creation of many unrelated smaller graphs. An option to connect these graphs into one is to use users who participate in more than one discussion. The use of this connection method is highly dependent on the number of users who have participated in more than one discussion. When there are not enough of them, separate communities are obtained in the graph, which is a sign of controversy, although the topic may not be controversial at all. In [Rie21], the authors conclude that in this way, a sufficient number of connecting users cannot be achieved to guarantee the legitimacy of the graph.

For this reason, as well as the fact that there are more than 15,000 articles with more than 1,000 posts, we define a topic as a single article. In this way, we eliminate the risk of having different communities, without this implying polarisation.

3.3 Graph Building

Having considered the types of data that would be relevant to our research, the next step is to construct a network. The way to construct the graph is of great importance for the quality and reliability of the information that can be extracted from it, as well as for the level of controversy, which is calculated based on the graph.

In most works on the topic (quantifying controversy online), the users are represented as the nodes [GDFMGM18, OdZDGFB20]. A node is created for each user present in the data selected for calculation. Edges represent a different kind of connection, communication or relationship between users.

In the literature, investigations and measurements of social networks often focus on Twitter. For example, researchers [OdZDGFB20] frequently use the retweet graph to create connections between users. This graph visually represents the sharing of posts from other users, operating on the assumption that retweeting signifies approval. In other words, the retweet graph creates connections between users who endorse or agree with each other's opinions. Therefore, when modeling social interactions in such a graph, it is preferable to represent an edge as an indication of endorsement or agreement. In this manner, groups of users who not only interact with each other but also share similar opinions can be clustered together, as their mutual agreement indicates a common perspective. This approach enables the identification of communities within the network where users are interconnected through shared endorsements, reflecting their aligned viewpoints and fostering a sense of collective agreement.

Such a method of creating the graph always results in a partially connected graph. For this reason, we consider only the largest connected component. This does not affect the results, as in all the graphs we constructed, the largest connected component accounts for at least 90% of the entire graph.

3.3.1 Unweighted Postings Graph

The posts that users have the opportunity to write under the various articles are suitable for analysis, as they carry a lot of meaning and may show different structures of agreement or disagreement.Particularly noteworthy is the capability for discussions to evolve through the initiation of responses by users to one another's comments, thereby providing a fertile ground for detailed analysis.

In this graph, each vertex represents an individual user involved in the discussion below the article. Every user who leaves a comment is assigned to a unique vertex. The existence of an edge between two vertices v_1 and v_2 indicates a single interaction between the corresponding users. A single interaction between two users, represented by an edge, can be defined as a reply from user u_1 to user u_2, or vice versa. If there are multiple interactions between two users, multiple edges will be present between their respective vertices.

By identifying the different communities in the discussion, we would notice different groups communicating with each other. But, as specified above, when the edge represents only communication between users, it cannot be unequivocally judged whether they agree with each other and simply reaffirm their opinion, forming so-called echo chambers. Or they completely disagree and argue, presenting opposing arguments.

Henceforth, it can be deduced that employing an undirected and unweighted graph, predicated solely on iterative interactions among users, may not offer a robust or accurate foundation for quantifying the degree of controversy within a discourse.

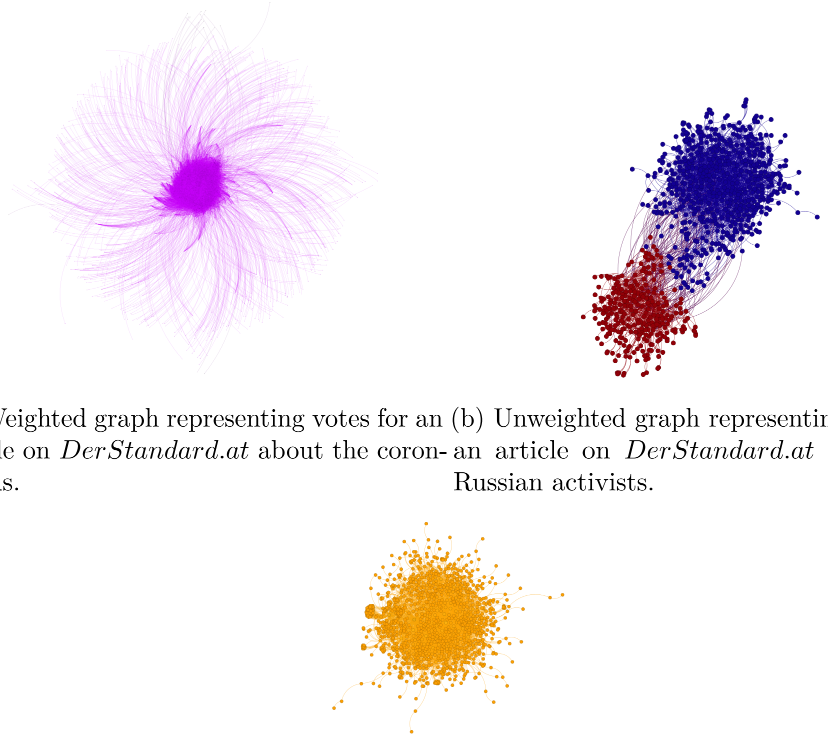
3.3.2 Votes Graph

Creating a graph based on user votes for each post is a promising approach, given the definite positive or negative connotations associated with votes. As demonstrated in the provided statistics, votes on "der Standard" are approximately four times more numerous than comments. Consequently, articles that generate substantial user engagement can accumulate an exceptionally high number of votes on the posts beneath them.

However, despite the clear-cut nature of votes, their capacity to represent genuine commitment or complete agreement/disagreement with a post is uncertain. The simplicity of casting a vote with a single click diminishes its value as an indicator of real engagement with the topic, in contrast to the effort and thought involved in writing a comment.

Each user who left a vote or whose post was voted for is represented as a node. As a first attempt at creating a graph based on votes, we tried to represent every positive and negative vote associated with a specific post. This was implemented as a weighted graph where edges were assigned a weight of 1 for positive votes and 0 for negative votes. However, this approach proved inadequate for producing meaningful graphs, as it contravenes the principle that each edge should signify a form of endorsement or interaction. Unusable graphs are obtained in which the positive votes are clustered in one place and the negative ones circle around them. An example of such a graph is depicted in Figure 3.3a.

Because placing a single vote does not demonstrate sufficient engagement to be considered significant, an edge is not created for each individual vote. Creating an edge for each



(a) Weighted graph representing votes for an article on *DerStandard.at* about the coronavirus.

(b) Unweighted graph representing votes for an article on *DerStandard.at* about pro-Russian activists.

(c) Unweighted graph representing votes for an article on *DerStandard.at* about favorite movies

Figure 3.3

vote, where one user votes on another's post, results in graphs that are excessively large and lack representativeness. That's why we decided to create a graph from only positive votes. By creating an edge between two users n1 and n2 only if n1's votes to n2's posts are only positive or vice versa. This shows strong agreement and lack of disagreement, which helps identify different communities. An example of such a graph can be seen in Figure 3.3b. It represents the votes of an article that we would rather qualify as controversial. Two distinct communities can be clearly discerned.

Figure 3.3c illustrates the application of the graph building method to an article that does not raise any controversy, as it is about sharing favorite movies in one sentence. Certainly, the lack of controversy and polarisation is evident purely visually in the graph, giving us confidence that this method of graph construction succeeds in capturing diverse communities and is accurate enough to be used for computation.

It is important to note that the graph is colored only based on the output of the Force Atlas 2 algorithm and not on other clustering algorithms.

3.3.3 Content-based Graph

The thesis [Rie21] presents the use of content-based methods as a very promising area for future research. For this reason, we tried to create a graph based on the content of the posts. The idea is to use sentiment analysis to classify posts as positive, neutral or negative and create a graph based on that.

To facilitate sentiment analysis in our study, we utilized the German Sentiment Classification model [GSBB20] based on Google's BERT architecture. This model, trained on 1.834 million German-language samples from diverse domains such as Twitter, Facebook, and various reviews, enables us to classify texts into positive, neutral, or negative sentiments with high accuracy. By incorporating this model, we aim to leverage its robust sentiment classification capabilities to analyze the sentiment of the posts, thereby enabling the creation of a sentiment-based graph for our research.

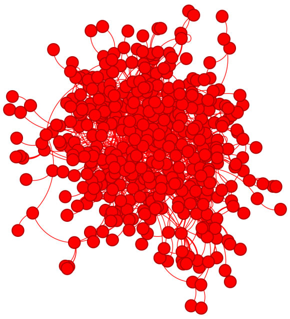


Figure 3.4: Content-based graph representing positive postings an article on *derStandard.at* about pro-Russian activists.

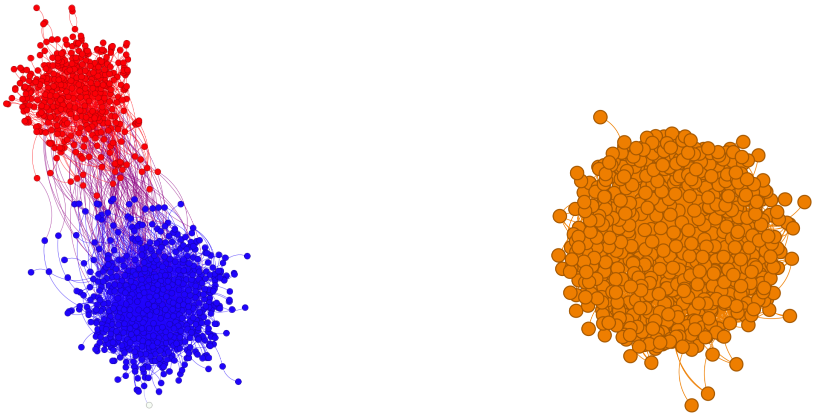
The approach is similar to that presented in Section 3.3.1. That is, for each user involved in the discussion, we create a node. An edge between n_1 and n_2 is raised if n_1 responded to n_2's post (or vice versa) and this response was classified as positive or neutral by sentiment analysis. We include both positive and neutral texts because the model requires a text to explicitly and strongly convey positivity and agreement for it to be classified as positive. Consequently, our graph construction is somewhat based on the absence of negation. The main difference with the approach in Section 3.3.1 is that we respect the idea that an edge between two users must represent some form of approval.

In [GDFMGM18], the authors attempt to construct a graph based on content analysis. However, both in their work and ours, this approach does not align well with our intuitive understanding of the various topics. This discrepancy is evident in Figure 3.4, where a graph is generated from a highly controversial topic. While the Votes graph in Figure 3.3b clearly demonstrates the presence of polarisation, the content-based graph fails to accurately reflect the polarized nature evident in many societies.

3.3.4 Hybrid votes and postings graph

A graph created only from votes is left for future research in the master's thesis [MTT17]. The reason they don't analyze vote graphs is because the people who leave votes aren't engaged enough in the discussion. That is, the vote by itself does not express enough commitment, therefore approval or disapproval. This is a thesis with which we agree, users who only vote and do not leave a post or reply to a post do not provide enough legitimacy.

That's why we offer a hybrid graph approach. By combining two methodologies, we aim to achieve greater legitimacy in our results. The two approaches we combine are the votes graph and the postings graph.



(a) Hybrid graph for an article on *derStandard.at* about pro-Russian activists.

(b) Hybrid graph for an article on *DerStandard.at* about favorite movies.

Figure 3.5

In our model, a node represents a user. An edge between two nodes n_1 and n_2 exists if n_1 has voted for n_2's post(s), all votes from n_1 to n_2 are positive, and n_1 has posted a comment under the given article. Since the graph is undirected, the conditions apply vice versa: n_2 can have voted for n_1's post(s), all votes from n_2 to n_1 are positive, and n_2 has posted a comment under the given article.

$$\text{Edge}(n_1, n_2) \iff \begin{cases} n_1 \text{ has voted for } n_2\text{'s post(s)} \\ \text{All votes from } n_1 \text{ to } n_2 \text{ are positive} \\ n_1 \text{ has posted a comment under the given article} \end{cases}$$

By incorporating the criterion that a user must have written a post under the article, we exclude users who are deemed to be insufficiently engaged. The rationale behind this is that once an individual has contributed a post related to the specific topic, their votes can be regarded as sufficiently representative and legitimate.

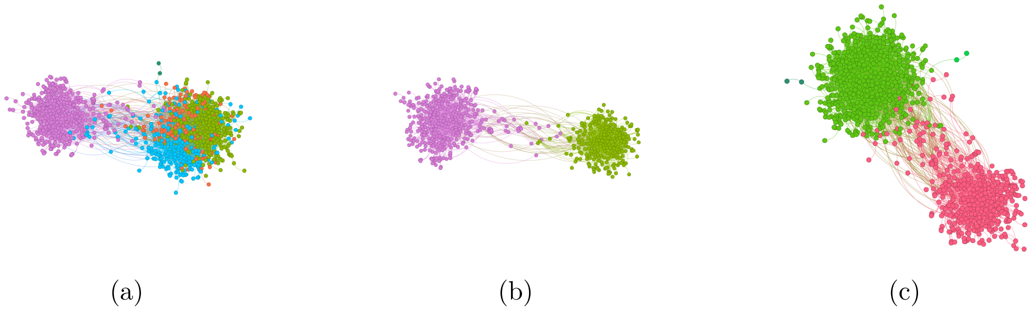
The graphs in Figure 3.5 are visualized in the same manner as those in Figure 3.3, utilizing the ForceAtlas2 algorithm. While there is no significant visual difference between the hybrid and ordinary votes graphs, it is important to note that the hybrid graphs exclusively include users who have written a post under the article. This inclusion criterion makes the hybrid graphs more reliable.

By using this hybrid method, we can better identify influential users and key opinion leaders within the community. Users who both post and receive positive votes are likely to have a significant impact on the discussion, and their interactions can reveal important patterns of influence and agreement. This enriched analysis can inform strategies for fostering engagement and managing community dynamics more effectively.

3.4 Graph Partitioning

To identify distinct communities within a graph, employing an algorithm is imperative. Various algorithms, such as METIS and Louvain, offer effective solutions. In this context, the choice between them often depends on factors like graph size, computational resources, and specific requirements.

For instance, the [GDFMGM18] utilizes METIS, while [OdZDGFB20] opts for Louvain. We opt for Louvain due to its widespread popularity and efficacy in graph clustering tasks. Louvain stands out as a grid-based algorithm renowned for its ability to handle large graphs seamlessly, without compromising on memory usage or computational speed.



(a) (b) (c)

Figure 3.6: Community structures with different resolutions: (a) 6 communities with resolution 1.00, (b) 2 biggest communities with resolution 1.00, (c) 2 communities with resolution 1.50.

In essence, the Louvain algorithm offers a parameter known as resolution, allowing for control over community size and consequently, the number of communities identified. By default, this parameter is set to one. Interestingly, despite this default setting, when employing visualization tools like Force Atlas 2, the Louvain algorithm often detects more than two communities.

However, adjusting the resolution slightly, such as to 1.5, tends to yield two communities. Yet, this approach poses risks as it may bias the algorithm towards finding exactly two communities, potentially compromising the legitimacy of the outcome.

The best option is to use Louvain with a default resolution of 1 and then take the two largest communities, because removing the smaller communities does not affect the final results. This method ensures that the analysis remains robust and unbiased while maintaining the integrity of the detected community structures. By focusing on the two largest communities, we can effectively simplify the graph without losing significant information, thereby facilitating a clearer and more manageable analysis.

3.5 Embedding Phase

As we have mentioned several times so far, the main goal of this work is to find a way to combine the methods related to analyzing the graph (Social Network Analysis) and the methods based on the textual content of the users' posts (Content-based methods). This combination is noted as the most promising possibility for future work in the Master's thesis [Rie21].

The initial concept aimed to integrate sentiment analysis into graph creation to enhance outcomes. Regrettably, this approach did not yield the expected improvements. In fact, employing sentiment analysis to construct the graph as outlined failed to capture polarisation, at least from a visual standpoint.

As a result, we have decided to adopt a different approach. Each user will now be embedded into a corresponding vector, and these embeddings will be utilized to calculate the degree of controversy. We will try to apply the approach used in this paper[OdZDGFB20], adapting it to the specific differences of *derStandart.at*, because there is a big difference between Twitter, where users write individual posts on different topics, and discussions to an online article.

First, we gather all posts made by a specific user within the discussion. These posts are preprocessed where unnecessary symbols are removed to ensure they are suitable as training data. Subsequently, we encode these cleaned posts into a single vector per user. This vector encapsulates both syntactic and semantic features of the text, precisely aligning with our goal of content analysis within posts.

There are different embedding techniques, ranging from simple methods to deep language models. Simple techniques include Bag-of-Words models and TF-IDF, which are based on the frequency of words. On the other end of the spectrum are deep language models such as BERT or GPT, which rely on neural networks and can capture complex syntactic and semantic relationships within the text.

For our approach, we have selected FastText because it strikes a balance between simplicity and performance. FastText, developed by Facebook's AI Research (FAIR) lab, is an extension of the Word2Vec model. Unlike traditional word embeddings that treat each word as a single entity, FastText breaks down words into n-grams (subword units). This allows the model to generate embeddings for words based on their constituent parts, which is particularly useful for handling out-of-vocabulary words and capturing morphological nuances of a language.

The model works by learning vector representations for both words and their subwords. For example, the word "playing" might be broken down into subwords like "play," "lay," "ayi," and "ing." This approach helps in creating more meaningful embeddings, especially for languages with rich morphology or where compound words are common.This helps in handling rare words and misspellings more robustly. This makes it a suitable choice for analyzing user posts on *derStandard.at*, where language can be diverse and nuanced.Additionally, FastText has pretrained models [GBG+18], that are trained on Common Crawl and Wikipedia for German, which is highly beneficial for our analysis of posts on *derStandard.at*. These pretrained models provide a solid foundation for embedding the textual data, ensuring that both common and rare words are represented accurately.

3.5.1 Model Training

Finetune pre-trained model

To achieve better results, we need to fine-tune the FastText model. While pretrained models provide a strong starting point, fine-tuning allows us to adapt the embeddings to the specific characteristics and nuances of the text data from *derStandard.at*.

For the time being, the FastText model that is pretrained can only be fine-tuned in a supervised manner. This means that to adapt the pretrained model to our specific dataset from *derStandard.at*, we need to use labeled data during the training process. For this purpose, we will use the results of the graph partitioning phase. Users' posts will be lebeled based on their respective communities, forming the structure of our training dataset.

```
__label__c1 Ihre Kritik an der aktuellen Politik ist berechtigt.
__label__c2 Diese Regierungspolitik ist ein Segen für unser Land.
```

Figure 3.7: Example of training data

To ensure that the model is not biased towards the larger community, we will balance the dataset by reducing the size of the larger community to match that of the smaller one. This will involve randomly sampling posts from the larger community until it has the same number of posts as the smaller community. By doing so, we create a more balanced training dataset, which helps prevent the model from being skewed towards the more dominant community and ensures a fairer and more accurate fine-tuning process.

Training the Model from Scratch

Another option is to create and train a model from scratch using fastText. We utilize the same data for training the model as mentioned in the fine-tuning section above. It's essential to train the model for an adequate number of epochs, especially since the size of the training data is typically not very large.

3.6 Controversy Score Computation

In this section, we will use the results obtained from the previous stages to calculate the level of controversy, representing it as a positive number.

3.6.1 HITS Algorithm

Firstly, we will employ the **HITS algorithm** to compute the authority and hub scores of users. The **HITS algorithm**, also known as Hyperlink-Induced Topic Search, evaluates the importance of nodes within a graph based on two metrics:

- **Authority score**: Estimates the node's importance within the network.
- **Hub score**: Measures the value of its relationships to other nodes.

After calculating these scores for all users, we will select the top 30% of users based on their hub score and the top 30% based on their authority score. These selected users will be referred to as central users.

In summary, the HITS algorithm helps us identify central users by quantifying their authority and hub scores, reflecting their importance and connectivity within the network of users.

3.6.2 Centroids

In this subsection, we are using the embeddings of users to calculate the centroids of each cluster c_1 and c_2, and a global centroid c_{glob}. Each user is represented as $x_i \in \mathbb{R}^k$ and $y_i \in \{1, 2\}$ represents the community of the user (C1 or C2).

The centroid c_j for cluster C_j is calculated as:

$$c_j = \frac{1}{|C_j|} \sum_{i:y_i=j} x_i$$

The global centroid c_{glob} is calculated as:

$$c_{\text{glob}} = \frac{1}{|C_1| + |C_2|} \sum_i x_i$$

Centroids play a crucial role in clustering analysis by providing representative points that summarize each cluster's characteristics. They serve as reference points for cluster interpretation, similarity comparison, and further analysis of user behaviors within distinct communities.

3.6.3 Distances

Let x_i denote the embedding vectors and c_j denote the centroids, where $i = 1, 2, \ldots, n$ and $j = 1, 2$.

The Euclidean distance $ed(x_i, c_j)$ between an embedding x_i and a centroid c_j is given by:

$$ed(x_i, c_j) = \sqrt{\sum_{k=1}^{d} (x_{ik} - c_{jk})^2}$$

where d is the dimensionality of the embeddings.

The sum of distances D_j for centroid c_j is defined as:

$$D_j = \sum_{i:y_i=j}^{n} ed(x_i, c_j)$$

where n is the number of user embeddings.

We define D_{glob} as the sum of distances between all the embeddings and the global centroid c_{glob}.

$$D_{\text{glob}} = \sum_{i}^{n} ed(x_i, c_{\text{glob}})$$

3.6.4 Controversy Score

We define the controversy score r as follows:

$$r = \frac{D_1 + D_2}{D_{glob}}$$

The score represents the extent to which the clusters are separated. If they overlap, the result should be close to 1, if they are clearly separated, that is, the topic is controversial, the result should be different from 1. By definition, r should be positive because D_1, D_2, and D_{glob} are positive as well.

CHAPTER 4

Experiments

In this section, we will apply our proposed method for measuring the polarisation level on real articles from der Standart.at. Then we comment and analyze the measurement results. We also offer various variants to possibly improve the accuracy of the results.

4.1 Article Selection

To be able to understand whether our method makes a real difference between controversial and non-controversial discussions, we will select 10 controversial and 10 non-controversial ones, for which we will calculate the controversy score.

It's important to acknowledge that determining whether the discussion under a given article is controversial relies heavily on our intuition. This involves reading through the posts in the discussion. However, this method has its limitations. Firstly, it is inherently subjective, as it depends on individual interpretation. Secondly, it is impractical to read every post in its entirety, given that discussions often contain hundreds or even thousands of posts. A more effective approach could involve determining whether a discussion is polarized through empirical analysis by a larger group of people. By aggregating assessments from a diverse set of individuals, we can achieve a more objective and accurate evaluation of the discussion's nature.

Some of the selected articles, as well as their controversial score, can be seen in Table 4.1

4.2 Network Creation Method

Given that the graphs created for an article significantly influence the final results, it is crucial to choose an effective method for their creation. This is particularly important for embeddings generated from FastText, as we rely on the communities identified by the Louvain algorithm to train the model.

The best way in terms of accuracy as well as representativeness is the hybrid graph. Detailed information on how it is created can be found in Section 3.3.4. For all calculations, we use this particular method to create the graph, as it provides the best results in initial graph tests.

article_id	**article_title**	**intuition**	**r**
2000075994165	Shitstorm gegen Strache, nachdem er NS-Verbrechen verurteilte	yes	0.5426
2000038582887	SP-Klubchef Schieder: "Keil zwischen Lopatka und die FPÖ getrieben"	yes	0.3001
2000004413943	Separatisten bestätigen militärische Unterstützung aus Russland	yes	0.1705
1397522511244	Eltern gegen Kirchenlieder im Musikunterricht	yes	0.3196
2000126868214	Wie hoch ist die österreichische Covid-19-Impfbereitschaft?	yes	0.1826
2000125419254	SPÖ Burgenland verhindert geplante Bundesrats-Blockade für neues Covid-Gesetz	yes	0.3174
2000122301927	Servus TV und das Futter für die Covidioten	yes	0.3653
1348286007406	Musikzeitschrift "Guitar World" kürte den besten Gitarristen aller Zeiten	no	0.7932
2000121487115	Die schönsten Film-Happy-Ends?	no	0.9625
1397521819673	Musik made in Austria – welche Bands hören Sie?	no	0.4342
2000112621927	Merry gegen Happy: Alljährlicher Weihnachtskrieg in den USA	no	0.7159

Table 4.1: Results from some Experimental Articles

4.3 Embeddings

We offer two methods for embedding user posts. The first method utilizes a pretrained FastText model for German. The second method involves training a model from scratch specifically for each article.

We tried both methods and concluded that to achieve significant results, a new model specialized for the specific article must be used. Using a pretrained model on a large corpus of information fails to differentiate between posts under a specific article, as they appear very similar due to their shared topic. When using the pretrained model, we consistently obtained results close to 1, indicating that it classifies every article as non-controversial. This failure to capture the differences between the two communities results in the distance between the two centroids remaining almost zero. This issue is also exacerbated by the lack of sufficient posts from each article to fine-tune the model effectively.

For these reasons, for these experiments we create a new model for each individual article, and train it with posts from it.

4.4 Controversy Score Results

We conducted an analysis on a balanced dataset, comprising an equal distribution of controversial and non-controversial articles. Analyzing the statistics in Figure 4.1, it is evident that our method effectively distinguishes between controversial and non-controversial discussions. It's worth noting that the more controversial or polarized a discussion is, the lower its score, whereas less polarized discussions tend towards a score closer to 1. However, there are opportunities for improvement, which we discuss in the next section.

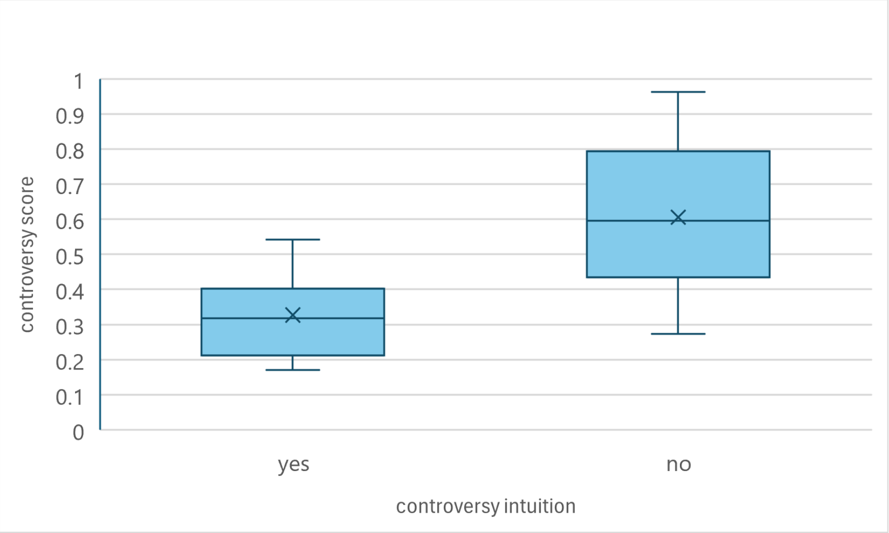


Figure 4.1: Comparison between Controversial and Non-Controversial Articles and Their Results

CHAPTER 5

Conclusion

In conclusion, the analysis of controversy and polarisation in online discussions plays a crucial role in understanding the dynamics of social media platforms. These platforms serve as arenas where diverse opinions converge, influencing public discourse and shaping societal perspectives. Controversy, while sometimes associated with negative outcomes such as hate speech and harassment, also serves a vital function in fostering meaningful debate and challenging echo chambers. Studies underscore the risks of polarisation in digital spaces and highlight the importance of measuring controversy to enhance media literacy and promote balanced consumption of information. By assessing the degree of disagreement and understanding its implications, we can mitigate the adverse effects of polarisation and cultivate environments where diverse viewpoints are respected and understood.

We conducted an extensive literature review focusing on online polarisation, addressing various related issues and exploring different measurement methods. These methods include those based solely on graph structure, those focused exclusively on content analysis, and hybrid approaches that integrate both structural and content-based metrics. Each method offers unique insights into the complex dynamics of online polarisation, highlighting the importance of considering network structure, ideological content, and their interactions for a comprehensive understanding of this phenomenon.

In this thesis, we proposed a method for measuring the polarisation level of discussions on *derStandard.at* articles using a controversy score. Our approach aimed to distinguish between controversial and non-controversial discussions by analyzing user posts through graph-based techniques combined with content-based methods. The controversy score defined as $r = \frac{D_1+D_2}{D_{glob}}$, provided a quantitative measure of the polarisation within the discussion.As we demonstrate in Section 4.4, it provides real results, succeeding in distinguishing controversial from non-controversial debates.

Furthermore, we have proposed four different graph creation methods, with varying degrees of success. Since the graph is the basis of the whole method, it is very important for it to be able to capture the real communities in a discussion. As the best of these methods, we determined the hybrid method, which uses both the votes and the posts of the users, managing to capture the different communities in a reliable and representative way.

We also discussed different approaches to defining topics. In our thesis, we treated each individual article as a topic, though we explored alternative definitions. Broadening the scope of topics could potentially enhance results, making it a promising avenue for future research.

CHAPTER 6

Future Research and Limitations

In this section, we outline potential avenues for enhancing our method and expanding its capabilities. We discuss various strategies to improve the accuracy and effectiveness of our approach.

6.1 Different Embedding Techniques

For embedding user posts, we currently use two methods, both based on FastText. Given the critical role that post embeddings play in the final result, employing different techniques could significantly impact the outcomes. In the paper [OdZDGFB20] whose method we apply, the authors tested their approach not only with FastText but also with BERT.

There are numerous word embedding techniques that could be explored. For example, GloVe (Global Vectors for Word Representation) is a popular method that generates word embeddings by analyzing word co-occurrence statistics in a large corpus. Another advanced technique is ELMo (Embeddings from Language Models), which generates context-dependent embeddings for each word, capturing the word's meaning based on the surrounding text. Transformer-based models like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers) also offer powerful embedding techniques that can understand context and nuances in text more effectively.

Testing these different embedding techniques could lead to significant improvements in our method by enhancing the model's ability to capture the subtleties in user posts, ultimately improving the accuracy of distinguishing between controversial and non-controversial discussions. This exploration presents a promising direction for future work and the potential enhancement of our approach.

6.2 Topic Definition

Section 3.2 justifies our focus on a single article rather than broader definition of a topic. However, we do not rule out the possibility of analyzing and obtaining results with a broader topic definition, such as multiple related articles based on keywords. The primary challenge with broader topics lies in constructing the graph, as connecting separate graphs into one cohesive structure presents difficulties. This limitation pertains specifically to the graphs we currently utilize. The graph in our method is mainly used for training the model; therefore, with a new and improved graph construction technique, the issue of integrating individual articles can be resolved. We consider the analysis of larger groups of articles a promising avenue for enhancing our method.

6.3 Enhanced Content-based Graph

To create the content-based graph, we use a pre-trained model trained for German sentiment analysis [GSBB20]. When we talk about sentiment analysis, we refer to classifying a text as positive, neutral, or negative. One reason this graph-based method might not work well is the nature of the content being analyzed. Political media posts, unlike product reviews, cannot be easily categorized as simply positive or negative. Furthermore, the sentiment of each post might not be standalone but rather contextually dependent on whether the post agrees with or opposes the preceding comments. This complexity makes the sentiment analysis of political media posts more challenging, as it requires understanding the interplay and progression of sentiments within the discussion.

If the model used for the sentiment analysis is fully trained on the text of posts in *derStandart.at* and the posts are considered in context, this would significantly improve the results of this graph creation method.

List of Figures

List of Tables

Bibliography

[cam] Cambridge Dictionary. `https://dictionary.cambridge.org/`.

[DGL12] Pranav Dandekar, Ashish Goel, and David Lee. Biased assimilation, homophily and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, September 27 2012.

[GBG$^+$18] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[GDFMGM18] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.

[Gil00] Rosalind Gill. Discourse analysis. In Martin W. Bauer and George Gaskell, editors, *Qualitative Researching with Text, Image and Sound*, volume 1, pages 172–190. Sage Publications, 2000.

[GSBB20] Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. Training a broad-coverage german sentiment classification model for dialog systems. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1620–1625, Marseille, France, May 2020. European Language Resources Association.

[KZN$^+$21] Juhi Kulshrestha, Muhammad Zafar, Lisette Noboa, Krishna Gummadi, and Saptarshi Ghosh. Characterizing information diets of social media users. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1):218–227, Aug. 2021.

[MMT18] Cameron Musco, Christopher Musco, and Charalampos E. Tsourakakis. Minimizing polarization and disagreement in social networks. 2018.

[MTT17] Antonios Matakos, Evimaria Terzi, and Panayiotis Tsaparas. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, 31(6):1480–1505, 2017.

[MZDC14] Yelena Mejova, Amy X. Zhang, Nicholas Diakopoulos, and Carlos Castillo. Controversy and sentiment in online news. Technical report, Qatar Computing Research Institute, University of Maryland, MIT CSAIL, 2014.

[OdZDGFB20] Juan Manuel Ortiz de Zarate, Marco Di Giovanni, Esteban Zindel Feuerstein, and Marco Brambilla. Measuring controversy in social networks through nlp. 2020.

[Rie21] P. O. Riemer. Measuring polarization in an online news forum, 2021.