Explaining News Spreading Phenomena in Social Networks

From Data Acquisition and Processing to Network Analysis and Modelling

vorgelegt von

M.Sc.
Daniel Thilo Schroeder

an der Fakulät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr. Ing. -

Promotionsausschuss:

Vorsitzender:	Prof Dr. Jan Nordholz
Gutachter:	Prof Dr. Odej Kao
Gutachter:	Prof Dr. Anis Yazidi
Gutachter:	Prof Dr. Johannes Langguth
Gutachterin:	Prof Dr. Giulia Boato

Tag der wissenschaftlichen Aussprache: 19. Januar 2022

Berlin 2022

Acknowledgments

First of all, I am indebted to Prof. Kao, not only for supervising this thesis but also for the time I was allowed to work as a student assistant at the chair for distributed systems at the Technical University of Berlin. Prof. Kao's recommendation only made all this possible, for that thank you. Furthermore, I have to thank Prof. Yazidi and Prof. Boato, who agreed to act as reviewer.

I now work 3.5 years on my Ph.D. at Simula, and during this time, I was allowed to meet numerous beautiful people, many of whom I have taken deeply into my heart. First of all, I would like to thank Johannes for allowing me to write this thesis within the UMOD project and mentor me during this exciting period of my life. Thank you, Johannes, it is a pleasure to work with you! Moreover, I would like to express my gratefulness towards the entire HPC group at Simula Research Laboratory, especially Petra and Konstantin. Moreover, Michael Kreil deserves a special mention for sharing his knowledge with us and providing guidance throughout my entire journey.

There were a few moments when I doubted. I have to thank the entire HOST group at OsloMet but especially Pål and Michael, for being there for me in those moments and welcoming me into their group. Without you, I probably would not have been able to finish this. Thank you, Saeed, for your hospitality in Iran!

My gratitude goes to the CIT or DOS group at TU-Berlin. I have the feeling to know some of you all my life. Especially Sören, Marcel, Anton, and Florian have to be mentioned here. Thanks for the first paper Flo, that really helped a lot! And thank you Kevin for the many pleasant conversations. Thanks, Tobi, for everything I learned from you. Finally, thank you, Jana, for the many years you helped with simply everything.

Special credit goes to Aigars, Andreas, Emil, Emily, Ferdinand, Haseeb, Kaspara, and Max, who decided to write their thesis with me. For me, it was, and still is, a great time, and I wish you all only the best! I would especially like to thank Luk, who has been by my side since day one. I truly appreciate you and your work and hope to watch you grow for a long time to come.

Before stopping to express my appreciation towards the people I have worked with, I want to emphasize two of my mentors in particular: Pedro and Carsten. Pedro for the discussions and ideas and especially for helping me with my first journal, and Carsten for giving me advice and support all along the way. I will never forget that; sounds like a threat, but its not :-)! In my little world, and most likely not just there, you guys are by far the best seniors alive.

Finally, I thank my family, especially my father, Heino, who has stood behind me all my life, and my two brothers, Simon and Jonas, simply for being alive. I could not imagine better siblings. Thank you Yannick for being my best friend for the past 100 years and for washing my head from time to time. Thank you Karen; by the time you read this, we will both be married, and I want you to know that was the best decision I ever made in my entire life. I have never met anyone that made me feel so secure, and I look forward to every day, we both wake up next to each other.

Abstract

When a high-ranking British politician was falsely accused of child abuse by the BBC in November 2012, a wave of short messages followed on the online social network Twitter leading to considerable damage to his reputation. However, not only did the politician's image suffer considerable damage, moreover, he was also able to sue the BBC for £185,000 in damages. On the relatively new media of the internet and specifically in online social networks, digital wildfires, i.e., fast spreading, counterfactual or even intentionally misleading information occur on a regular basis and lead to severe repercussions. Although the example of the British politician is a simple digital wildfire that only damaged the reputation of a single person, there are more complex digital wildfires whose consequences are more far-reaching. This thesis deals with the capture, automatic processing, and investigation of a complex digital wildfire, namely, the Corona and 5G misinformtionsevent - the idea that the COVID-19 outbreak is somehow connected to the introduction of the 5G wireless technology. In this context, we present a system whose application allows us to acquire large amounts of data from the online social network Twitter and thus create the database from which we extract the digital wildfire in its entirety. Furthermore, we present a framework that provides the playing field for investigating the spread of digital wildfires. The main findings that emerge from the study of the 5G and corona misinformation event can be summarised as follows. Although published work suggests that a purely structure-based analysis of the information spread allows for early detection, there is no way of predictively labelling spreading information as probably leading to a digital wildfire. Digital wildfires do not emerge out of nowhere but find their origin in a multitude of already existing ideas and narratives that are reinterpreted and recomposed in the light of a new situation. It does not matter if ideas and explanations contradict each other. On the contrary, it seems that it is the existence of contradictory explanations that unites supporters from different camps to support a new idea. Finally, it has been shown that the spread of a digital wildfire is not the result of an information cascade in the sense of single, particularly influential short messages within a single medium. Rather, a multitude of small cascades with partly contradictory statements are responsible for the rapid spread. The dissemination media are diverse, and even more so, it is precisely the mix of different media that makes a digital wildfire possible.

Zusammenfassung

Als ein hochrangiger britischer Politiker im November 2012 von der BBC fälschlicherweise des Kindesmissbrauchs beschuldigt wurde, folgte eine Welle von Kurznachrichten im sozialen Online-Netzwerk Twitter, die zu einer erheblichen Schädigung seines Rufs führte. Doch nicht nur das Image des Politikers erlitt erheblichen Schaden, er konnte die BBC auch auf 185.000 Pfund Schadensersatz verklagen. In den relativ neuen Medien des Internets und speziell in den sozialen Online-Netzwerken kommt es regelmäßig zu sogenannten Digital Wildfires, d.h. zu sich schnell verbreitenden, kontrafaktischen oder sogar bewusst irreführenden Informationen, mit schwerwiegenden Folgen. Obwohl es sich bei dem Beispiel des britischen Politikers um ein einfaches Digital Wildfire handelt, welches nur den Ruf einer einzelnen Person schädigte, gibt es komplexere Digital Wildfires, deren Folgen weitreichender sind. Diese Arbeit befasst sich mit der Erfassung, automatischen Verarbeitung und Untersuchung eines komplexen Digital Wildfires, nämlich dem Corona- und 5G-Misinformationsevent - der Idee, dass der COVID-19-Ausbruch irgendwie mit der Einführung der 5G-Mobilfunktechnologie zusammenhängt. In diesem Zusammenhang stellen wir ein System vor, dessen Anwendung es uns ermöglicht, große Datenmengen aus dem sozialen Online-Netzwerk Twitter zu erfassen und so die Datenbank zu erstellen, aus der wir den digitalen Flächenbrand in seiner Gesamtheit extrahieren. Außerdem stellen wir ein Framework vor, welches das Spielfeld für die Untersuchung der Ausbreitung von Digital Wildfires bietet. Die wichtigsten Erkenntnisse, die sich aus der Untersuchung des 5G- und Corona-Misinformationevents ergeben, lassen sich wie folgt zusammenfassen. Obwohl veröffentlichte Arbeiten darauf hindeuten, dass eine rein strukturbasierte Analyse der Informationsausbreitung eine frühzeitige Erkennung ermöglicht, gibt es keine Möglichkeit, die Informationsausbreitung im Voraus als wahrscheinlich zu einem digitalen Flächenbrand führend zu kennzeichnen. Digital Wildfires entstehen nicht aus dem Nichts, sondern haben ihren Ursprung in einer Vielzahl bereits bestehender Ideen und Erzählungen, die im Lichte einer neuen Situation neu interpretiert und zusammengesetzt werden. Dabei spielt es keine Rolle, ob sich Ideen und Erklärungen gegenseitig widersprechen. Im Gegenteil, es scheint, dass gerade das Vorhandensein widersprüchlicher Erklärungen die Befürworter aus verschiedenen Lagern dazu bringt, eine neue Idee zu unterstützen. Schließlich hat sich gezeigt, dass die Ausbreitung eines digitalen Flächenbrandes nicht das Ergebnis einer Informationskaskade im Sinne einzelner, besonders einflussreicher Kurznachrichten innerhalb eines einzigen Mediums ist. Vielmehr ist eine Vielzahl von kleinen Kaskaden mit zum Teil widersprüchlichen Aussagen für die rasante Ausbreitung verantwortlich. Die Verbreitungsmedien sind vielfältig, mehr noch, es ist gerade der Mix aus verschiedenen Medien, der zu einem Digital Wildfire führt.

Contents

List of Figures

List of Tables

List of Abbreviations

GDELT Global Database of Events, Language, and Tone

UMOD Understanding and Monitoring Digital Wildfires

NLP Natural Language Processing

OSN online social network

FACT Framework for Analysis and Capture of Twitter Graphs

APP Developer App

BFS breadth-first search

DFS Depth-first search

LM Louvain Method

SIR Susceptible Infected Recovered

LT Linear Threshold

IC Independent Cascade

VPN Virtual Private Network

API Application Programming Interface

CC Connecting Consciousness

DW Digital Wildfire

Chapter 1

Introduction

The rapid expansion of the internet and the associated increasing interconnectedness [1, 2] have led to news spreading globally in a matter of seconds [3–5]. Simultaneously, developments such as affordable mobile network connectivity and push message protocols led to more and more individuals being exposed to a seemingly endless stream of information. The speed and quantity with which this information pours in leaves less and less room for interpretation and reflection and thus influences the opinion-forming processes.

While public news agencies employ journalists, and thus content is subjected to scrutiny before publication [6,7], the same does not apply to social media. Here, unlike in news agencies, whose size seems modest, the number of potential sources is equal to the number of potential consumers [8]. Moreover, the protocol the dissemination follows differs. News is not only broadcast from a single source to those who consume it, but simultaneously shared by consumers. By considering any consumer as a source, one enters an ambit where news can spread exponentially fast, wide, and far. We know that circulating information is not always factual, but it also contains misinformation (see Definition 2.1.3). The dissemination of misinformation has harmful consequences and frequently poses a danger, with cases ranging from damage to reputation over damage to property to threats to human life. In fact, there are numerous documented cases of dangerous misinformation of which we present a selection in Chapter 2. We argue that the frequent occurrence of misinformation events and the circumstance that news consumers act as potential news sources lead not just to an exponentially fast, wide, and far spread but also implies an exponential growth of these potential threats.

Every user of social media services is an actor within a social network, and each such actor is a potential source of false information, fake news, as we will call it from

now on, misinformation. And although the causes for dissemination are many, for example, false news can be generated to influence certain opinions or serve narratives, but also simply out of ignorance and by chance, the consequences are far-reaching and often unpredictable [9–15]. Actors having the ability to act as source and sink simultaneously, which furthermore implies the existence of interaction patterns. These interaction patterns depend on the underlying social network structure, which may form the base for phenomena like echo chambers [16] or filter bubbles [17].

In simple terms, the increasing ubiquity and connectivity to the world's electronic networks, which allow anyone to consume and disseminate information in fractions of seconds, inevitably results in a flood of data that humans cannot manually analyze[1][2]. It is simply impossible to validate the entirety of tweets, messages, status updates, etc., by hand. Implementing mechanisms that allow reporting supposedly questionable content and subject it to a supervisory authority has proven weaknesses [18–20].

On the one hand, this is because the one who checks the content needs to be familiar with the dissemination context. Appropriately assessing the damage potential seems impossible otherwise. On the other hand, there is a time window for the assessment and countermeasures to mitigate harmful consequences effectively. Ultimately, all procedures of this kind are always on the borderline of censorship or even beyond and thus require careful consideration and public discourse.

In order to understand the spreading patterns and develop measures to stop misinformation, it is necessary to identify it in the first place. However, there are several challenges associated with the identification [21]. The same content can be interpreted differently, depending on whether it is consumed by a comedian's followers or those of a far-right party. However, differentiating between irony or satire and harmful misinformation is not the only challenge. Moreover, the danger that the dissemination of a certain content entails depends on the current external situation. For example, false reports about alleged terrorist attacks in Sweden [22] have become the subject of Islamophobic agitation precisely because the so-called refugee crisis in central Europe occurred simultaneously (2016). Thus, determining the author's intention, based purely on the content, becomes extraordinarily difficult. Especially if an author only shares the content instead of creating it, a message may be lifted out of context and misinterpreted by those that consume it, leading to the accusation of spreading misinformation against the author.

[1]Twitter in Numbers: `https://bit.ly/2RHy5cK`
[2]Facebook in Number: `https://bit.ly/3tIQCCG`

In order to stop the spread of dangerous misinformation before it leads to dangerous consequences, the development of automatic detection methods is inevitable. There are essentially four different approaches for automated detection of misinformation. First, one can assign a label or weight to the source itself. Therefore, keeping track of a source's or a source neighborhood's record of questionable content is necessary. Methods that generate danger assessment based on sources are called source-based methods [23–26]. Second, there are methods that analyze content at the text level. Here, unique linguistic characteristics or idioms serve as input for classification. Methods that fall into this category are known as style-based methods [27–29]. Third, the knowledge-based methods extract meaning from a particular statement and compare it to a knowledge graph [30, 31]. This method intends not just to classify but fact-check questionable content. In the fourth and last group are the propagation-based approaches which attempt to recognize the dubious content based on the way it spreads.

This dissertation aims to investigate news spreading phenomena, starting with the acquisition of large amounts of data from social media through its processing and analysis. A particular focus lies on the extraction and evaluation of data describing the spread of so-called Digital Wildfires (DWs). DWs are false news that spread remarkably quickly and have consequences in the real world. We find a DW, the so-called 5G corona conspiracy, study it qualitatively, and present a model to study its spread.

1.1 Defining Research Question and Problem Areas

The topic of this thesis is the design of methods to capture, detect, and understand spreading phenomena in online social networks (OSNs). Here, the focus is on analyzing the characteristics of the social networks in which the phenomena spread. The research question of this thesis is:

How do Digital Wildfires spread, and how to analyze their behavior?

We conceive the process to find answers to this research question as a triad involving the data acquisition, DW identification and analysis, and social network modeling.The latter includes the development of a model for dissemination analysis. Figure 1-1 shows the contributions made in each of these three fields, which we explain in more detail below.

Data Acquisition includes the evaluation and selection of suitable data sources. Sources are, e.g., online news agencies, online social media, and online forums. Moreover, it is important to target data acquisition in such a way that entire DWs can be captured. A large part of data acquisition is the processing and acquisition of large amounts of data. Data acquisition is particularly challenging in OSNs because there are often access restrictions or privacy settings that make access difficult. Furthermore, the structured storage of content from all aforementioned data sources is a challenge in itself. Appropriate relationships and interrelationships between data and across data sources need to be established to store data efficiently. Exploiting data sources also means, as in the case of Twitter, developing methodologies that allow collecting data at a large scale during the limited time when it is available.

Digital Wildfire Identification and Analysis deals with the problem of extracting complex DWs from a large amount of social media data. A complex DW in this context differs from a simple one in the multitude of message sources involving a variety of different actors as well as a multitude of different messages that promote the same narrative. In contrast, simple DWs start from a single source. To the best of our knowledge, there is no publicly available dataset containing a complex DW in its entirety. We argue that this is mainly due to the need to understand the content of large data sets. By definition, a DW is characterised by causing negative effects in the "real world" (see Definition 2.1.5). Thus, it is only possible to recognize a DW after real-world consequences occur. Therefore, identifying and extracting the data belonging to such a complex DW implies a feedback process, alternating between understanding the content, adjusting the filtering and detection criteria, and extracting further data. The word *understanding* in this context does not include the examination of network dynamics but should answer questions like: *Given the data before, after, and during the emergence of a DW, who is responsible for its creation? From which narratives do new narratives emerge that ultimately lead to the emergence of a DW? At what point do real-world consequences begin to emerge?*

Social Network Modeling deals with the process of developing models helping to understand the spread of a complex DW. Knowing how a DW spreads means knowing how a DW behaves compared to "everything else". This problem leads to the demand for network-based models that represent the reality outside of DWs and how to compare them with these spreading phenomena. The problem falls in the category of graph building and should be formulated as follows: *How and based on*

which properties can we model networks that allow deducing information exchange properties from OSNs? One should notice that these models cannot be based on the relationships actors have chosen to relate to each other, .i.e. the friend relationship on Facebook or follow relationship on Twitter. Instead, they must be designed based on information exchange derived from interactions.

1.2 Contributions

The main contributions of this thesis are in the fields of computational social science, complex networks, data mining, and applied machine learning. In addition, work in algorithms, programming languages, and high-performance computing has taken place in this thesis. Although there is work intersecting with high-performance computing, systems engineering, and programming languages, the primary focus of this thesis is in the area of computational social science with a focus on the study of complex networks. The main focus is the investigation of the diffusion dynamics of dangerous content, especially misinformation, DWs, and conspiracy theories in the arena of OSNs.

The Understanding and Monitoring Digital Wildfires (UMOD) project that provided the framework for this dissertation states in its project description:

> *Our project aims to develop improved prevention and preparedness techniques to counteract this type of misinformation.*

We claim that there is a logical way to approach the development of countermeasures for DWs. Naturally, the phenomenon of DWs needs to be studied first. Furthermore, we should determine the difference between the normal state and a DW. Only in the course of such a study can we understand the causes, spreading, and consequences. Based on the knowledge gained, characteristics can be derived that form a definition, which then distinguishes DWs from other information spreading patterns. It is this knowledge, in turn, that allows the development of methods for automatic detection. We have to consider that DWs occur in a space where time exists, and exponential growth can occur. It follows that the earlier we can detect a DW, the more likely it is that we can prevent real-world consequences. Therefore, we propose the following steps to approach the final goal: **Analyze**, **compare**, **understand**, **detect** and **develop measures**.

The contributions made in this dissertation belong to the first three categories, i.e., analysis, comparison, and deduction. However, starting from scratch, the main task

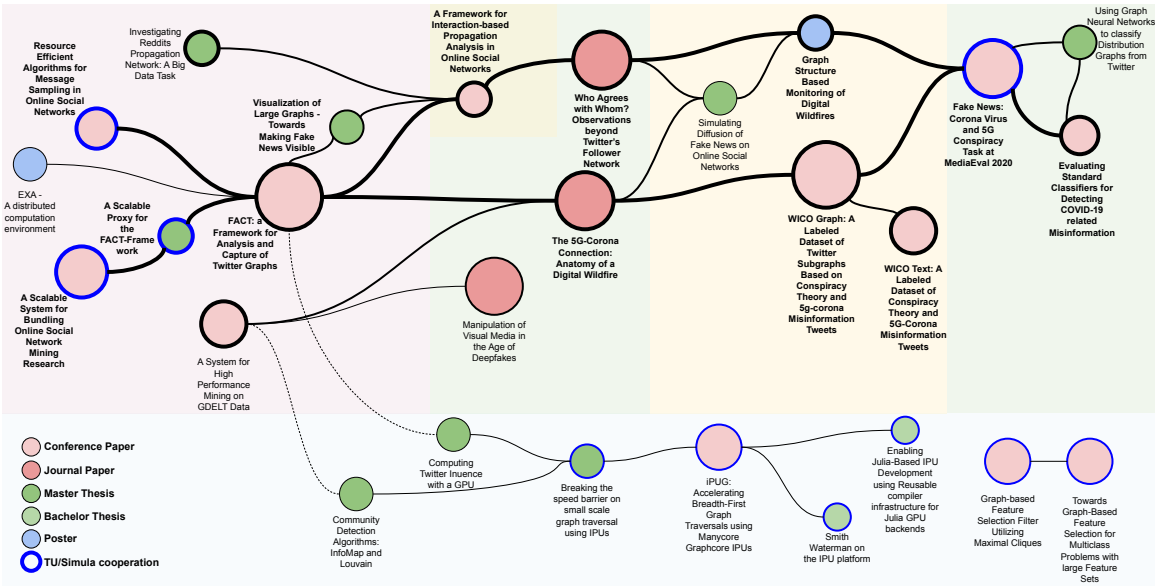

Figure 1-1: shows publications and theses delivered as part of this dissertation. The color red represents written publications in the form of conference papers (light red) or articles in journals (dark red). Blue circles are poster presentations, and the green nodes represent theses. Here, light green nodes present Bachelor theses, and dark green nodes present Master theses. Since the program in which this doctoral thesis took place is a collaboration between the Simula Research Laboratory and the Technical University of Berlin, all publications made in collaboration with researchers or students of the Technical University of Berlin have a blue border. The thickness of the edges and node borders represents the relevance of the respective contribution to this dissertation. The background colors illustrate to which areas the contributions belong to. A light pink background marks contributions in the area of data mining the data acquisition. The yellow area stands for data processing and open data sets, while the green area stands for data analysis and social network modeling. It is clear to see which scheme was used to create the contents of this dissertation. The publications on the left side show the evaluation of several data sources and the implementation of a system to acquire and analyze large amounts of data from the OSNs Reddit and Twitter and the news metadata service GDELT.

is to obtain the relevant data. This very first obstacle is often underestimated and takes a not insignificant part of the working time. This is also true in this case.

Data Source Evaluation and Data Acquisition for Digital Wildfire Examination

The selection of data sources for the investigation of DWs and especially for investigating their spread is not trivial. The most important factors are: *spread granularity*, *data accessibility* and *data quality*. Here, spread granularity refers to the degree of

refinement based on which the spread can be examined. If, for example, a network is constructed from the OSN Reddit, the exchange of information from actor to actor cannot be traced because no history of the conversations between the actors is apparent.

In the context of this thesis, we evaluate two OSNs, namely Reddit and Twitter, and a metadata collector for news sources, namely the Global Database of Events, Language, and Tone (GDELT). Continuously, GDELT provides metadata on nearly all news articles from major news agencies. Since the data is available for download, the main challenge left is processing the data flood. With "A System for High Performance Mining on GDELT Data" [32] we have presented a system capable of handling this flood on the supercomputer located at Simula Research Laboratory. The same is true for Reddit. Here, too, data is openly available and only needs to be processed. We made the first steps towards exploring this data source in the master thesis "Investigating Reddit's Propagation Network: A Big Data Task" by investigating the subreddits network from 2005 - 2019.

In regard to data quality, Twitter data is a reasonable choice because here, unlike Reddit or GDELT, we can track information distribution on an individual level. On the other hand, access is difficult, as Twitter's data is only accessible to a limited extent and it is besides, not in the appropriate format for postprocessing diffusion. Twitter allows access to its data only for a time window of two weeks. Given that the real-world consequences of a DW are not known until later, and that we have the ambition to capture a DW from its inception, this implies that the "hunt" for a DW takes place pro-actively in huge amounts of data where a lot of data is collected but never becomes part of a DW itself. The consequence: We must collect data in large quantities. Therefore, we introduce the Framework for Analysis and Capture of Twitter Graphs (FACT) for collecting, storing, and processing these large amounts of data. The following work was published in context of data acquisition and FACT in particular.

1. **FACT: A Framework for Analysis and Capture of Twitter Graphs** *DT Schroeder, K Pogorelov, J Langguth* The 6th International Conference on Social Networks Analysis, Management and Security (SNAMS-2019) [33]

2. **Resource Efficient Algorithms for Message Sampling in ONSs** *L Burchard, DT Schroeder, S Becker, J Langguth* The 7th International Conference on Social Networks Analysis, Management and Security (SNAMS-2020) [34]

3. **A Scalable System for Bundling ONS Mining Research** *L Burchard, DT*

Schroeder, K Pogorelov, S Becker, E Dietrich, J Langguth small The 6th International Workshop on Online Social Networks Technologies (OSNT-2020) [34]

4. **A System for High-Performance Mining on GDELT Data** *K Pogorelov, DT Schroeder, J Langguth, P Filkukova* Parallel and Distributed Processing for Computational Social Systems (ParSocial-2020) [35]

Digital Wildfire Identification and Analysis: The Interplay of Understanding and Extracting

After evaluating various data sources, we concluded that Twitter is the most suitable for distribution analysis (see Section 3.1). Given the FACT system and thus a large amount of continuously growing data, the question of how to process this data arises. Catching a DW on Twitter means, first of all, to be "lucky." Since Twitter allows accessing large amounts of historical data only for a limited period, the challenge is to collect the correct data before the DW starts. In the course of the covid pandemic and the related flood of misinformation, we managed to collect, identify, and publish an almost complete DW, the so-called corona and 5g conspiracy. The process of capturing is a two-way interaction of incremental understanding and filtering or detecting data belonging to the DW. Here, the term understanding refers to a qualitative analysis including reading the published content and doing background research. We summarized the understanding part in a qualitative analysis of the corresponding data in the following publication.

5. **COVID-19 and 5G conspiracy theories: long term observation of a digital wildfire** *J Langguth, DT Schroeder, K Pogorelov, P Filkukova* the International Journal of Data Science and Analytics (JDSA-2021) [36]

Finally, as part of the MediaEval Challenge 2020, we hosted a challenge task focusing on the automatic detection of DWs. The task here was to be able to classify tweets and dissemination graphs from Twitter into one of the following categories: 5G conspiracy, No conspiracy, or Other conspiracies. Fourteen groups participated, including one student group led by us. The results are summarized in the following two publications.

6. **FakeNews: Corona Virus and 5G Conspiracy Task at MediaEval 2020** *K Pogorelov, DT Schroeder, L Burchard, J Moe, S Brenner, P Filkukova, J Langguth* MediaEval Multimedia Evaluation Benchmark (MediaEval-2020) [37]

7. **FakeNews: Corona Virus and Conspiracies Multimedia Analysis Task at MediaEval 2021** *K Pogorelov, DT Schroeder, S Brenner, J Langguth* MediaEval Multimedia Evaluation Benchmark (MediaEval-2020) [38]

8. **Evaluating Standard Classifiers for Detecting COVID-19 related Misinformation** *DT Schroeder, K Pogorelov, J Langguth* MediaEval Multimedia Evaluation Benchmark (MediaEval-2020) [39]

Open Datasets: Invitation for Collaboration to Scientists in Related Fields

We were able to extract the data that belongs to the corona and 5G conspiracy mentioned in the previous section. To the best of our knowledge, this is the first data set of this kind. We have published both tweet content and associated induced subgraphs as open datasets to encourage researchers in machine learning, graph processing, and related fields to study misinformation and its spread. Furthermore, we provide a series of baseline experiments using both Graph Neural Networks and state-of-the-art Natural Language Processing (NLP) classifiers.

9. **WICO Graph: A Dataset of Twitter Subgraphs based on Conspiracy Theory and 5G-Corona Misinformation Tweets** *DT Schroeder, F Schaal, K Pogorelov, J Langguth* International Conference on Agents and AI (ICAART-2021) [40]

10. **WICO Text: a Dataset of Conspiracy Theory and 5G-Corona Misinformation Tweets** *K Pogorelov, DT Schroeder, F Schaal, J Langguth* ACM Conference on Hypertext and Social Media (ACMHT-2021) [41]

Social Network Modeling: Interaction Networks and their topological Properties

We study the corona and 5G conspiracy as a representative for a DW. In the Twitter-based case, this implies that an investigation of the Twitter sphere should precede a DW investigation. We did exactly this by analyzing a Twitter dataset of about one billion tweets and retweets. As part of this analysis, we introduced so-called interaction networks. In an interaction network, the connection strength between actor pairs is based on the reaction time and the frequency of information exchange. The outcome is a network including more than thirty million users that

allows studying information propagation at a large scale. Thus, with the network obtained, we present a framework based on which the propagation of any kind of phenomena on Twitter can be studied compared to the normal state.

11. **A Framework for Interaction-based Propagation Analysis in Online Social Networks** *DT Schroeder, PG Lind, K Pogorelov, J Langguth* The 9th International Conference on Complex Networks and their Applications (CNA-2020) [42]

12. **he connectivity network underlying the German's Twittersphere: a testbed for investigating information spreading phenomena** *DT Schroeder, J Langguth, L Burchard, K Pogorelov, J Langguth, PG Lind* Nature Scientific Reports (SR-2022) [43]

Other Publications created as Part of this Thesis

Moreover, we published several other publications in this thesis that did not fit appropriately in the overall story. Nevertheless, this work is related to the News Spreading Phenomena topic or derived from work on this topic. It is also essential to mention the ten master and bachelor theses written in this thesis's context. These theses allowed the author to try new things, learn a lot and possibly publish them in the future in one form or another and thus let others participate. A very promising branch of new research includes the work on the IPU hardware, done in cooperation with TU Berlin students. Furthermore, the work on Graph-Based Feature Selection with members of TU Berlin's DOS group is promising and should be further developed. The final papers can be found in Figure 1-1. Further publications are listed here.

13. **Don't Trust Your Eyes: Image Manipulation in the Age of DeepFakes** *J Langguth, K Pogorelov, S Brenner, P Filkukova, DT Schroeder* Frontiers in Communication 2020 [44]

14. **Graph-based Feature Selection Filter Utilizing Maximal Cliques** *DT Schroeder, K Styp-Rekowski, F Schmidt, A Acker, O Kao* International Conference on Internet of Things: Systems, Management and Security 2019 [45]

15. **iPUG: Accelerating Breadth-First Graph Traversals using Manycore Graphcore IPUs** *L Burchard, J Moe, DT Schroeder, J Langguth, K Pogorelov* International Supercomputing Conference (ISC-2021) [46]

1.3 Outline of this Thesis

This thesis is further structured as follows.

Chapter II is the Background and Related Work section. We approach chronologically after the UMOD project's course (see above) the necessary basics to understand the remaining thesis. After reading this chapter, it should be clear why data sources were considered so valuable. Besides, we give a brief overview of news phenomena with a particular focus on digital wildfires. We also present the most important approaches for the automatic detection of these contents. The third part deals with the technologies needed for the FACT framework presented in the approach part. Besides, we present the most important technologies for classifying.

Chapter III deals with data acquisition. First, we evaluate the three data sources that we considered valuable: GDELT, Twitter, and Reddit. Here we discuss the possibility of creating networks (Section 3.1.1), access to data or data quantity (Section 3.1.2), and data quality (Section 3.1.3). After identifying Twitter as a promising data source, we discuss this data source in more detail and finally introduce the FACT framework. FACT is a system that helps to overcome the shortcomings of digital wildfire-oriented data collection. The framework subdivides into four layers, which we introduce in Section 3.3.2 - 3.3.4. Using the FACT framework, we generated three data sets, on the one hand, the basis for social network modeling in Chapter IV. On the other hand, both the graph and text data for the Corona and 5g misinformation event.

Chapter IV presents a method that enables the derivation of a network where interactions reflect more than binary edges labeling acquaintances. In particular, we define relations between pairs of Twitter users based on the frequency of shared content and the time elapsed between publication and sharing (Section 4.2.1). We argue that both frequency and response time are indicators of approval, and we later, in Section 4.2.2, compose what we call a large-scale interaction network based on these properties. Moreover, we present a first thorough topological analysis of the derived network in Section 4.3, from which it is possible to raise a hypothesis about specific groups of users, such as famous individuals or users who are more robust towards untruthful content.

Chapter V deals the 5G-Corona event, a real-world digital wildfire in which the claim that 5G wireless technology is related to the COVID-19 outbreak is stated and that we were able to observe in its entirety. We start with introducing the event and defining its scope in Section 5.1, including a timeline (Section 5.2) of the events we

considered to be part of the DW. In Section 5.3 we present the attacks on 5G equipment that happened as a consequence. We continue with a manual examination of the misinformation narratives in Section 5.4.2) and investigate the early Twitter posts in Section 5.4.1 to understand where the 5G-Corona misinformation event came from. In Section 5.5, we look into the role that other online sources played in the 5G-Corona misinformation event by analysing Youtube and other video platforms (Section 5.5.1 - 5.5.2 as well as news articles related to the event on GDELT (Section 5.5.4).

Chapter VI presents the quantitative analysis. In contrast to Chapter V where we only analyse parts of the 5G-Corona event in a qulitative manner, in Chapter VI we work with the entire dataset in order to investigate the development over time (Section 6.1.1), a location analysis (Section 6.1.2) and a sentiment analysis (Section 6.2). Finally, we explain the attempts for automated digital wildfire detection in Section 6.3. The latter does not aim to contribute to the field of misinformation detection. Instead, detection acts as a method for data enrichment, helping to identify as much context as possible for the introduced conspiracy.

Chapter VII summarizes and concludes the work presented and identifies directions for future work.

Chapter 2

Background

Contents

In this chapter, we first introduce the terms necessary to distinguish DWs from other misinformation. We approach the concept of DWs through a discussion of the terms fake news, disinformation, and misinformation. Furthermore, we introduce complex networks and their role in computational social science. In this context, we first introduce two types of networks, the follower network, and the interaction network, and discuss models to simulate information diffusion, i.e, the spread of

information in these networks. Here, the most important models are the Linear Threshold (LT) and the Independent Cascade (IC) model. We also present the main graph algorithms used in this thesis. These include breadth-first search (BFS), which is used as a subroutine, especially in Chapter 4 as well as community detection and graph centrality. In addition, we give a brief introduction to the operation of OSNs and Twitter in particular. Twitter, unlike other large OSNs such as Facebook or Instagram, provides extensive access to its data. At the same time it is popular enough to represent a sufficiently large cross-section of social events at a global level. Large parts of the FACT frameworks described in Chapter 3 have therefore been developed specifically for Twitter.

2.1 Definitions

2.1.1 Misinformation vs Disinformation vs Fake News

Scholars have discovered that up to 60% of social media content is shared without being opened [47]. As a result, educational curriculum materials for media literacy emphasize the need for users to read or view content before sharing [48]. However, not just the unwillingness to check content before sharing is a threat, but also the sheer amount of content being published. Because news content is no longer shared by news agencies or newspapers only, but by individuals that are part sources and part consumers, the amount of information including false information that each individual is exposed to has increased dramatically. Allcott and Gentzkow [49], for example, were able to show that in the run-up to the US-elections 2016, almost every U.S. citizen was exposed to what they called "fake news".

Even thought Allcott and Gentzkow use the term fake news to describe misleading or false information, there is no consensus on the precise definitions of misinformation, disinformation, and fake news [50]. David Lazer, who coined the term computational social science [51], defines fake news in his 2018 article "The science of fake news" [52] as

Definition 2.1.1 (Fake News) *fabricated information that mimics news media content in form but not in organizational process or intent.*

Lazer further argues that 20th-century information dissemination technologies have evolved from the set of rules that emerged in response to the propaganda disseminated during the First World War. Today, however, these rules are no longer enforceable

due to the sheer size of the Internet. As a result, news outlets which are not complying with these standards compete with those that do.

Nevertheless, the term fake news was reduced to absurdity during the presidency of Donald Trump and since then has become a colloquial term. Its inflationary use during this era has undermined the legitimacy of the press and ultimately led to the fact that some people lost confidence in it. Thus, they eventually lack the source that gives them the opportunity to distinguish "truth from fake news" [53]. However, in recent publications, the authors even go so far to assert that the term fake news lost its meaning entirely [4].

The European Association for Viewers Interests proposes ten categories (see Figure 2-1) while Wardle [54] argues that the reason we struggle with the term fake news is because it is about more than simply news. She presents an expanded definition of the term that provides a breakdown based on the following criteria. She concludes that there are seven types of potentially problematic content in our information ecosystem, namely:

1. **Satire**: Content that was not created to cause harm, but can be misleading.
2. **Misleading**: Information that is used in a misleading way to frame someone.
3. **Fraudulent**: Sources that merely purport to be authentic.
4. **Invented**: Overwhelmingly false and created with the intent to deceive or cause harm.
5. **Wrong context**: Authentic content associated with false information.
6. **Revised content**: Authentic content or images reworked with the intent to deceive.
7. **Incorrect entanglements**: Headlines, visual assets or captions do not match the content.

Other scholars, while agreeing that the term "fake news" developed mainly during the Trump era, describe the phenomenon as akin to propaganda [55] and thus impute intent to the author.

Definition 2.1.2 (Propaganda) *is the strategic use of communication and information to influence public opinion.*

Hobbs [50], for example, claims that the word propaganda is better suited to describe content constructed with the intention of changing the attitudes and behavior of large groups of people.

Other arguments pointing in favour of using the term propaganda for what is colloquial labeled as fake news are that propaganda is not necessarily political [56,57] and that it can be both truthful and "full of lies", i.e., the criterion for the definition is indifference to truth. Taylor [58] states

> "Propaganda is really no more than the communication of ideas designed to persuade people to think and behave in a desired way ... [by] persuading people to do things which benefit those doing the persuading, either directly or indirectly"

and aims towards a broader definition. Propaganda does not necessarily pursue only harmful goals but can appear, for example, in the form of satire [59] or advertising [60]. Finally, the latter is

> designed to influence the receiver of the message toward the point of view desired by the communicator and to act in some specific way as a result of receiving the message.

While the term fake news has only recently gained popularity, at least in the sense of imitating news media content, the terms misinformation and disinformation have been established for a long time.

The Wikipedia page for misinformation defines it as: "false, inaccurate, or misleading information" which is the same definition many scholars seem to adopt [52]. Unfortunately, the connection between information and misinformation is less discussed [61]. Fox [62] stated already in 1983 that information does not have to be true and further, that misinformation is a subset of information and still fulfills the purpose of information, which is to inform. Furthermore, the possibility that misinformation is just incomplete information is discussed [63] and additional types of misinformation like ambivalence and distortion are added. To the best of our knowledge, a definition that clearly delimits misinformation independently of information is still missing. Nevertheless, we adopt the following definition.

Definition 2.1.3 (Misinformation) *Statements about reality that are not actually true.*

Furthermore, Wikipedia defines the term disinformation as "false or misleading information that is spread deliberately to deceive", referring to the historical context [64] which sees the term as a derivative from the Russian word "dezinformatsiya"

Figure 2-1: The ten types of misleading information released by the European Association for Viewers Interests[1].

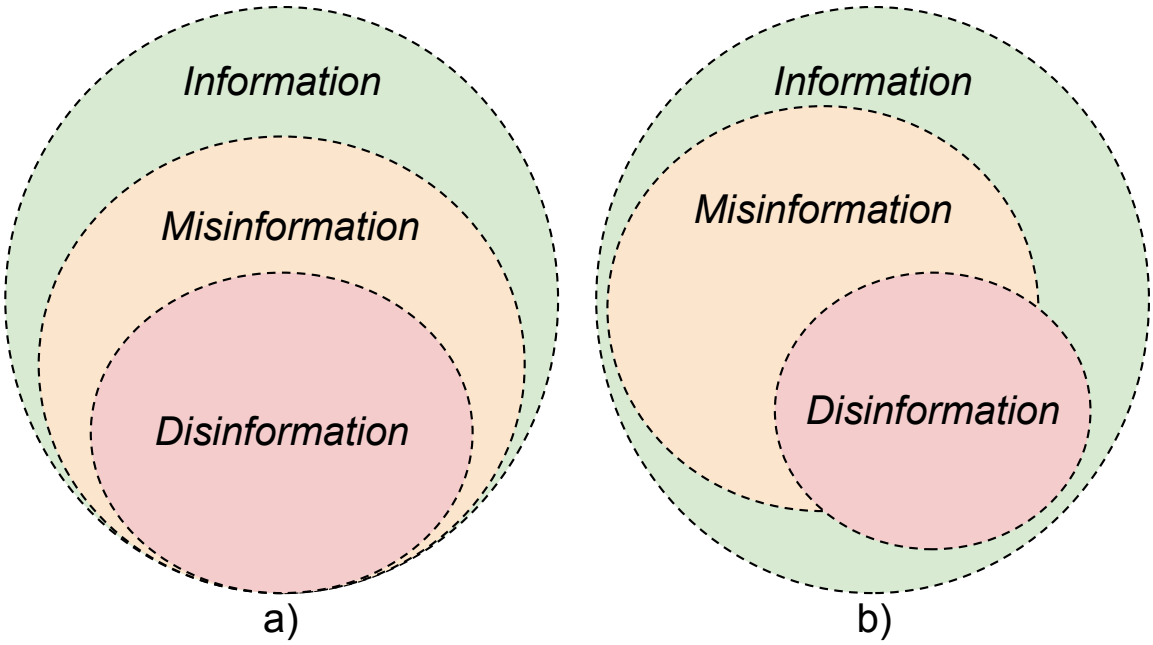

Figure 2-2: a) The subset relationship of information, misinformation and disinformation that we have chosen and that is generally accepted. b), the subset relation discussed in [61]. Since disinformation is not necessarily false here, it follows that disinformation is not a true subset of misinformation.

which first appeared around 1900. The intentions behind the spread of disinformation are many and varied. Control of the population, manipulation of shares or crypto prices, or simply the concealment of own misfortunes are only some examples. Traditionally, disinformation is understood as a subset of misinformation. However, according to [61, 65], disinformation does not necessarily have to be inaccurate but merely misleading. This idea implies that disinformation is not a true subset of misiniformation after all. Figure 2-2 depicts both definitions. In this thesis, we use the following:

Definition 2.1.4 (Disinformation) *Misinformation that is intentionally used to deceive.*

2.1.2 Digital Wildfires

The World Economic Forum places DWs, i.e. fast-spreading inaccurate, counterfactual, or intentionally misleading narratives that quickly permeate the public consciousness, among the top global risks in the 21st century [8]. The experts claim that the spread of misinformation on OSN may cause economic and political unrest.

Definition 2.1.5 (Digital Wildfires) *Social media events in which provocative content spreads rapidly and broadly, causing significant harm.*

Examples

Like a visit to the cinema where a false fire alarm goes off, and people are trampled to death at the exits before they realize that there is no fire, digital wildfires usually occur when misinformation spreads faster than its correction. Another reason for the emergence of a digital wildfire is the refusal to correct facts. The following two examples have been considered DWs and are excerpted from the World Economic Forum's report *Digital Wildfires in a Hyperconnected World*[2].

> **United Airlines** When a musician travelling on United Airlines had his claim for damages denied on a guitar that baggage handlers had allegedly broken, he wrote and performed a song – "United Breaks Guitars" – and uploaded it to YouTube, where it has been viewed more than 12 million times. As the video went viral, United Airlines stock dropped by about 10%, costing shareholders about US$ 180 million [9, 10].

> **Senior politician and child abuse** In November 2012, the BBC broadcast an allegation that a senior politician had been involved in child abuse, which transpired to have been a case of mistaken identity on the part of the victim. Although the BBC did not name the politician, his identity was easily discovered on Twitter, where he was named in about 10,000 tweets or retweets. On top of pursuing legal action against all people who spread this false information on Twitter, the injured politician settled on £185,000 in damages with the BBC.

2.2 Information Distribution

The essence of DWs is its rapid dissemination and thus spread of information. Information distribution can be modeled by considering individuals as stateful entities frequently exposed to information accumulated and processed to form opinions, which again serve as new input [66]. The connections through which the exchange of information and thus individuals' influence occur are in many cases part of the core

[2]World Economic Forum Report: `https://bit.ly/3skaI7s`

functionality of an OSN and are often called *friend-* or *follower* relationships. However, these edges are unweighted and thus signify only the presence of interest. They do not reflect how individuals actually interact or influence each other. Therefore, there is a need for obtaining weighted relationships from the communication that occurs between individuals. When looking at the diffusion of information or ideas in social networks, there is always a separation between the underlying network and the diffusion model acting on it.

2.2.1 Underlying Networks

A network or graph $G = (V, E)$ is a tuple including a set of vertices or nodes V and a set of edges connecting these vertices $E \subseteq V \times V$. Graphs are either directed, meaning the edge between two vertices v_i and v_j is a tuple (v_i, v_j) or undirected meaning the edge is a set $\{v_i, v_j\}$.

Definition 2.2.1 (Network) *$G = (V, E)$ with $E \subseteq V \times V$*

In a directed network, we consider the set of incoming neighbours for any vertex $v \in V$ as $n^{in}(v)$ and the set of of outgoing neighbours as $n^{out}(v)$. Correspondingly, v's in-degree is $|n^{in}(v)|$ and v's out-degree is $|n^{out}(v)|$. In an undirected graph, the degree of a node v is its number of edges $|n(v)|$ that connect to v. The total number of edges L in an undirected graph with N vertices is then

$$L = \frac{1}{2} \sum_{v \in V}^{N} n(v). \tag{2.1}$$

Many networks, especially those of natural origin, show structures that do not arise randomly. Rather, the degree distribution, the clustering coefficient, or community structure (see Definition 2.3.1) follow a certain pattern. We call these networks complex networks.

Definition 2.2.2 (Complex Network) *is a network with nontrivial topological properties, i.e., properties that do not occur in simple networks such as lattices or random graphs.*

Social networks are complex networks with a high clustering coefficient but a small diameter. In the following, we present the most important social networks for this work. After this, we focus exclusively on Twitter and its underlying social networks. We refer to Section 3.1 in which we explain this decision in detail.

Follower Networks

The follower network is specific to Twitter because Twitter calls the connections that two users have in between each other follower connections. However, the concept can be found, albeit in variations in other OSNs. Facebook, for example, offers a friend network according to their naming of the undirectional relationships users chose to have with each other. The follower network appears to be the most obvious way to construct networks from Twitter. After all, its structure is intrinsically integrated into the functioning of the OSN. As the name suggests, a follower network consists of a set of users representing the nodes of the network and a set of edges formed by the decision of one user to follow another. The mechanism enables a subscription to another user's content. The reasons for subscribing are manifold. In the following, we present reasons that cause one user to follow another.

For example, Pasman, in his book "The Social Logic of Likes" [67] found that it is more likely to follow people when there have been real-world encounters. This means there must be a social pressure to follow. In fact, this pressure is sometimes so intense that the content being shared plays a subordinate role; it is the expression of interest that leads to the decision to follow in this case.

In addition, Twitter interventions play a role in selecting who a user follows. For example, Twitter implements a recommender algorithm [68] that suggests new accounts for following users according to their interests. When creating a Twitter account, the onboarding, a process that has evolved over the years [69], initially preselected 20 accounts to follow automatically. By now, the user is prompted to actively select accounts to follow from a list sorted by topic.

Follower trains are another way to influence the user's decision of who to follow. Follower trains are tweets including a variety of mentions on different Twitter accounts. The deliberately polarizing spam-like tweet content catches a reader's interest and suggests that the mentions point to users with similar opinions (see Figure 2-3). According to Gupta et al. [70], follower trains are often used for political manipulation on Twitter. For example, it has been shown that pro-Trump follower trains have been highly effective in inflating follower numbers for the train riders,i.e., the accounts involved.

The practice of aggressive friending [71], which involves following arbitrary accounts in the hope of being followed back, also distorts the "purity" of a follower edge. To limit this practice, Twitter introduced a follow limit of first 2,000 and later 5,000 follows. If 5000 follows are reached, the user has to wait until he or she reaches

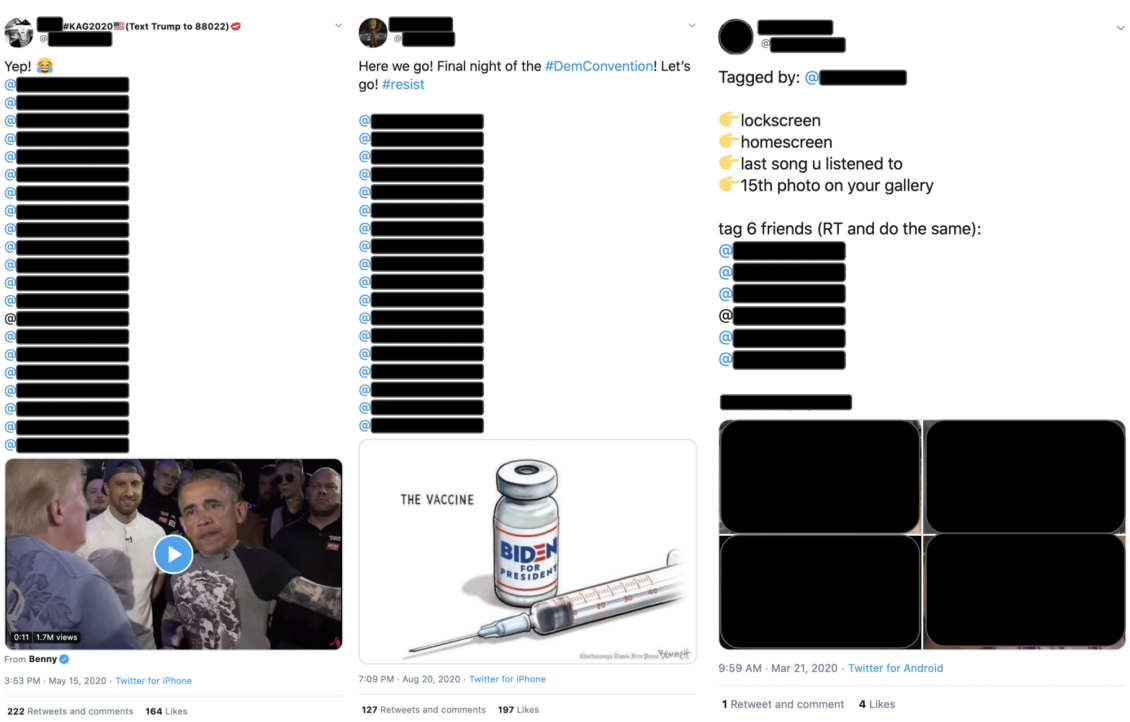

Figure 2-3: Figure from the Torres-Lugo paper on "The Manufacture of Partisan Echo Chambers by Follow Train Abuse on Twitter" [70]. Three tweet trains are shown, (left) a pro-Trump train conductor, (center) an anti-Trump train conductor, and the tagging game (right). The latter is a request to the tagged person to fulfill a certain task like posting a screenshot and tagging multiple additional users to do the same. Some account information is redacted to protect privacy.

more followers[3]. Moreover, the daily limit of follower requests is restricted to 400.

While in the early 2010s, it was still possible to examine Twitter's entire network of followers on request, Twitter no longer allows access to this data. In response, several data scraping methods and tools were developed. At times, methods like snowball scraping were used [72, 73]. Snowball scraping starts with an initial set of users and iteratively visits the respective neighbors of the users, noting which users have already been visited. Optionally, different criteria can be taken into account when selecting the initial accounts, such as the use of a certain hashtag or the specification of a place of residence. In addition, there are variations of snowball scraping where filters are applied during the process to consider users with specific characteristics. An example of a follower network is given by the study of the Norwegian Twittersphere [74].

[3]Follow limit: `https://bit.ly/3ctDmMw`

Interaction Networks

While follower networks do not account for user activity, one can also build networks based on interaction. Typical interactions on Twitter are replies, mentions, and retweets. Similar to the follower network, the nodes of an interaction network are accounts, while the edges represent communication. The main difference between the follower and interaction network is that one can assume that somebody who has either retweeted, replied, or be mentioned by somebody else has actually consumed the counterpart's content and has therefore been influenced by it. In a follower network, on the other hand, there is no guarantee that the one who has made the decision to follow is actually still aware of the counterpart's content. In fact, he or she would, in many cases, not even notice when the counterpart is not active anymore. Compared to follower networks, interaction networks may have weighted edges. One possibility for defining edge weights could be to simply count the number of interactions for a certain timeframe. We introduce a more sophisticated approach in Chapter 4 that takes into account not just the communication frequency but also the response time.

Collecting interaction networks from Twitter is less costly because one is not constrained by the very low Application Programming Interface (API) limit for follows (15 accounts per 15 minutes), but the necessary information is already included in the dataset of collected tweets, retweets, replies and mentions. The challenge in capturing interaction networks is the selection of tweets.

Bruns and Moon [72] evaluated a single day of tweets in 2019 based on their beforementioned Australian dataset by extracting mentiones as well as retweets. Despite the small period, individual clusters were visible that could be specifically assigned to individual events on that day. This shows that a network evaluation via interaction can be influenced by events in the study period and does not necessarily reflect the networking of the active accounts

Van Geenen et al. [75] evaluated a dutch mention network in which the mentions of accounts within tweets and retweets are used for network building. Using a combination of replies, mentions, retweets, and follows, Kelly et al. [76] built an interaction network depicting the Russian Twittersphere. To do this, they collected 50 million tweets in 2010 by using an existing index of Twitter accounts from the search engine Yandex. However, bias cannot be ruled out because the criteria used to create this index were not addressed.

Tie Strength

Substantial efforts have been made to understand the implications of tie strength. For instance, Granovetter, in his work "The Strength of Weak Ties" distinguished between the strong and weak ties of an individual and concluded that as the name suggests, weak ties are important [77]. However, what a weak tie is is difficult to define. There are two distinct approaches. One is more established from a network perspective. Weak ties are bridges not only between individuals but between communities. Therefore, a bridge is something like an ear into another world. Another is what one can call the first-person perspective. Here, the strength of a connection is defined by the characteristics of the two individuals. Pairs of persons who exchange information more often and talk longer are more strongly connected than those who exchange information less often.

Implications of the strength of the connections are numerous and far-reaching. For example, it has been shown that people with many weak ties have an easier time finding jobs [78], are more creative [79], and are often better situated [80]. On the other hand, strong ties are related to persons with stronger acquaintances, e.g., friends and family. An important fact that must always be taken into account deals with the *information content* when pairs of people communicate. This content is strongly correlated with the strength of the ties binding pairs of persons. People who tend to communicate over strong ties tend to know each other very well and therefore need less complex protocols and less communication accuracy in order to express themselves [81]. In other words, since there is a greater common prior knowledge, information needs less context to be understood.

Communication "short-cuts" also promote communication skills and intelligence. For instance, it has been reported that people with weaker ties, who need to explain complex content, must use more elaborate language skills and strategies [82]. Thus, it is of utmost importance to assess the strength of ties that bind pairs of individuals exchanging information.

2.2.2 Spreading Models

In networks, several dynamic processes occur, which we can describe as diffusion. The network serves as a medium over whose edges, depending on the application, different "things" flow. For example, the edges in a friendship network are "highways" for rumor or gossip [83] while the connections between computers are mediums for transporting data or computer viruses [84]. In a similar manner, virologists study the

spread of diseases through contact networks [85].

The basis for the various models that are used to explain spreading in networks is the Susceptible Infected Recovered (SIR) model, which we describe in the following paragraph. However, these simple spreading models do not take into account the structure of the underlying network. The two universally accepted models that do take into account the underlying network are the LT model and the IC model.

The basis, at least for the IC model, is the SIR model [86]. As the name implies the SIR model is based on the classification of actors into three pigeonholes of Susceptible S, Infected I and Recovered R. The SIR model is used to model processes such as diseases that the affected person can contract only once in his or her entire lifetime. However, the network structure and the complex human behavior both play an important role in the diffusion of these entities. Keeping this in mind, information diffusion models were proposed which took into account both of these parameters.

In the following, we consider a network $G = (V, E)$ with V as a set of nodes and E as a set of directed edges. Furthermore, we describe the incoming edges of a node v_i with $n^{in}(v_i)$ and the outgoing edges with $n^{out}(v_i)$. Both the LT and IC models operate in discrete time steps which we will denote as t. Each node $v_i \in V$ has a state denoted as infected or healthy, active or inactive, depending on the application. Since we are mainly concerned with social networks in this work, we will stick with active and inactive. If a node is inactive, it can be activated. The other way around, a change of the stautus from active to inactive is not possible, at least in the basic versions of the LT and IC models. Although there are non-progressive models, which allow a change of the state in both directions, here we present only the base models. We denote the set of active nodes at time t as X_t. The set of inactive nodes is then $V \setminus X_t$. Both, the LT and IC model have in common that the state change of a node v_i in some way dependeds on $n^{in}(v_i)$.

Linear Threshold Model

Regarding the diffusion of misinformation the SIR model does not reflect the idea of being convinced but not "infected" anymore. Somebody who believes in UFOs is likely to not change his or her mind quickly. In the Linear Threshold model there is no *R*ecovered state, here each edge $e_{u,v} \in E$ is assigned a weight given by the function $w : E \to [0, 1[$. It also holds $\forall e_{u,v} \in E | w(e_{u,v}) \leq 1$. Simultaneously the the sum of

incoming edge weights for a node v_i is less than or equal to 1, too.

$$\sum_{u \in V} w(e_{u,v_i}) \leq 1 \tag{2.2}$$

Furthermore, each node v_i selects a threshold Θ_{v_i} where Θ_{v_i} lies in $[0, 1]$ and is drawn from a uniform random distribution [87].

We investigate the dynamics over time where t is a particular point in time. X_t is the set of nodes that are activated at t or earlier. An inactive node becomes activated when

$$\sum_{u \in V} w(e_{u,v_i}) \geq \Theta_{v_i} \tag{2.3}$$

Each node is checked for activation in each time step t by adding up the weights of all its active neighbors. If the sum exceeds the threshold assigned to the node, it is also activated and contributes to the activation of its inactive neighbors in the next time step.

Independent Cascade Model

The Independent Cascade [88] model is a generalization of the SIR model. In contrast to the SIR model, however, the IC model does not assign a probability to each actor that determines whether it is going to be assigned to one of the SIR bags. Here, a bag refers to one of the categories *susceptible*, *infected* or *removed*. Instead, each edge $e_{u,v} \in E$ connecting nodes u and v is assigned a probability given by the function $P : E \to [0, 1[$. At each time t, from the set of nodes activated at time $t-1$, $X_{t-1} \subseteq V$, the outgoing edges are selected. So $n^{out}(v_i)$ for $v_i \in X_{t-1}$.

At each time t, the outgoing edges of the most recently activated nodes, i.e., at time $t-1$, are considered. Each of these edges $e_{v,u}$, if connected to a node u that has not yet been activated, then activates it with $P(e_{v,u})$.

2.3 Algorithms

2.3.1 Breadth-first Search

Breadth-first search and Depth-first search (DFS) are the most fundamental ways of traversing graphs. For sequential execution, the BFS algorithm is essentially defined by the data structure used to store the graph, as its fundamental operation is to iterate over the edges of a given vertex. However, parallel implementation of BFS,

particularly on distributed memory systems, is far more complicated. Consequently, there are far more possibilities for algorithm design and performance optimization.

While parallel BFS has been studied earlier [89], the topic gained widespread interest in the previous decade on distributed memory computers [90, 91], on shared memory [92,93], and on GPU systems [94]. The establishment of the Graph500 benchmark [95] in 2010 marks a turning point, since it encouraged direct comparability of results. This increased activity on the topic further, resulting in a large number of publications on that topic [96–100]. Furthermore, BFS implementations for GPUs have also received considerable attention in the recent years [101–104]. In addition to the parallel implementation, algorithmic improvements have been presented in the last decade. Possibly the most important among those was the introduction of direction optimizing searches [105]. At the same time, efficient parallel algorithms for BFS and DFS were also developed in the context of other graph problems, such as parallel matching algorithms [106–108].

2.3.2 Community Detection

Community detection, also known as graph clustering, is a form of identifying groups of nodes based on the topological properties of a graph [109]. In practice, community detection has a variety of applications. Netflix, for example, uses community detection methods in order to provide recommendations based on content users have watched before [110]. Facebook and Twitter use similar methods for friend or follower recommendations. Even though community detection has a long history reaching back to the early 70s, it was not until the rise of social media that the popularity of the field reached its peak.

Finding an appropriate definition for a community is still a challenge. While there is no universal definition, intuitively, one considers a set of nodes with more edges in between each other than to the "outside" as a community. Moreover, one could assume that the node properties in themselves play a significant role when assigning a node to a community. In this case, distances in Euclidean spaces, cosine similarities, or the Pearson correlation provide an adequate tool to assign nodes to communities [109, 111].

Definition 2.3.1 (Community) *is a subset of a network vertices in which the links between vertices are denser than in the remaining graph.*

Modularity

Although it is still not agreed upon how "good" and "bad" communities are distinguishable, it is widely accepted that there must be functions evaluating the quality of a community and with it the algorithm that identified the community in the first place. By far, the most established quality function is Girvan and Newman's modularity [112] which is given by the number of edges within a group, minus the expected number of edges in a similar graph, placed at random [113]. The more clearly communities emerge, the higher the modularity score. When comparing two community detection algorithms with the help of modularity, the algorithm with the highest score leads to communities that are more "defined". The highest possible modularity score is 1. Furthermore, modularity is defined as

$$Q = \frac{1}{2m} \sum_{i,j} \Big(A_{ij} - \frac{w_i w_j}{2m} \Big) \delta(c_i, c_j), \tag{2.4}$$

with $m = |E|$, w_i the sum of edge weights for node i and w_j correspondingly for node j. A_{ij} is the the entry in the adjacency matrix A for ij or the edge weight between i and j. c_i and c_j are the communities to which i and j are assigned. δ is the Kronecker delta function and returns 0 in case $c_i = c_j$ and 1 otherwise. Thus, δ is only one if both communities c_i and c_j are the same. In other words, Q is the the fraction of edges that fall within group c_i or c_j, minus the expected number of edges within the communities c_i and c_j for a random graph with the similar node and degree distribution. Barabási [114] made several observations using this definition.

Despite the popularity and versatility of the modularity score, there are some weaknesses, the most important of which should be mentioned. First, modularity is poor at scoring small communities. In addition, modularity is not robust. The result of the analysis often depends on single edges. Furthermore, the definition presented in Equation 2.4 assumes potential edges for each arbitrary pair of nodes i, j. The latter implies i.e. in case of large social networks, that each actor i knows about its potential counterpart j and is able to interact with it. If one assumes, for example Twitter's follower network, this is by far not true, in fact, the opposite is the case. Most actors will never meet or even know about each other.

Louvain Method

The Louvain method [115] is an approximation algorithm that tries to maximize the modularity score (see Equation 2.4). Furthermore, the Louvain method is capable of

determining the number of communities on its own. Since modularity maximization clustering is NP-Hard [116], different approximation algorithms based on heuristics were introduced [117, 118]. Clauset et al. [117] discusses several approaches and was able to show graph clustering for graphs up to ~ 120 million nodes. This constituted a significant improvement from previous clusterings which were limited to about 5 million nodes. The algorithm proposed by Clauset et al. works by recurrently merging communities in a way that the resulting modularity is maximized.

The Louvain method computes:

$$\Delta Q_{i,j} = \frac{1}{2m}\left(d_{i,j} - \frac{d_i d_j}{m}\right) \tag{2.5}$$

for each pair of communities. Here, m is the number of edges in the graph (see Equation 2.1) while i and j are communities with d_i and d_j the inner degree of i respectively j. The inner degree is defined as the set of edges that connect only nodes within this community. Then $d_{i,j}$ is the set of edges that connect nodes between i and j.

The Louvain algorithm can be divided into two phases that are repeated iteratevly. Starting with every node i being in its own community, in the first phase or modularity optimization phase, for each i in a random order, i is virtually assigned to every neighbour j's community. After assigning i to j's community we chose the i, j combination with the maximum $\Delta Q_{i,j}$. In case all $\Delta Q_{i,j} < 0$, we leave i in its own community. In other words, we try to assign i to the neighbouring community in a way that returns the highest modularity. We repeat phase one until no node gains by moving.

In the second phase, or community aggregation phase, the communities from phase one are merged into community nodes. The number of edges between two communities, before merging, becomes the edge weight in between the new community nodes. Edges from within the communities, previous to phase two, become selfloops with the previous number of internal edges as edge weights. This process is repeated until the number of communities does not change, i.e., $\Delta Q_{i,j} = 0$.

Leiden Algorithm

The Leiden algorithm is an improvement of the Louvain method, which according to Traag et al. [119], is prone to yield arbitrarily badly connected communities up to a degree where communities may even be internally disconnected. The Leiden

algorithm, compared to the Louvain method, is proven to result only in internally connected components as well as a better runtime performance.

The Louvain method assigns nodes to new communities in every but the last iteration. A node can even be assigned to a new community when before, in its old community, it acted as a bridge. Removing such a node from its old community, then leads to its disconnection. The appropriate reaction to such a disconnection should be a split into two (different) communities which were previously held together by the "bridge". Traag et al. were able to show that this phenomenon occurs not only in theory and argue further, that the issue of disconnected communities is only the "most extreme manifestation of the problem of arbitrarily badly connected communities".

The Leiden algorithm, named after the university in the Netherlands, addresses these issues and ensures that communities are well connected. The algorithm uses the smart local move [120], fast local move [121, 122] and random neighbour move [123] technique, by using heuristics for moving nodes between clusters in a novel way. This improves the modularity as well as the speed of the algorithm.

The algorithm can be roughly divided into three phases: First, local moving of nodes, the second refinement of the partition, and the third aggregation of the network based on the refined partition. When we compare these three phases with the two of the Louvain method, we recognize that Leiden introduces a second phase in between the moving and aggregation phase, called the refinement phase. Instead of aggregating a new representation from the partition that results immediately from the local moving phase, the Leiden method uses the partition resulting from the refinement phase. Dealing with two different partitions, the regular and the refinement partition, the refinement partition is calculated as follows. In the first place, like in Louvain, each node forms its own community. Later, only single nodes can be merged into a existing communities. Moreover, nodes in the refinement partition are only merged if connected in the regular partition, too. Then, in the refnement phase, nodes are randomly merged into a community that leads to an increase in the modularity function and not to the community leading to the highest increase. The randomness when selecting a community to merge with allows for a broader exploration of the partition space.

Markov Clustering

The main idea of Markov Clustering is to keep the edges that belong to communities and discard the edges that do not. Introduced in 2000 by van Dongen [124] as

part of his dissertation, Markov Clustering implements the idea that the probability a random walk ends in the same cluster is higher than the probability of ending outside its cluster. Moreover, Markov Clustering allows for clustering of graphs with edge weights, does not require prior knowledge of the cluster structure, is easy to understand, and cannot be misdirected by edges between different clusters.

The algorithm operates as follows: Initially, the adjacency matrix M is converted into a stochastic or flow matrix M^F. Here, for each node, the incoming and outgoing edges are counted and then a weight of $\frac{1}{\#Edges}$ is assigned to the entry for the corresponding edge.

Subsequently, so-called expansion and inflation steps alternate. In an expansion step, the flow matrix M_t^F at t is multiplied with itself $M_{t+1}^F = M_t^F \cdot M_t^F$ which enhances the flow to well-connected nodes, i.e., nodes within a community.

In the inflation phase, M^F is column-wise manipulated with an inflation operator $\Gamma_r(M_F)$ with $r \in \mathbb{R}^+$. For each entry M_{ij}^F we calculate $\Gamma_r(M_{ij}^F)$ with $\frac{M_{ij}^r}{\sum_{k=1}^N M_{kj}^r}$ or in other words we exponentiate the M_{ij} and normalize with the weight of neighbours. This increases the inequality in each column and reduces the flow across communities.

After each iteration, edges with a weight below a certain threshold are pruned out, leaving a sparse matrix including only edges with strong sinks. These islands are then interpreted as clusters or communities. Satuluri [125] introduced an improved version of Markov Clustering tackling the weakness Dongen's primary introduced algorithm that identifies many clusters which are too small.

2.3.3 Centralities

Centralities are indicators that assign importance to the nodes of a network. In general, the literature distinguishes between centralities that consider the number of specific paths originating from the node (radial) and those that take into account the number of paths that pass through a particular node [126] (medial). The goal of this section is to give a short overview of the essential centralities, namely, Degree Centrality, Closeness Centrality, Betweenness Centrality, and Page Rank.

Degree Centrality

We start with the easiest to understand and calculate [126] of the centralities presented here, the degree centrality. In computing degree centrality, the edges of a node are counted. As Wasserman [127] already stated, degree centrality is a radial metric since it considers the number of paths starting from a node.

Closeness Centrality

Closeness centrality, or simply closeness, was first introduced by Bavelas [128, 129] in the 1950s and describes how close a node v_i is to all other nodes V in a network. Thus, the more central a node is, the closer it is to all other nodes. The closeness of a node v_i is given by the function $C_c(v_i)$ and is the inverse of the average length of the shortest paths to all other nodes. In general, the normalized form of closeness centrality is used by multiplying the result of Bavela's closeness by the number of nodes $|V|$ without the considered node v_i. The latter allows for the comparisons between nodes in graphs with different sizes. Formally, we define

$$C_c(v_i) = \frac{1}{L_{v_i}} \text{ with } L_{v_i} = \frac{\sum\limits_{v_j \in V \setminus v_i} d(v_i, v_j)}{|V| - 1} \tag{2.6}$$

where $d(v_i, v_j)$ is the length of the shortest path between v_i and v_j.

Taking distances from or to all other nodes is irrelevant in undirected networks. In contrast, it can produce different results in directed graphs e.g., websites in the World Wide Web may have a high closeness from an outgoing connection and low closeness from incoming connections simultaneously.

Betweenness Centrality

Betweenness centrality, or simply betweenness, is a medial metric that quantifies how often a node v_i acts as a bridge between other nodes. Intuitively, nodes with a high betweenness centrality can also be understood as "bottlenecks" since many shortest paths pass through them. The betweenesss centrality was introduced by Freeman [130] in 1977 to give a measure for the influence an actor has on the communication of other actors in a social network. Nowadays, betweenness centrality $C_b(v_i)$ is usually defined as

$$C_b(v_i) = \sum_{s \neq v_i \neq t \in V} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{2.7}$$

where σ_{st} is the number of shortest paths between the nodes s and t, and $\sigma_{st}(v_i)$ is the number of shortest paths between s and t passing through the node (v_i). The most used algorithm is the one of Brandes [131] which works by first calculating all shortest paths from s to t, in a second step calculating the fraction of shortest paths going through v_i and, in a last step, summing up the fractions.

Page Rank

Page Rank is probably the most famous of the centralities presented here. Its existence is the basis for the success of the Internet giant Google. The English Wikipedia page on Page Rank[4] refers to Google's description of the algorithm as:

> "PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites."

Page Rank was introduced by Page et al. [132] in 1999 as an algorithm to determine the importance of web pages based on the important of their neighbours. The PageRank score for a node is aggregated from its neighbour PageRank-score. We define the PageRank for a given node v_i as

$$C_p(v_i) = c \cdot \sum_{(v'|v_i,v') \in E} \frac{R_{v'}}{d_{v'}}. \tag{2.8}$$

Here, E is the set of undirected edges while $\frac{R_{v'}}{d_{v'}}$ represents the neighbour v''s PageRank score. Calculating the PageRank for each node in the network means repeating this process iteratively until the PageRank scores converge.

2.4 Twitter

2.4.1 Twitter Volume

Twitter is one of the leading online social networks next to Facebook; it started its service in 2005 with 5,000 tweets a day and increased by magnitudes to 35 million tweets per day in 2010 to almost 500 million tweets per day in 2013 [133]. Twitter released these numbers in 2014. It can be assumed that these numbers have not changed much because Twitter's active user count has not increased since 2014. However, during record tweet events like TV airings, a twenty-fold increase was recorded with a peak of over 140 thousand tweets per second.

[4] Wikipedia PageRank: `https://bit.ly/2SYqirL`

2.4.2 Twitter API

On average, the entirety of Twitter users creates six thousand tweets each second [133]. To access this vast amount of data, Twitter offers the developer API to interact with its underlying services across multiple endpoints. Thus, every functionality offered on Twitter's website is accessible via its API, too. Nevertheless, Twitter does not allow researchers to retrieve its entire data at once. Enterprise customers have access to a so-called Decahose [134] stream, which includes 10% of all Twitter data in real-time. Furthermore, there is also a nonlisted option called *firehose*[5], which streams the total Twitter data in real-time. However, this option is not available to the general public or research. There are some approaches to reconstruct the firehose access by reverse engineering Twitters Snowflake IDs[6], but this is not the scope of this work as it violates Twitter's current terms of service.

Quotas

Despite Twitter's public open nature, it does not allow mass exporting of data outside its enterprise API[7]. For ordinary purposes like third-party Twitter apps, users can request an API token to act and view content on their behalf. The REST-API is rate limited under a 15 minute time window that resets the individual endpoints' contingents to a constant amount. Every URL path group has its rate limit, which we will call quota from now on. A URL path group is a matching regex path such as `/user/:id` where `:id` is a matching variable. In case that the quota is exceeded, the API returns an HTTP "429 Too Many Requests". A time frame is anchored to the time of the first request on a given path. Furthermore, a user-generated token has no global upper usage limit or rate limit, meaning that we can continuously use the quota on an isolated path for analysis and calculation.

Authentication

Each request to Twitter's API must contain the corresponding authentication credentials in the form of a tuple consisting of an API token and a secret token [8]. Here, Twitter uses OAuth 1.0a[9] to authenticate apps, acting as a user on behalf of a user. OAuth is an authentication delegation protocol that generates tokens that carry the

[5]Twitter Firehose: `https://bit.ly/3uYMPmn`
[6]Twitter snowflake id: `https://bit.ly/3yiTBW6`
[7]Twitter enterprise API: `https://bit.ly/3f1erln`
[8]Twitter authentication: `https://bit.ly/3hzRE1q`
[9]Oauth 1.0: `https://bit.ly/3uXqbe2`

ability to authenticate as a target without knowing any sensitive user data such as the password. Twitter requires this OAuth token as an HTTP header field in every request to the API.

API Responses

Every Twitter REST-API response returns additional meta data including the current path rate limit and other instrumentation related fields, like the maximum possible usage amount on the current path, currently remaining requests, and the epoch timestamp at which the used token quota gets reset. Here, a path refers to an API endpoint. All this metadata refers to the information associated with the developer account whose token was sent as part of the corresponding request. The fact that it is linked to one specific account is essential to understand the approach we are going to introduce in Chapter 3. It has turned out that it is possible to use more than one account simultaneously.

2.4.3 Twitter Data Processing Tools

The popularity of Twitter as a research platform has led to the development of many tools capable of capturing and analyzing Twitter data in the last decade. An overview of earlier tools is presented by Gaffney and Pushmann [135]. However, most of these systems [136, 137] use the Twitter Streaming API, which provides Twitter data in real-time. This has the significant disadvantage of requiring researchers to select their area of interest beforehand. Thus, it is essentially impossible to look at events or people which turn out to be significant in hindsight. Some larger companies use the *firehose*[10] bandwidth of Twitter's Streaming API, which allows them to retrieve and store the entire content posted on Twitter. However, doing so requires an immense infrastructure that is neither feasible nor affordable for most academic users. Moreover, firehouse is only available for industrial-level customers at extraordinary prices. Furthermore, some tools [138] are no longer compatible with Twitter's terms of service after they changed significantly in 2018 [139]. Finally, in 2018, the U.S. Library of Congress withdrew from its earlier policy of collecting the complete Twitter archive[11].

As the extraction of information on trends and opinions has significant commercial value, there is a sizable number of commercial tools and services for doing so. Unlike

[10]Twitter Firehose: `https://bit.ly/3uYMPmn`
[11]Library of Congress: `http://bit.ly/2RNvgCX`

research tools, these programs typically focus on ease of use for a well-established set of analytics [140]. Other research focused tools such as *IndexedHBase* [141] and *DMI-TCAT* [142] follow similar goals.

The original paper for *DMI-TCAT* provides a detailed description of the underlying motivations and assumptions made in the development of that project. Their key observations are that a) only access to social media data by independent researchers guarantees reproducible, independent science. b) the design choices made in such an analysis tool influence the science that is based on it and, as a consequence, c) the tool should stay as close as possible to the raw data rather than selecting aggregates.

Chapter 3

Data Acquisition

Contents

This chapter explains the data acquisition process, starting from analyzing data sources, through the design and implementation of a system for data acquisition, to the final process of data set collection. While the Global Database of Events, Language, and Tone (GDELT) allows unlimited access to its data, there are projects like Baumgartners Pushshift [143], collecting Reddit data in its entirety and on a regular basis. However, both Reddit and GDELT do not allow for investigating the spread of information on an individual to individual basis. Due to their intrinsic design principles, there are no traceable notions of user-to-user or news-agency-to-news-agency communication. Reddit's main entities are topics called subreddits with users subscribing to them, leaving interaction in the form of comments to a post as

the only viable option for tracing information diffusion. GDELT, on the other hand, is a service that acquires and processes news data and has no intrinsic concept of interaction at all, resulting in a set of automatically generated metadata that does not meet any criteria for tracking the dissemination of information. We discuss both the services and the possibilities of building networks (see Section 3.1.1) from the available data on which the spread of information can be tracked in detail.

In Section 3.1, we initially evaluate the three candidates for potential data sources. In this context, the aspect of graph building plays a crucial role. We conclude that Twitter is the most suitable option and consequently take a closer look at this remaining candidate by discussing its limitations and, with this, primarily, access restrictions for data retrieval.

Later, in Section 3.3, we propose the FACT framework to solve the problems arising from Twitter's limited accessibility. We motivate the core mechanisms, derive framework components from these, and present them. Finally, we give a glimpse into FACT's operation from late 2019 to mid-2020.

3.1 Data Source Evaluation

Being aware of the constantly growing online social network landscape, we have to make a pre-selection. Instagram's [144] and Facebook's [145] privacy concepts lead to obstacles in collecting and analyzing user-related data. In addition, the complexity associated with video analysis leads us to exclude platforms like TikTok and Youtube. Thus, we focus on Twitter, GDELT, and Reddit. We continue examining these three candidates based on the following criteria: *graph building*, *data quality*, and *data accessibility*.

Studying news spreading phenomena on a network basis implies accessing data sufficient to represent information spread. We argue that the quality of the conclusions drawn about information spreading depends on a network's granularity, i.e., the type of entity represented by a network's node and the associated edges. Therefore, we rank networks with respect to *graph building*. *Graph building* describes the extent to which the available data is suitable for obtaining network representations. We discuss the types of provided entities that reflect network properties and evaluate the networks built from them. In the discussion of data quality, we evaluate our data sources to determine whether, as in the case of GDELT, we deal with automatically generated metadata, whether users use plain names, or as in the case of Reddit, browser plugins manage multiple accounts to allow participation in exclusive Subreddits, or whether,

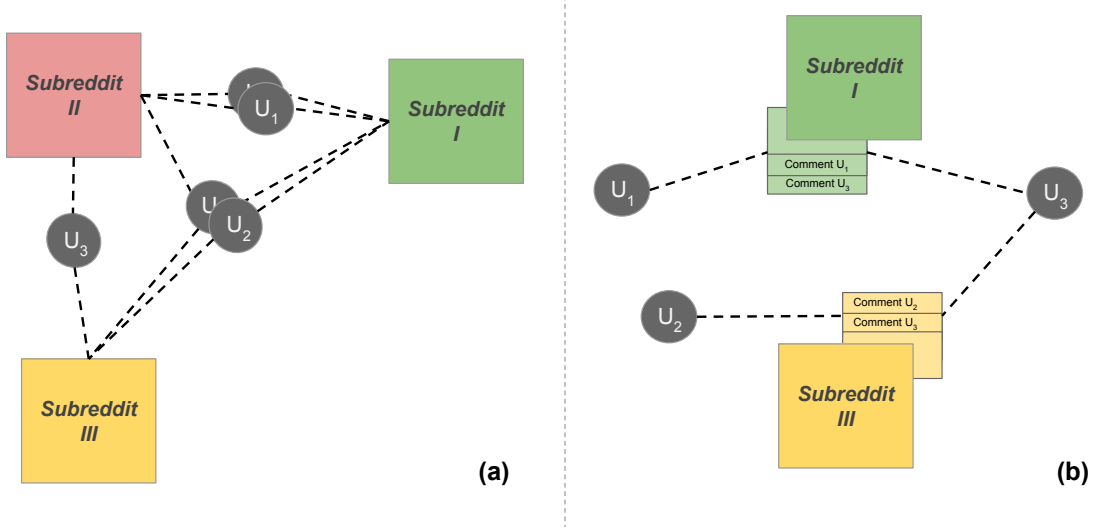

Figure 3-1: illustrates two ways to build networks from Reddit data. Above, based on subbreddits as data sources and below based on comments under posts.

for example, auto-translations falsify the data. Finally, we discuss *data accessibility* and investigate restrictions in the form of volume limits implemented by Twitter's API or privacy settings that prevent us from retrieving parts of the data at all.

3.1.1 Graph Building

We define graph building as the following problem.

Problem 1 *Given data from an online social network, how can we derive a graph representation on which the diffusion of information is traceable on an individual level?*

Reddit

By December 4, 2019[1] Reddit reported more than 130.000 active communities which are called Subreddits and which we further refer to as V^S. A Subreddit [146] is an openly accessible [147] forum-like community that operates as a discussion platform. Nevertheless, there are few private Subreddits [148] some of them even "quarantined", to hide them from new users [149].

Each Reddit user can become a member of such a community and therefore acquire the right to contribute with posts and comments on others' contributions. We refer to the set of Reddit users as V^U and indicate user affiliation to Subreddits with directed

[1]Reddit stats: `https://bit.ly/2S74Dgj`

edges

$$E^A \coloneqq \{(v_i^U, v_j^S) \mid v_i^U \in V^U \wedge v_j^S \in V^S \wedge i, j \in \mathbb{N}\}. \tag{3.1}$$

Here, V^S is the set of Subreddits.

A graph of type

$$G_0 \coloneqq (V^S \cup V^U, E^A) \tag{3.2}$$

seems to be the most obvious. However, G_0 is not more than a "universe" of star graphs and as such not a suitable solution to Problem 1. Here, the diameter

$$\delta \coloneqq max(s(x_1, x_2)) \text{ with } x_{1,2} \in V^S \cup V^U \tag{3.3}$$

and $s(x, y)$ referring to the shortest path, is always $\delta \leq 1$.

In Reddit, user interaction takes place only in a Subreddit's comment section. We continue discussing two different graph-building approaches (see Figure 3-1) that address G_0's $\delta \leq 1$ problem.

In the first case, we consider the number of Subreddits a pair of users participates in as an edge weight. We define:

$$G_1 \coloneqq (V^U, E^S), \tag{3.4a}$$

$$E^S \coloneqq \{(v_i^U, v_j^U, \psi(v_i^U, v_j^U)) \mid v_i^U, v_j^U \in V^U\}, \tag{3.4b}$$

with

$$\psi \colon V^U \times V^U \longrightarrow \mathbb{N}_0 \tag{3.5a}$$

$$\psi(v_i^U, v_j^U) \longrightarrow |\{v^S \in V^S \mid (v_i^U, v^S) \in E^S \wedge (v_j^U, v^S) \in E^A\}|. \tag{3.5b}$$

Even though this seems obvious, it appears to be more of a measure for similar interests than a for interaction. Note that no indicator points to the amount of content absorbed from v_x^S. Besides, there is no notion of user pairs interacting with each other, i.e., commenting on the same post.

In the second case, we assume two users interact with each other when they contribute to the same comment section. Defining a contact based on the mutual interactions in comment sections seems to be a more appropriate criterion for determining a contact than only the membership. However, just as the first approach, no guarantee exists that mutual interaction took place. We define the resulting graph as G_2. We point out that Reddit comments can follow a tree structure and that it is possible to make assumptions about users that can be assumed to have read

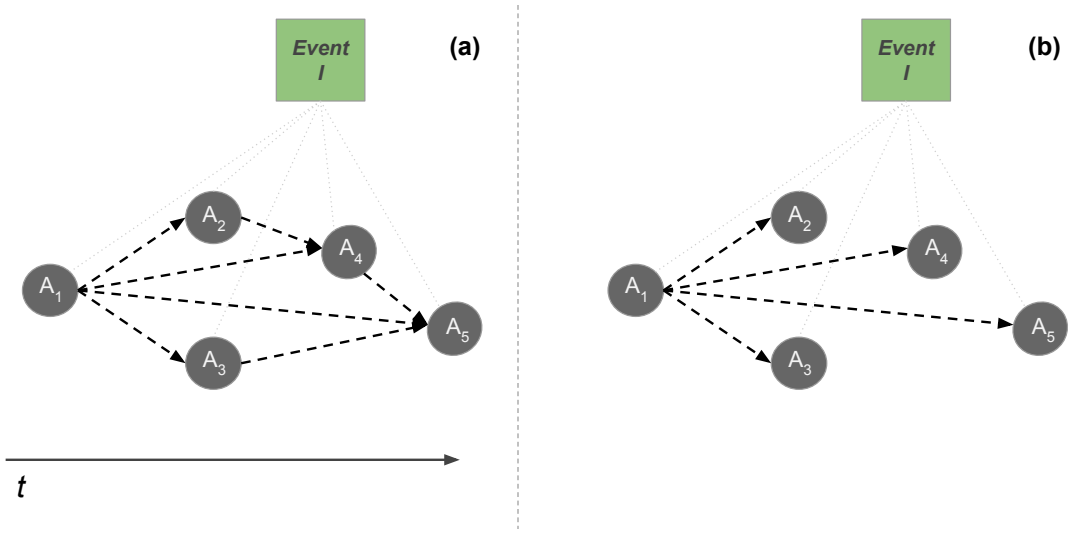

Figure 3-2: illustrates two ways to build networks from GDELT data. Left based on time properties and on the right hand side based on references.

comments. Even thought these assumptions are not verifiable and, therefore, nothing more than assumptions. It seems likely that the people commenting on a comment have probably not just read the comment they are commenting on but also the main post itself.

GDELT

The GDELT system obtains news data in 15-minute intervals and publishes the data as two tables, containing so-called *Mentions* and *Events*. While gathering, the system performs text processing to assign news articles to *Events*- and *Mentions* table. The *Mentions* table thus contains the URLs of the articles along with supplemental information.

As a service for creating and retrieving news metadata, GDELT [150] uses AI-techniques in order to determine which news article V^N belong to which event V^E. Similar to the proposed approach in Equation 3.2 GDELT's data forms a "universe" of stars. In this case

$$E^G := \{(v_i^N, v_j^E) \mid v_i^N \in V^N \wedge v_j^E \in V^E \wedge i, j \in \mathbb{N}\}. \tag{3.6}$$

An example of such an event would be the US Capitol attack on January 6th, 2021. Moreover, GDELT creates metadata, including the time of publication, sentiments and cross-references.

Like in Reddit, we identified two suitable graph-building approaches. Firstly,

referring to the references that news articles might have in between each other.

$$G_3 := (V^N, E^N) \tag{3.7a}$$
$$E^N := \{(v_i^N, v_j^N) \mid v_i^N, v_j^N \in V^N\} \tag{3.7b}$$

One can assume that one article's author must have read another article when his/her article refers to it. This reference alone makes an appropriate edge. However, indicating cross-referencing in the first place is not mandatory for the author of an article.

Second, a time based-approach. For each pair of articles v_i^N and v_j^N with

$$\exists v_x^E \mid v_x^E \in V^E \wedge (v_i^N, v_x^E), (v_j^N, v_x^E) \in E^N \tag{3.8}$$

and their timestamps $t^{v_i^N}$, $t^{v_j^N}$ indicating the publishing date. If $t^{v_i^N} < t^{v_j^N}$ we can assume a possibility that v_i^N's author might have been aware of v_j^N's content. Considering the sum of news articles news agencies published, it seems reasonable to build a network based on this property.

Both of these approaches have obvious weaknesses. In neither case, there is a certainty as to whether an author read its corresponding counterpart or got even influenced by it. It is nevertheless conceivable that in the future, advanced AI will redefine the playing field and approaches that, for example, identify the writing style of an author and thus makes tracking possible with an acceptable probability of success. GDELT version 3, which is scheduled to be released in the near future, is expected to offer improvements in this technology.

Twitter

Compared to graph building on GDELT and Reddit, there is a distinct advantage to Twitter. Twitter operates on the granularity of individual users V^F following each other and subscribing to each other's content. A user $v_i \in V^F$ can follow any other user and receive updates about that user's recently authored or shared content. These decisions naturally form a network

$$G_4 := (V^F, E^F), \tag{3.9a}$$
$$E^F := \{(v_i^F, v_j^F) \mid v_i^F, v_j^F \in V^F \wedge i \neq j\}, \tag{3.9b}$$
$$(v_i^F, v_j^F) := v_i^F\text{'s decision to subscribe to } v_j^F\text{'s content.} \tag{3.9c}$$

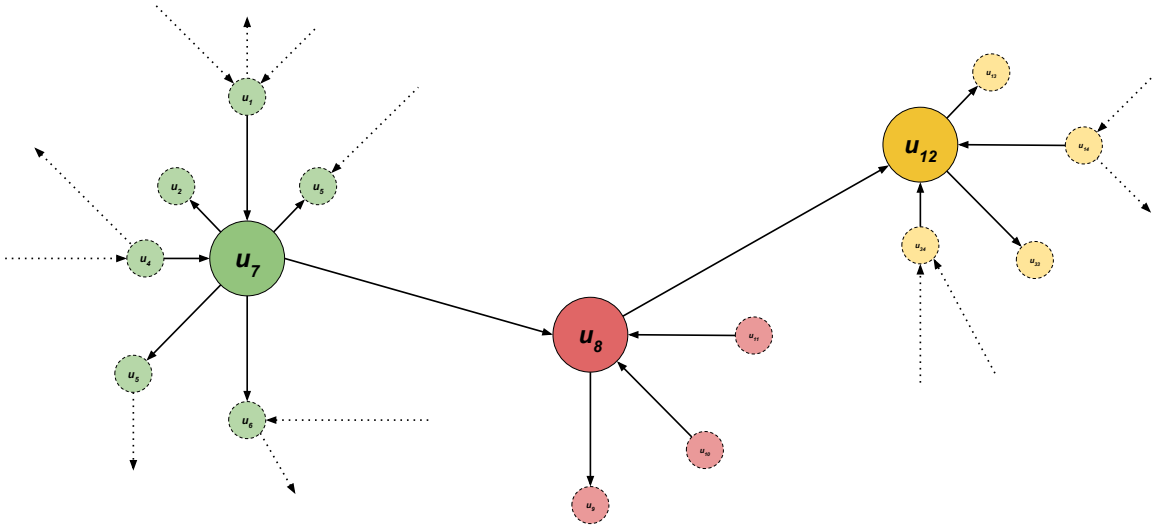

Figure 3-3:

It is essential to notice that follower networks are based only on users' decisions to follow another but provide little information about how they interact. Although content is displayed to those who follow, it depends on a follower's activity whether he or she inspects it. In addition, Twitter implements a recommender system that displays content in a particular order based on metrics such as view frequency or the number of likes.

Furthermore, there is the possibility of sharing and commenting what others posted. In addition, content and connections between users are publicly visible and accessible at a large scale. The fact that an individual is an entity with a timeline, i.e., a kind of accessible activity history, enables graph building on two levels: the most obvious, G_4 itself, and, moreover, network building based on interactions.

Each user v_i^F can share another user's content. We call this process retweeting. Because both retweets and tweets are retrievable, it would be conceivable to derive network representations based, for example, on the frequency of the retweets exchanged. The advantage of these types of network representations is that it is certain that an interaction took place. We introduce a dedicated procedure to build interaction networks in Chapter 4.

3.1.2 Data Accessibility

Due to its open API and the presence of substantial historic datasets [143] Reddit data is retrievable without significant effort. The Rest API surrounds a multitude of

established libraries [151] such as praw[2].

Since aggregating and storing Reddit in its entirety seems complex and expensive, the Pushshift project appears to be a valuable alternative. Pushshift aims at obtaining and exposing Reddit data through its custom API. Compared to Reddit's API, Pushshift's API allows querying historical data efficiently. There are three API endpoints to obtain information about subreddits, submissions, and comments. Pushshift's API provides extensive functionality to filter and search based on the content and to restrict the result ranges by ID or date. Because Reddit status IDs are sequential, this is a simple way to query, e.g., missing comment ranges. Supplementary to its API, Pushshift provides Reddit data dumps periodically[3]. Their compressed size is around 200 GB for all the submissions and 654 GB for comments from June 2005 until September 2019.

Pushshift dumps are, compared to API queries, easy to obtain and can be processed more rapidly, since it is possible to process them on hardware. Unfortunately, Pushshift dataset releases are occasionally delayed for months. Moreover, Pushshift dumps do not provide time-series information; thus, there is only one consistent snapshot at crawling time[4].

According to its website, GDELT [150]

> *"monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages and identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day, creating a free open platform for computing on the entire world."*

GDELT's version 2.0 monitors news sources, with archives going back to 2015 with articles in 65 languages auto-translated in real-time, making GDELT the system with the broadest reach that is publicly available [32].

While GDELT metadata is available for public download and thus is entirely accessible, its sheer quantity leads to challenges in data processing.

Twitter, like many other OSNs, offers a publicly accessible REST API. This API is protected through a user login and rate limits that define the allowed number of interactions with the service per API user. Such rate limits are commonly defined as a bucket over a sliding time window, starting with the first API call. This limits

[2]PRAW: The Python Reddit API Wrapper `https://bit.ly/2QtDUdF`.
[3]Pushshift Datasets `https://bit.ly/33QDvEX`.
[4]Pushshift dumps: `https://bit.ly/3eTyEcu`

access to a fixed amount, where each bucket gets reset after a given time period. The most commonly used time span for Twitter is a 15-minute frame combined with an optional daily time window.

3.1.3 Data Quality

Data quality states in which quantity, for which period of time and from whom data is available.

GDELT, despite its openness and ability to retrieve, large amounts of data, offers only aggregated metadata. News events are automatically assigned to one of 300 categories and stored as metadata on a 15 minute basis. Each entry has a global id as well as a timestamp indicating when it was collected for the first time. Moreover, GDELT aims to automatically extract what they call actors, i.e., persons of interest that play a significant role in the according article. All articles are summarized to so-called events. Even thought GDELT's approach seems to be promising, there are significant disadvantages regarding the results of the automated metadata extraction. During the period working with GDELT we have stumbled across a variety of articles including wrong metadata. Furthermore, GDELT does not give any access to the content of the articles, which makes manual downloads the only option for access. This in turn requires parsing a large number of media sources and penetration of paywalls.

3.1.4 Result

After evaluating Twitter, GDELT and Reddit based on *graph building*, *data quality*, and *data accessibility*, we conclude Twitter to be the most appropriate choice for studying news spreading phenomena.

The decisive factor for this choice is Twitter's individual-based nature, leading to information spreading traceable at a user-to-user granularity. Besides, there are concepts suitable to model interaction and influence. Twitter is also reasonably large and represents a cross-section of society. Many celebrities and institutions maintain a Twitter account and thus contribute to our analysis. However, these benefits come at the price of access restrictions. Unlike GDELT, which is built around the idea to make data available, Twitter allows access to its data only via its API.

3.2 Building the Haystack to find the needle

Despite the aforementioned advantages of using Twitter data to examine news spreading phenomena (see Section 3.1.4), data acquisition has significant limitations.

First off, there is the issue that Twitter, or rather its API, is not designed for querying news dissemination phenomena as a whole. Twitter does not provide labeled data with potentially harmful content, and, at the moment, Twitter's search API only allows for querying content that is up to two weeks old[5]. Although there is the opportunity to buy historical content, this is only affordable for small amounts of data. Therefore, in this thesis we do not rely on commercially acquired data.

We state that the two-week window is a reversible limitation that results in searching for the proverbial needle in a haystack, with no certainty that this needle exists at all. Keeping the haystack, i.e., the amount of collected data as large as possible, seems to be the only reasonable strategy for maximizing the chance to observe digital wildfires in their entirety.

An additional motivation for such a large-scale data collection is examining the comprehensive data related to digital wildfires. We argue that studying the context of a phenomenon leads to new insights that allow us to better understand the phenomenon itself. Here, the context includes tweets, retweets, and user networks that directly relate to the phenomenon, those that lead to the phenomenon as well as those that result from the phenomenon. Assuming the data collection is comprehensive enough to include at least one "needle in the hay" and its surroundings, we face the challenging task of identifying suspicious data. To uncover what to search for requires careful reading of so-called "alternative media" and content circulating in conspiratorial circles. If we finally find ourselves in this situation and have become aware of potential candidates, as in the example of the 5G Corona conspiracy (see Chapter 5) we are facing additional challenges, namely, the corresponding network structure that allows for tracking the spread of the "needle" (see Section 3.1.1) must be fetched, maintained and kept up to date. These network structures are constantly changing due to newly appearing friend and follower relations and the emergence and elimination of accounts. Considering the API's access restrictions, it seems unattainable to capture a consistent image of Twitter's state at an arbitrary point in time. Furthermore, Twitter's API design forces us to outsource this process to a separate procedure that we will call Follower Network Job (see Section 3.3.4). In summary, the "hunt" for news spreading phenomena and the need to study their distribution on

[5]Search API: `https://bit.ly/3dOuz9a`

the social networks underlying Twitter presents us with the following challenges:

1. Twitter API Restriction.
2. Constantly changing network structures.
3. Store / Maintain large data collections.
4. Changing data (likes, retweet counts, ...).
5. Processing of huge amounts of data (limited in time).
6. Combination of multiple collection strategies.

To address these challenges, we introduce the FACT framework. FACT's design and development started in early-2019 and was the topic of three publications. The first publication [33] includes the overall concept and architecture. In contrast, the second publication [152] proposes a mechanism to extend the data restrictions assigned by Twitter using a crowd-based approach. Finally, the third contribution [34] introduces the algorithms and data structures necessary to optimize news spreading oriented data collection. In the following, we give an insight into the main concepts, the structure, the hardware used, and the algorithms. Besides the mentioned conference papers, Andreas Huber [153] and Haseeb Rana [154] have developed parts of the system as part of their Master's theses. These theses were written in the context of this dissertation and are listed individually in Section 1.2. The main purpose of FACT is to collect, store, and process large amounts of data from the online social network Twitter and and thus make Twitter a suitable data source for the investigation of large-scale news spreading phenomena.

We developed the data collection process with two main objectives. The first objective is to collect what we call "dense user data" on a large scale. Dense user data aims to be as coherent as possible, i.e., we are interested in data published by users that are "close" to each other. Here, close refers, on the one hand, to the number of hops in Twitter's follower network and, on the other hand, and the existence and frequency of interactions, i.e., the number of retweets or comments. The second main goal of the data collection is a form of discovery that collects tweets and retweets based on topics independent of the underlying network structure. The idea is to capture subsets of news complexes that are potential sources for news spreading phenomena. Here, network density plays a subordinate role. An example of this is the US elections, COVID-19 or Black lives Matter. Maintaining and servicing a

distributed environment with these requirements is costly and labor-intensive. Even though developed in a research project, FACT is a production system, which has been running and continuously improving over more than five months, while being distributed over two research data centers and a cloud provider.

The FACT framework consists of six different components (see Figure 3-4). These components are distributed over five different systems: The Wally Cluster of the Technical University of Berlin, the Simula's eX^3 infrastructure, the Simula Lizhi Server, several AWS EC2 instances, and Amazon's Elastic Beanstalk. The main requirement was to provide inexpensive storage, which we could use for an extended period. This was especially true for the Wally cluster and the Lizhi server. In the following, we will go over the FACT framework's components and try to explain the purpose of each of them in detail.

3.3 FACT: A Framework for capturing Twitter Data

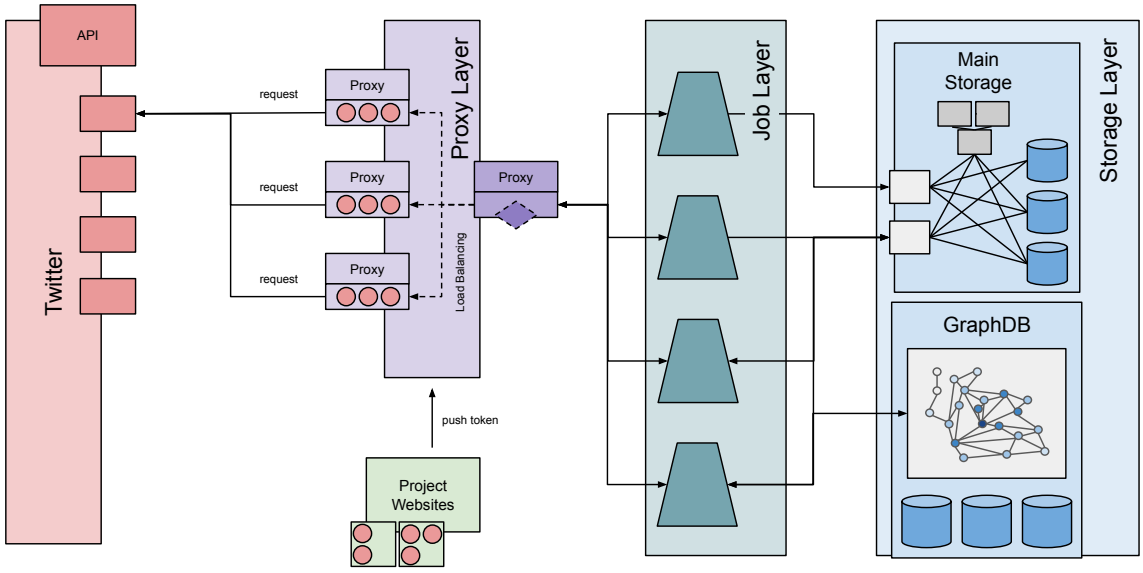

Figure 3-4: The figure illustrates the overall architecture of the FACT framework including the six main components. A basic distinction is made between the proxy layer, crowdsourcing, scraping jobs, the storage engine and the data processing pipelines.

Probably the most crucial component of the FACT framework and, at the same time, the part that has driven the development to its current state is the proxy layer which presents the gateway from FACT to the Twitter API. Based on the idea that it is possible to donate API quota from individual Twitter users and thus increase

the number of requests to Twitter's API above the standard level, we first designed a simple token balancer. Initially, the token balancer was developed as a single component (not distributed) and tightly coupled with Twitter4j[6], a Twitter client for Java. Because of the lack of scalability and the need to run multiple clients simultaneously, which would have to share the tokens, we decided to outsource the token balancing to a dedicated layer (see Section 3.3.2 & Figure 3-4). In order to understand the whole system, it is crucial to comprehend proxies as mirrors to Twitter's API running inside our hermetically sealed system and making the Twitter data restrictions transparent. The proxy layer connects to the project website, which offers the possibility to donate API tokens. Donated tokens are retrieved at regular time intervals and automatically integrated into the proxies so that the newly gained quota is immediately available (see Figure 3-6). Proxies are the gateway to the Twitter API and the only components that talk directly to it. All other components are hidden behind a Virtual Private Network (VPN). Furthermore, proxies are interconnected hierarchically in a tree structure so that only the leaves talk to the API. The roots of this proxy structure run on the same nodes as the jobs. The nodes on which the corresponding jobs run request the data locally via their dedicated access proxy and then process them to store them in the local Mongo Gateway [155] or the Neo4j database [156]. The main link connecting the proxy layer and the storage layer is the job. Jobs are Java programs executed over an extended period of time and described in more detail in Section 3.3.4. Each job operates on a dedicated machine that runs a local Mongo gateway and a local proxy, respectively. Since we usually deal with continuous data streams, scraping jobs are programmed using the Java Reactive Stream Framework, allowing us to work on an unpredictably large data flow.

3.3.1 Crowd-Based Approach

To increase the total amount of Twitter data that can be collected, we use a crowd-based approach. The UMOD project website contains a donation button (see Figure 3-5). If the button is clicked by a Twitter user, this user transfers his or her data contingent to the UMOD app in the form of an OAuth token. Currently, more than 90 users support the project. This means that the amount of data that can be collected is ninety times higher then the single user quota.

[6]Twitter4j: `https://bit.ly/3qcWmod`

Figure 3-5: This figure shows both the English and German versions of the project website. Clearly visible in green or red is the Donate button with the help of which any Twitter user can donate their quota.

3.3.2 Proxy Layer

Twitter permits access to its developer API after successfully applying for a Twitter developer account. Applications must include a detailed explanation and reasons justifying the access to Twitter's data while promising to handle the data in a responsible manner. After the access is approved, developers can create up to one hundred Twitter Developer Apps[7]. Access to Twitter's API requires authentication in the form of a token consisting of an API key and a secret. This token must accompany each request. For authentication, Twitter uses either OAuth 2.0 Bearer Token[8] or OAuth 1.0a with key and secret. The entity generating these tokens is the Developer App (APP). An APP aims to organize projects by encapsulation. Moreover, each APP offers the ability to issue authenticated requests on behalf of the APP itself, as opposed to on behalf of regular Twitter users. The latter uses the 3-legged OAuth flow, which allows users to authorize an APP and thus provides the app developer with a token to use on behalf of the user.

[7]Twitter Developer App: `https://bit.ly/3wKux97`
[8]Bearer Token: `https://bit.ly/3qcybGm`

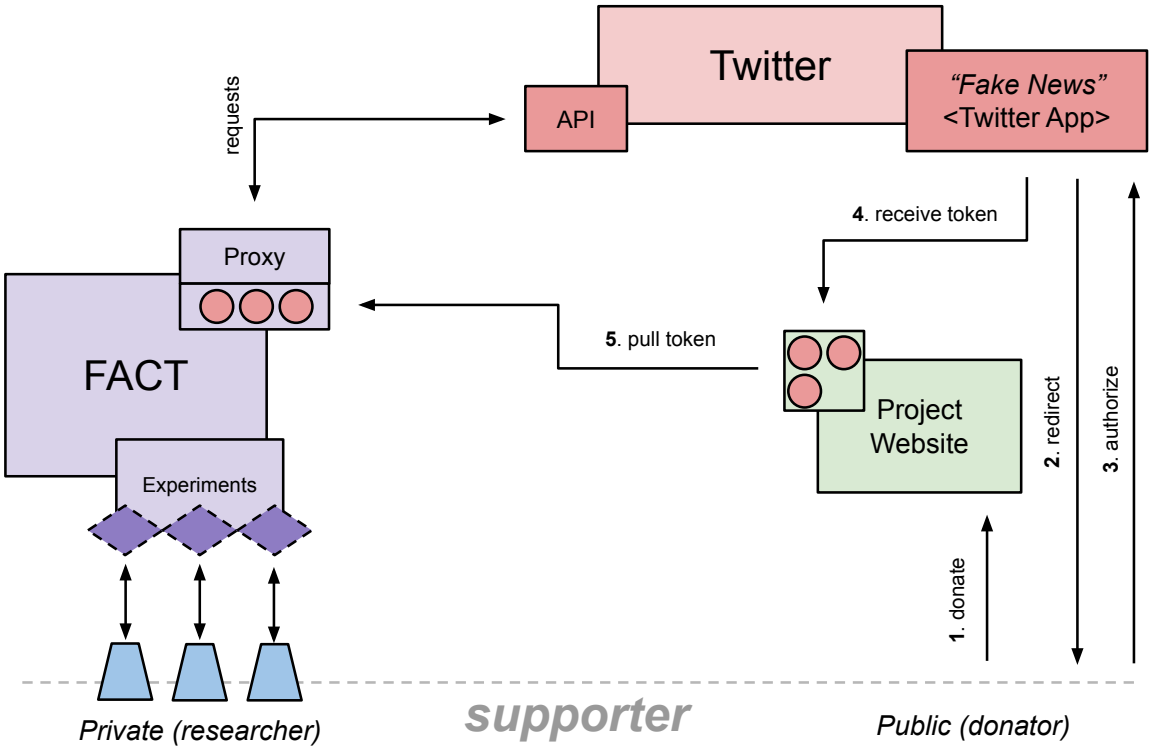

Figure 3-6: The diagram of interaction within the system's environment including Twitter, the project website, and the participants. Steps from one to five describe the process of donating a token. After pressing the donate button on our project website (1), an OAuth authorization is initiated. The donor is first forwarded to Twitter (2) where they authorize our "Fake News App" with read-only authorization (3). Then our web server obtains an OAuth-token (4) which is finally collected by the proxy (5).

Twitter's API offers several endpoints, each of which provides access to a specific kind of data. For example, there is one endpoint for retrieving user data: tweets, retweets, followers, etc. Twitter restricts access to its API by assigning each access token a quota for each of these endpoints, maintaining the remaining quotas in server-side logs, and replenishing them in 15-minute intervals.

A layer of API proxies outsources request handling and load balances donated tokens while offering a mirrored API to internal services. Thus, the proxy layer makes Twitter's access restriction transparent. Each proxy server manages a contingent of access tokens donated previously through the website.

In case a Twitter user decides to donate quota, he or she is redirected to Twitter and requested to agree to his or her membership to the APP. Twitter then generates the token (key, id) and forwards it to the project website, where we keep it in an encrypted store until a CRON job from within the proxy layer fetches and decrypts

the tokens on a 15 minute basis.

Depending on the load and the number of donated tokens, additional proxy servers can be added. Scaling is possible horizontally and vertically across multiple project websites, developer accounts, and donor groups. In the latter case, a proxy server does not act as an API gateway but as a gateway to another proxy server layer. Thus, the proxy on the first level acts as a load balancer for the proxies on the subsequent layers.

Overall, the entire proxy layer provides scalable internal access to the Twitter's API, making Twitter access restrictions transparent. Furthermore, the proxy layer is directly connected to our crowd-sourcing website and can access the newly donated access token within a 15 min time limit.

3.3.3 Persistence Layer

We distinguish between three different types of data that we want to store:

1. Network data, which is data that contains the follower and friend connections between Twitter users.

2. User data is data belonging to a Twitter account, such as username, id, location, and the timestamp of the account creation.

3. Statuses are the content generated by a user. Statuses include tweets, retweets, quotes, and replies.

For each of these data categories, there are different methods of preservation. This section discusses the technologies and structure of this data and further describes how it fits into the big picture of the FACT framework. We start with the network data and continue with the statuses and the user data.

Network Storage

The Follower Job (see Section 3.3.4) describes how we visit Twitter accounts, update the followers and friends, and pass the results to the persistence layer. The need for persisting graph structured data in the expectation of simultaneous updates and writes suggests using graph databases.

We considered two options and evaluated Neo4j [156–158] and Arrango [159] with a final decision for Neo4j. This decision is based mainly on usability aspects. Neo4j offers a web interface with a graphical explorer and the query language called "Cypher",

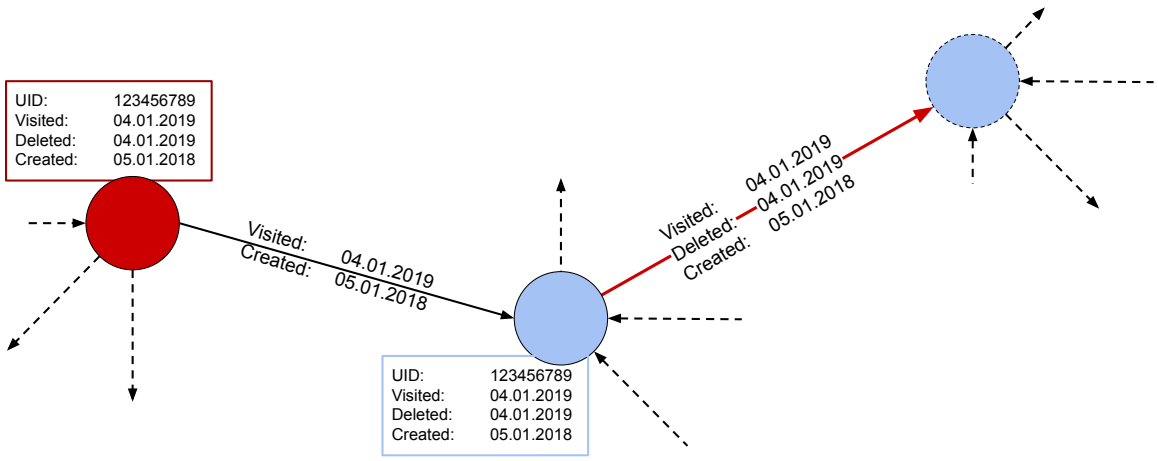

Figure 3-7: Follower network data model

which we considered very intuitive and easy to learn. Besides, excellent documentation is available. Furthermore, Neo4j can handle graphs containing over twenty billion edges and one billion nodes.

Figure 3-7 illustrates the graph structure persisted in Neo4j. It is important to note that the graph we maintain within our graph database is continuously increasing. We do not delete nodes, but instead, we label them as deleted. Each node contains only the Twitter user Id, a timestamp for the "last visited" date, and a label that indicates whether the node has been deleted or is no longer available. We do not distinguish the reason why the node is no longer available. Users that close down their account temporarily, delete it, or users that got suspended due to violations of Twitter's Terms of service are all labeled as deleted.

Even though Twitter internally stores new followers and friends according to the time they have been added, we collect the entire friend or follower list and compare it to our internal representation. This procedure not only discovers new followers or friends but at the same time allows for detecting missing connections. Thus, we can log if a friend or followership has been deleted between two visits. Each edge also contains the date of the last update and a label that indicates whether the edge still exists or has been deleted.

User and Status Storage

We store both the user data and tweets in a sharded MongoDB [155]. Sharding [160] refers to the praxis of distributing data across multiple machines and is, in fact, a form

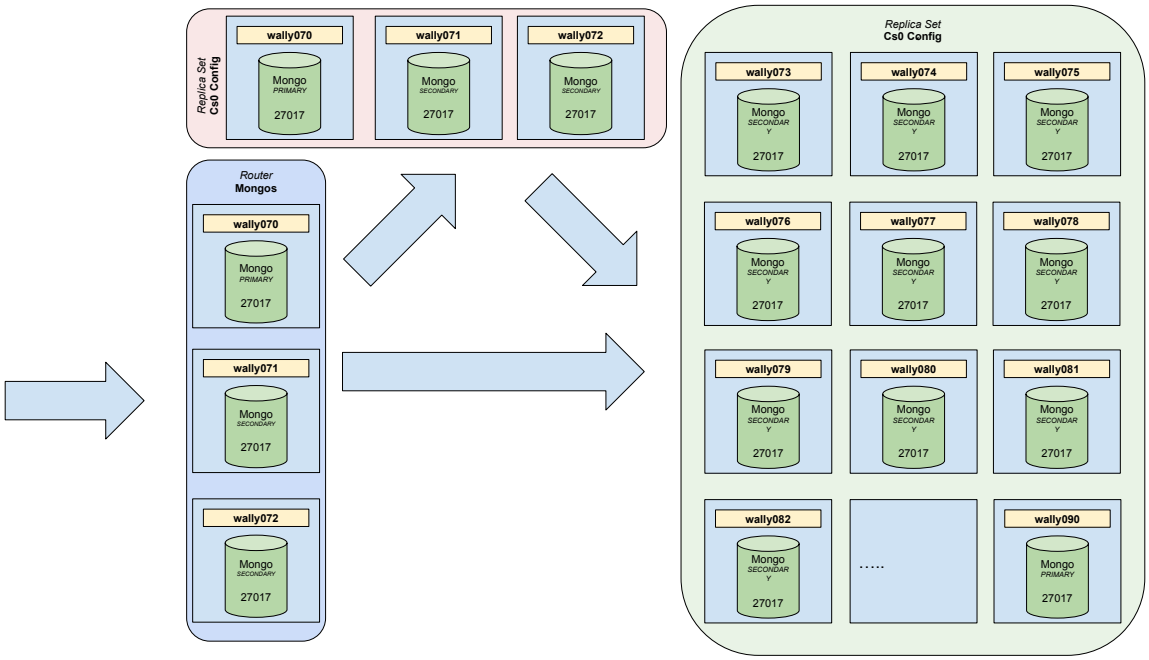

Figure 3-8: Distributed Sharded Mongo Cluster with twenty shards, configuration replica set and two mongos gateways. Each node is specified according to Table 3.1.

of horizontal scaling. Thus, we can divide datasets and workload over multiple servers and, on the fly, add resources in case of capacity constraints. MongoDB states in its documentation: While a single machine's overall speed or capacity may not be high, each machine handles a subset of the overall workload, potentially providing better efficiency than a single high-speed, high-capacity server. This promise seems to fit precisely our requirements - the FACT framework needs to collect data over extended periods, and most high-performance resources are only available for a limited period of time. Table 3.1 shows the characteristics of the available storage nodes and Figure 3-8 depicts our distributed sharded MongoDB Cluster including twenty shard nodes. We maintain two distributed collections, one for the user profile information and one for Tweets. Both collections are sharded according to their identifier.

Table 3.1: Resource of the Wally cluster at the TU Berlin

Resource	Details
OS Ubuntu	18.04.3
CPU	Quadcore Intel Xeon CPU E3-1230 V2 3.30GHz
Memory	16 GB RAM
Storage	3TB RAID0 (3x1TB disks, linux software RAID)
Network	1 GBit Ethernet NIC

User data describes user profiles. We aim to store this type of data in a separate Mongo collection[9]. A Snowflake Id[10], generated based on the creation date and Twitter's internal server configurations, uniquely identifies each user profile. Accordingly with Twitter's specification[11], we have kept a similar data structure with the following changes.

Each user profile contains a number of followers and friends; friends are followed by the owner of the profile while followers follow the corresponding. Understanding user data gathering as a process in which we visit user profiles at irregular intervals implies that these values might have changed between two visits. Representing these fields as lists in which we append new values for each visit if they change allows observing how the corresponding profiles change over time. The price for this powerful feature is the dynamic size. The Mongo database used in our case needs noticeably longer updates than in the case of a static, listless representation. Other changes include removing a handful of nonessential fields and replacing the last status update, which is also part of the profile, with a reference pointing to the tweet collection. Additionally, we indicate when a profile no longer exists, i.e., has been deleted or banned, and the time of the last visit.

To update user profiles regularly, we must receive the existing profiles ordered by the last visited timestamp. Only in this manner is it possible to update the profile visited longest in the past first and thus keep the entire collection consistent. The data in the user collection is therefore indexed as follows. We created a compound index[12] for the last visited/id field and a single field index[13] for the id field. Mongo maintains a memory hash table for each indexed field and thus ensures accessing indexed data with $\mathcal{O}(1)$. The compound index also allows retrieving the profile data in a stream, sorted by the last visit.

By status, we refer to a tweet, a quote, a retweet, or a reply. So we treat comments, tweets, and shares together. Again, as with the user objects, our internal data structure does not differ significantly from the Twitter API returns structure. Similarly, we indicate the point in time when the status is collected by adding the *last visited* field. Moreover, we replace the user object returned inside of every tweet with the corresponding snowflake id in the user collection.

[9]Mongo Collection: `https://bit.ly/3vCLWj1`
[10]Twitter Snowflake: `https://bit.ly/3370QSI`.
[11]Twitter User Object: `https://bit.ly/3vr12Iy`
[12]Compound Index `https://bit.ly/3uaM6hs`
[13]Single Field Index `https://bit.ly/2SaPGKg`

3.3.4 Job Layer

As mentioned in Section 3.2, there is no single API endpoint providing the entirety of data required to monitor the distribution of individual tweets through Twitter's underlying social network. Instead, it is necessary to access four different API endpoints and combine their results. First, the tweets themselves; second, the retweets related to a given tweet; third, the underlying social network (follower relationships) and last, the user's account information.

Jobs are the links between the persistent and the proxy layer. Here, the acquisition strategy, i.e., the algorithm that determines the scraping strategy, is implemented. Jobs are independent Java programs which we schedule on one node with a MongoDB client that acts like an actual instance. MongoDB makes use of these clients called *Mongos* to communicate with the configuration replica set, for caching and to talk to the actual shards that include the actual data (see Figure 3-8). Our data mining includes collecting and updating user profiles, tweets, retweets, and network data. Therefore, we query four API endpoints, thus creating the following four jobs:

1. Search API Job,
2. User Job,
3. Follower Network,
4. Timeline Job.

In the following, we present these four jobs. Because there is a dependency between profile and timeline jobs, we explain the timeline job last. Moreover, there is a discovery process with the help of which we can determine the set of users to be examined.

In regard to Twitter data mining, we face a problem that we will refer to as sampling under strict vs. ordinary cost constraints. Twitter is dynamic, and thus its data is subject to change. We define this change as either addition, deletion, or modification. The latter is assumed to be an infrequent event. The speed with which we attain the sampled data is limited by different factors, like access budgets or network throughput, making it challenging to measure the performance of a sampling algorithm because the change in the data can be greater than the sampling speed. Thus, we need to distinguish and find other ways of measuring the performance of our algorithm; when modeling the graph as an Monte Carlo Markov Chain, traditionally, the Gwecke Z Score [161] is used to determine the convergence of the underlying

Markov chain, but in a dynamic network, this convergence is not trivially definable. Therefore, we distinguish two cases:

Definition 3.3.1 (Sampling with ordinary cost constraints) *Given an online social network, creating new content happens slower than a sampling algorithm can track, obtain, and store it.*

Definition 3.3.2 (Sampling with strict cost constraints) *Given an online social network, creating new content happens faster than a sampling algorithm can track, obtain, and store it.*

We model Twitter's underlying social network as a directed graph

$$G^A = (V^A, E^A), \tag{3.10}$$

where V is the set of users and E the set of irreflexive follower relationships with

$$E^A \subseteq \{(v_i, v_j) \mid v_i, v_j \in V^A\}. \tag{3.11}$$

The first obvious step to make efficient use of quotas includes refining the observed population. Thus, we consider only the subgraph G of G^A with

$$G \coloneqq (V, E) \text{ with } V \subset V^A \wedge E \subset E^A. \tag{3.12}$$

Here, $v \in V$ is associated with a vector of user profile properties, including the most recent message, a message counter, and the profile language. The latter we consider a valuable filter criterion to reduce $|V|$. We suppose that users communicating in the same language tend to be connected and thus decrease $|V|$ with

$$V \coloneqq \{v \in V^A \mid \mathit{language}(v) = \text{german}\}. \tag{3.13}$$

Search Job

Twitter's search API[14] enables queries for sentence fragments, hashtags, or account information, including indicators for time intervals. Search queries cannot obtain tweets older than one week[15]. Additionally, there is no guarantee for returning the entire set of tweets matching the query. The same also applies in particular to retweets

[14]Twitter Search API: `https://bit.ly/3atm6WK`
[15]Twitter Search API Time Limit: `https://bit.ly/2Rgu92m`

which are to a high degree inconsistent. Both inconsistency and time constraints are limiting factors making the investigation of information dissemination difficult.

The task of the search API job is to collect tweets and retweets that contain terms related to specific news spreading phenomena. In the course of searching for these phenomena, we have looked into series of different search terms on topics such as COVID-19, black lives matter, etc. The goal here is not to capture whole phenomena or coherent data. Rather, topic-related data should be collected, which will be evaluated afterwards.

User Job

The User Job aims to update and extend V (see Equation 3.13). Therefore, we must distinguish between updates and discovery. Twitter's API endpoint for querying user profiles allows to query a total of 90,000 user profiles per user token for a 15 min window. The declared goal is to visit all $v \in V$ in regular time intervals, and to update their information. Because each user profile[16] includes the user's the most recently authored tweet in its entirety, updating a user's profile also extends the user's set of tweets.

Follower Network Job

This job connects to Twitter's follower and friend endpoint that receives a user id and returns pages of the respective follower or friend lists. The goal is to memorize large parts of the Twitter follower network and keep it as up-to-date as possible. The follower job receives a sorted stream of user ids starting with the users that were visited last. For each user id, the job checks whether the user still exists and then queries its entire friend and follower list to merge it with the graph representation in the FACT-Storage layer, i.e., the Neo4j database.

Timeline Job

In Section 3.3.4, we describe how to obtain the underlying social network using what we call the Follower Network Job, and in Section 3.3.4, we describe the implemented measurements allowing us to gain the corresponding user's account information, leaving us with the remaining parts, namely, the tweets and retweets. There are two different ways to obtain tweets and retweets either by collecting tweets on the user timelines or using the search API.

[16]User Object: `https://bit.ly/3zPvpue`

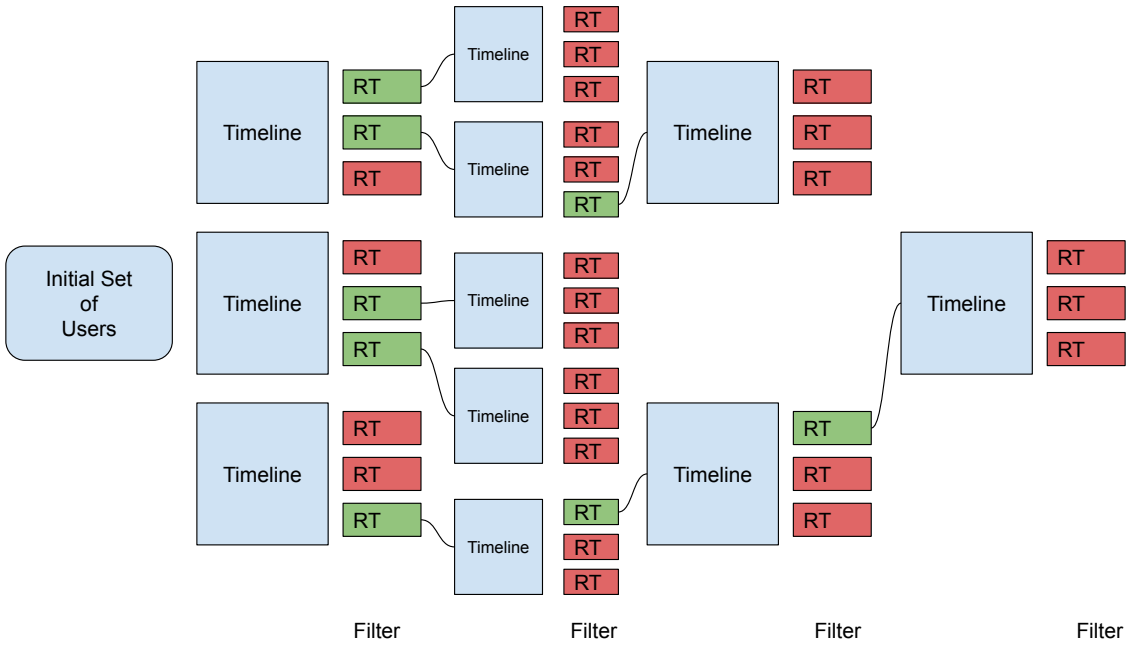

Figure 3-9: shows the iterative discovery process for finding users with a particular language property. Although Twitter's user profiles have a dedicated field for the language, this is not a sufficient criterion for assigning the profile to a specific language group. For example, it has been shown that older German profiles often use English as the profile language because in the early days of Twitter there was no option to select German as the profile language. Also, many international users are familiar with the English language. The discovery process shown here works like this. Starting with a small set of selected German users from Twitter lists[17] for politicians from parties across the spectrum (see Table 3), we started collecting the timelines up to the limit of 2000 tweets/retweets. These statuses were then filtered for retweets, with non-German retweets filtered out. The authors of all original tweets belonging to the retweet were then stored in our user collection. This process was repeated until the number of new users did not increase anymore.

Since Twitter's search API is time restricted and returns inconsistent data (see Section 3.3.4), the only remaining way to obtain data that is consistent or older than two weeks and thus meet the requirement of examining news spreading phenomena in their entirety is to walk Twitter's social graph and collect timelines. According to David Sayce [162] Twitter has approximately ($|V| \approx$) 330 million active users monthly, tweeting about 500 million tweets per day. Theoretically, it is possible to collect about one million tweets per day using just a single developer account because of the 900 requests per 15 minutes limit with up to 200 tweets per request when speaking to the standard tweet endpoint of Twitter's API. However, this is not the whole picture, as retweets, replies and quoted tweets remain unaccounted for resulting in data quantities

that are orders of magnitude higher. Moreover, the standard tweet endpoint expects a list of tweet ids that must have been discovered beforehand. According to Han [163], an average tweet is retweeted 17.75 times. Furthermore, we can't assume that API responses contain the maximum of 200 tweets, as most of the existing 1.3 billion accounts are inactive with 44% of new users leaving Twitter before publishing a single tweet[18]. The tweet distribution per user follows a power-law, so collecting exactly 200 tweets with one request is rather an exception than the rule. We define h as a function to compare the performance of the sampling algorithms with strict cost constraints (see Definition 3.3.2). To measure the utilization of the request budget, we describe the ratio of collected messages and consumed requests as a function

$$h(collected, budget) = collected/budget. \tag{3.14}$$

Notice h does not consider the age of the message.

We further assume that each user $v \in V$ can publish messages $p_i^{\langle v \rangle}$ where i denotes v's ith message. Here, a message is either a tweet, retweet, quote, or reply. Moreover, we refer to the entire set of v's messages as

$$\mathcal{M}^{\langle v \rangle} \coloneqq \{(p_1^{\langle v \rangle}, t_1^{\langle v \rangle}), (p_2^{\langle v \rangle}, t_2^{\langle v \rangle}), ..., (p_n^{\langle v \rangle}, t_n^{\langle v \rangle})\}, \tag{3.15}$$

where $t_i^{\langle v \rangle}$ is a timestamp. To simplify the notation, we continue using either t or p instead of $(p_n^{\langle j \rangle}, t_n^{\langle j \rangle})$.

Since retweets as well as replies carry up-to-date copies of their source, p is recursive. The set of p's derivations is then defined as

$$\mathcal{P}_n^{*\langle v \rangle} = \{p_n^{\langle v \rangle}, p', p'', ..., p^*\}, \tag{3.16}$$

and we call the subset of v's already collected messages $\mathcal{C}^{\langle v \rangle} \subseteq \mathcal{M}^{\langle v \rangle}$ with $n = |\mathcal{C}^{\langle v \rangle}|$ to denote the amount of collected messages at the time of observation. Consequently,

$$\mathcal{M} \coloneqq \bigcup_{v \in V} \mathcal{M}^{\langle v \rangle} \tag{3.17}$$

describes the entirety of Twitter's messages and subsequently $\mathcal{C} \subseteq \mathcal{M}$ depicts the set of collected messages over V. Hence follows that $\mathcal{T} = \mathcal{M}/\mathcal{C}$ is the set of not yet collected messages with $|\mathcal{C}|/|\mathcal{M}|$ as an indication for global sample completeness. Correspondingly, we get $\mathcal{T}^{\langle v \rangle} = \mathcal{M}^{\langle v \rangle}/\mathcal{C}^{\langle v \rangle}$ and $|\mathcal{C}^{\langle v \rangle}|/|\mathcal{M}^{\langle v \rangle}|$ specific to v. Following

[18]Twitter Stats: `https://bit.ly/3naV9MI`

Pfeffer [164] we assume

$$\forall t_j, t_i \in \mathcal{M}. \neg\exists. t_i = t_j \wedge j \neq i \text{ with } t_1 < t_2 < ... < t_n \tag{3.18}$$

which allows us to use timestamps t as identifiers for messages. We define the following problem.

Problem 2 *Given a period from t_i to t_j how to estimate $|\mathcal{T}^{\langle v \rangle}|/|\mathcal{M}^{\langle v \rangle}|$ or $|\mathcal{C}^{\langle v \rangle}|/|\mathcal{M}^{\langle v \rangle}|$ respectively and moreover keep track of sparse regions in v's timeline.*

Each request to Twitter's Timeline API returns a maximum of 200 messages. Since the API allows for specifying the starting point for these messages using either a message id or a timestamp, solving Problem 2 implies using requests in an optimal manner.

Since each user v's profile contains her or his last message as well as a message count, we can introduce a data structure that allows us to estimate sample completeness. We refer to this data structure as the block list

$$\mathcal{B}^{\langle v \rangle} = \{\mathcal{B}_1^{\langle v \rangle}, \mathcal{B}_2^{\langle v \rangle}, ..., \mathcal{B}_n^{\langle v \rangle}\}. \tag{3.19}$$

Here, a block is defined as

$$\mathcal{B}_i^{\langle v \rangle} = \langle t_n^{\langle v \rangle}, |\mathcal{C}^{\langle v \rangle} \cap [t_n^{\langle v \rangle}, t^{\langle v \rangle}_{n-\Delta\mathcal{B}_{i-1}^{\langle v \rangle}})|, \Delta\mathcal{B}_{i-1}^{\langle v \rangle} \rangle \tag{3.20}$$

with $\Delta\mathcal{B}_{i-1}^{\langle j \rangle}$ being the amount of messages since the previously defined block. Consequently $\mathcal{B}_i^{\langle v \rangle}$ is full when

$$|\mathcal{C}^{\langle v \rangle} \cap [t_n^{\langle v \rangle}, t_{n-1}^{\langle v \rangle})| = \Delta\mathcal{B}_{n-1}^{\langle v \rangle}. \tag{3.21}$$

In other words, a block is full when the number of elements in the intersection between the already collected messages and the existing messages is equal to the number of elements since the last block. If two subsequent blocks are full we can merge them by defining the sum of both blocks as

$$\mathcal{B}_i^{\langle v \rangle} + \mathcal{B}_{i-1}^{\langle v \rangle} = \langle t_i^{\langle v \rangle}, |\mathcal{C}^{\langle v \rangle} \cap [t_i^{\langle v \rangle}, t_{i-2}^{\langle v \rangle})|, \Delta\mathcal{B}_{i-1}^{\langle v \rangle} + \Delta\mathcal{B}_{i-2}^{\langle v \rangle} \rangle \tag{3.22}$$

To estimate the number of collected messages, we need to differentiate between the completed case, in which the blocks are full, and the case of incomplete blocks.

For simplicity, we assume a uniform message distribution for any given block. This means that: (a) When the beginning and ending interval of the period is the header of a full block, we can simply count the collected messages over the total amount of messages. (b) If the beginning and ending interval of the period is inside a block, the block's fraction is linearly interpolated by the exposure to the period and added to the sampled fraction.

To collect as many messages as possible from the user timelines, we combine three API endpoints that we evaluate on the basis of message yield, data granularity, and call volume.

First, the profile endpoint allows for querying up to one hundred user profiles per request, including the user's last message t_n, along with a message count $n^{\langle v \rangle}$ indicating the total amount of messages published. Therefore we consider

$$\mathcal{C}'^{\langle i \rangle} = \mathcal{C}^{\langle i \rangle} \cup \{(p_n^{\langle i \rangle}, t_n^{\langle i \rangle})\}, p_n^{\langle j \rangle} \in \mathcal{P}^{*\langle j \rangle} \tag{3.23}$$

with the derivative function (Equation 3.16), allowing access to a multiple of a hundred tweets. As $n^{\langle v \rangle}$, the total number of v's messages is known, we can use it as the ground truth for an ongoing sample and calculate the sample completeness from $|\mathcal{C}|/n^{\langle j \rangle}$. Because there are large request quantities available for the profile endpoint, we face high call volume, high yield, and almost no data control, since a profile contains only the last tweet.

In the literature, the second endpoint is often used to perform real-time topic detection and sentiment analysis [164–166]. The streaming endpoint allows for receiving real-time updates from up to 4000 users, simultaneously.

The timeline endpoint, which we discussed at the beginning of Section 3.3.4 returns 200 messages with a single request but not more than the past 3200 messages. The special feature here is that we can determine the start of the timeline block we want to query by means of a timestamp/message id t_n. The timeline endpoint has a medium call volume but offers the highest data control as it is possible to define a time interval with medium yield. Thus, we will examine the timeline endpoint further as it also offers more data granularity and the possibility to retrieve previous messages.

In general, users follow a statistically significant behavior, which can be approximated with suitable models, as shown in the literature [167–169]. Except for company accounts, most users post on Twitter according to some schedule. However, as the time between two sampling visits to an infrequently posting user can be weeks, we can assume that a user's posting frequency over a more extended period is linear.

Therefore, a seasonal frequency was chosen to predict the user messaging frequency. Ideally, we want to sample users immediately after they have created a new message or at some later point when a user has referenced the previous message, thus providing a copy of the message in a retweet or comment.

Our proposed algorithm draws users from a random distribution. Users selected in this manner will be fetched from the online network using one of three different fetching strategies. The algorithm's goal is to use the calls to the API under the given budget as efficiently as possible. It is trivial to see that the call budget should be used in its entirety as there is no cost associated with using the available budget during a reset time frame. Thus, our goal is to maximize the amount of new data obtained by each call.

We assume that the set of users V is known prior to sampling. As we only focus on retrieving messages, the set of users does not change over time. To sort and sample users utilizing their activities, the algorithm is given a function $\rho : V \mapsto \mathbb{Q}$, which assigns a posting frequency to every user. Thus, the average frequency of a user is known to our algorithm through ρ.

We make use of previous work by Bild et al. [168] and Mathews et al. [167] who found that the user tweet generation ρ follows a power-law distribution with a lognormal cutoff.

Under the assumption that the message creation behavior follows a power-law distribution, we can separate the users into three groups, with the heavy-tail group creating most tweets (*top*). The other groups are the active users (*intermediate*) and the least active users (*weak*). Naturally, users that only read but do not post are not analyzed by the algorithm.

We define two external parameters α and β which are used to set the lower and upper bound of the *intermediate* group. Using the bounds α and β we can separate users into three sets $\mathcal{V}|_{0,\alpha}$ as the *weak* users, $\mathcal{V}|_{\alpha,\beta}$ the *intermediate* users, and $\mathcal{V} \setminus \mathcal{V}|_{0,\beta}$ as the *top* users.

To get all messages of the *top* group the streaming endpoint is utilized to capture the messages. This causes the streaming call to get as many messages as possible. For the *weak* group, it is not cost-efficient to utilize the timeline endpoint for each user as this would require waiting until 200 new messages have been generated. Thus, the timeline endpoint is used for the *intermediate* group to maximize the amount of new data generated per call. The profile endpoint is primarily used to sample users in the *weak* group. The more messages are generated by a user the more efficient it is under the call budget to use the batched timeline endpoint for that user.

Our algorithm draws users from a random distribution Ψ and depending on the section the user is in we sample the profile using different endpoints.

Algorithm 1 Sectioned Sampling

$batchProfile \leftarrow budget$
$\mathsf{top} \leftarrow \mathcal{V} \setminus \mathcal{V}_{0,\beta}$
startFetchStream(top)
while true **do**
    $D \leftarrow \emptyset$
    $\mathcal{C}' \leftarrow \mathcal{C} \cup$ fetchStream(top)
    while $|D| < batchProfile$ **do**
        $D \leftarrow D \cup$ select($\Psi_{\alpha,\beta}$)
    $\mathsf{users} \leftarrow$ fetchProfiles(D)
    for all $u_i \in \mathcal{V}$ **do**
        $\mathcal{C}'^{\langle j \rangle} \leftarrow \mathcal{C}^{\langle j \rangle} \cup \{(p_n^{\langle j \rangle}, t_n^{\langle j \rangle})\}, p_n^{\langle j \rangle} \in \mathcal{P}^{*\langle i \rangle}(u_i)$
        if $u_i \in \mathcal{V}|_{\alpha,\beta}$ **then**
            $\mathcal{C}'^{\langle j \rangle} \leftarrow \mathcal{C}^{\langle j \rangle} \cup$ fetchTimeline(u_i)
end

Chapter 4

Social Network Modelling

Contents

In this chapter, we make use of the full information contained in tweet-retweet actions to explore a framework of social connections, which, while disregarding the specificity of text content, incorporates information beyond pure binary edge labelling as in the case of Twitter's follower relations. In particular, we define and weight social connectivity between pairs of Twitter users, keeping track of the frequency of shared content and the time elapsed between publication and sharing. Our framework is applied to one particular case of the Twitter network from which we derive a large-scale interaction (see Section 2.2.1). We depict a central part of this network in Figure 4-1. Moreover, we also present a preliminary topological analysis of the derived network using standard tools from network analysis [170]. Finally, we discuss how to apply this framework as a basis for investigating spreading phenomena of particular contents like the Corona and 5G conspiracy discussed in Chapter 5.

We archive this by introducing a quantitative metric that measures the strength of agreement for user pairs based on their communication and sharing behavior. In

particular, we define a weighting function based on the reaction time and frequency of retweets between pairs of users. We argue that social networks based on this metric are more appropriate for studying information diffusion than social networks based on unweighted expressions of interest that reflect how individuals have chosen to relate to each other like the follower relationships underlying Twitter's social network.

We begin in Section 4.1 by presenting the dataset used to build the network and continue in Section 4.2 with the methodology. In Section 4.2.1 we introduce the weighting function used to measure the interaction between pairs of users and present the properties resulting from applying it to all pairs of users in the dataset. Section 4.2.2 describes the building process for the entire network, which we discuss in detail in Section 4.3. The topological analysis of the network properties in Section 4.3 is further subdivided according to the degree of adjacency. Thus, in Section 4.3.1 we discuss the first neighborhood properties such as degree distributions and accumulated edge weights to a node. In Section 4.3.2, we study the neighborhood and the neighborhood of the neighborhood and their relation to each node. Finally, in Section 4.3.3, which is the last part of the topological analysis, we analyze the overall network properties from a bird's eye view.

4.1 Dataset

The data collection took place using FACT, including the scraping strategies we introduce in Section 3.3.4. The latter is required since we aim to fetch massive amounts of data that are as *dense* as possible. Here, dense refers to a comprehensive historical sample of tweets for each user. Furthermore, we target users that are likely to interact with each other and thus require a network-based scraping as introduced in Chapter 3.

The data collection started in late 2019 and stretched to mid-2020. During this time, we analyzed more than one billion historical posts from user timelines. Among them are 400 million tweets and 300 million retweets. All statuses in total are written or shared by about 30 million users. Figure 4-2 shows the distribution of tweets per user and a breakdown into languages. The goal was to collect and analyze as much tweet and retweet metadata as possible.

Based on a set of 2638 Twitter accounts of German personalities close to politics that we derive from the Twitter lists in Table 3 (Appendix), we first fetch a user's most recent 3,200 tweets and later their first and second neighborhoods in the follower network. Subsequently, we analyze the most recent tweets of the newly acquired users,

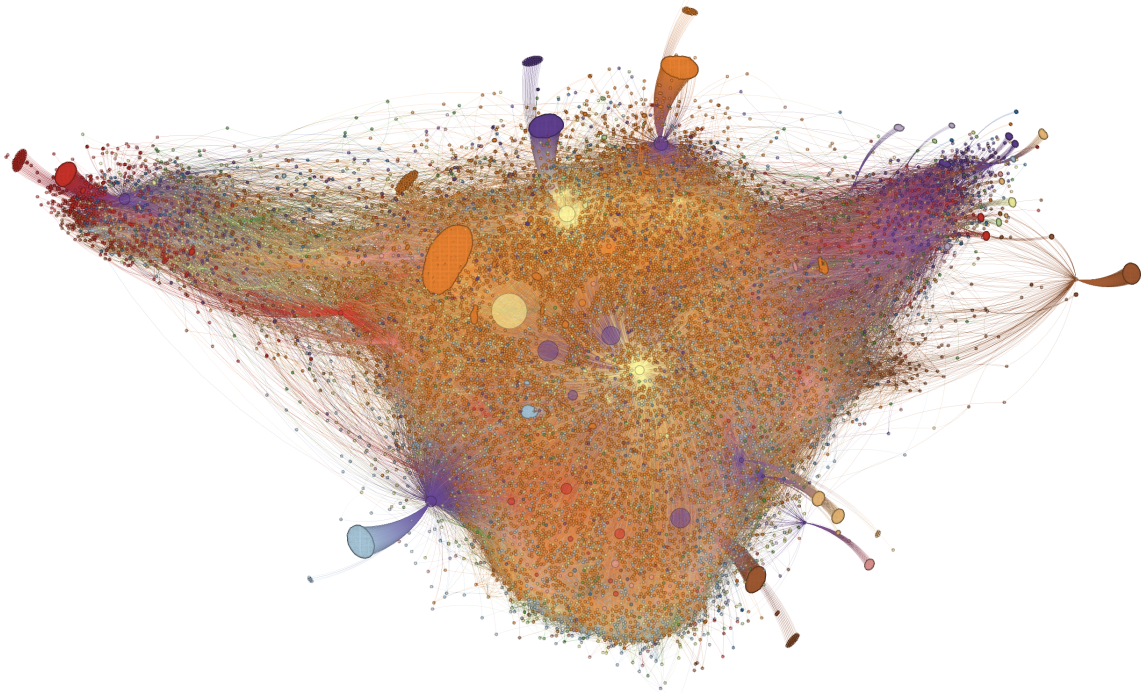

Figure 4-1: Illustration of a selection of the network that we present in the course of this work. The illustration contains approximately 25.000 nodes and 300.000 edges. The process for creating this illustration is as follows. First, we extract the main component, see Fig. 4-4 and transform it into an undirected network. Second, we perform 10.000 breath first searches (BFS) among a subset of 50.000 nodes of the main component. Each BFS starts from a unique randomly node. For each of the nodes visited in a BFS pass, we update a counter of the number of visits over all performed runs. The result is a statistic that keeps track of the number of visits, over all 10.000 BFS runs, for each node of the main networks. Later, in a third step, another batch of 10.000 BFSs is performed to determine the network that archived the highest average number of visited nodes, i.e. most representative connected sample. The node size reflects its degree, and its color corresponds to the number of visits during step one. The brighter the color, the more often the node was visited.

ignoring those that are not written either in German or in English. We repeated this process iteratively with users only tweeting in German until the number of newly added users decreased significantly.

We would like to point out once again that Twitter's limit of only 3,200 tweets that are accessible for each user timeline implies that only a small portion of the tweets authored by highly active users contribute to our dataset. Moreover, Twitter allows its users to set their profiles to private, i.e. allowing only direct followers to access that user's content. For that reason, data from private profiles is excluded from our consideration. Note that the work presented here examines the interaction

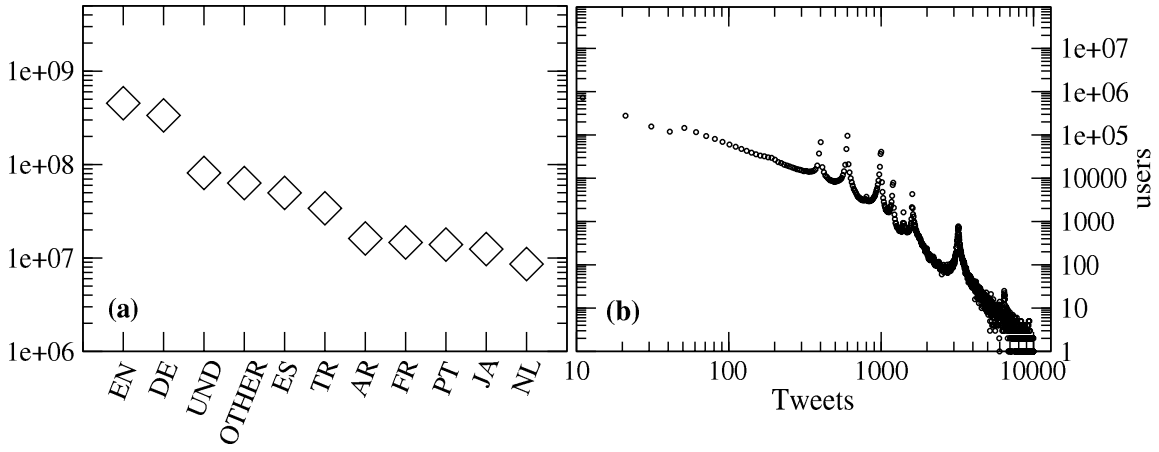

Figure 4-2: (a) number of tweets by language for the ten most tweeted languages. In total, there are tweets in 28 different languages. The languages shown here are starting from the left: EN = English, DE = German, UND = Undefined, OTHER = Languages not listed, ES = Spanish, TR = Turkish, AR = Arabic, FR = French, PT = Portuguese, JA = Japanese and NL = Dutch. English, German are the two dominating languages followed by tweets for which the language could not be determined (Undefined). The dominance of German-language tweets is due to the construction method (see Section 4.2.2), which was geared towards the German-speaking world. The eighteen less tweeted languages are summed up under the label OTHER. (b) Number of tweets per user is delineated on a log scale. The peaks can be explained by the collection process. Each request to Twitter's API for user-timelines returns a batch. The collection process is optimized for batch sizes and thus creates a binning.

that arises from the sharing of content. Users who have never retweeted another user are therefore not considered.

4.2 Building the interaction network

4.2.1 Assessing the intensity of pairwise interactions for information exchange

To assess the strength of connections between pairs of users, we derive two main properties from the Twitter dataset, namely, we define an average reaction time for a retweet and a so-called "tweeting rate". In this way, we postulate that the number of retweets and the reaction time with which two users exchange information are the fundamental properties for describing their connectivity.

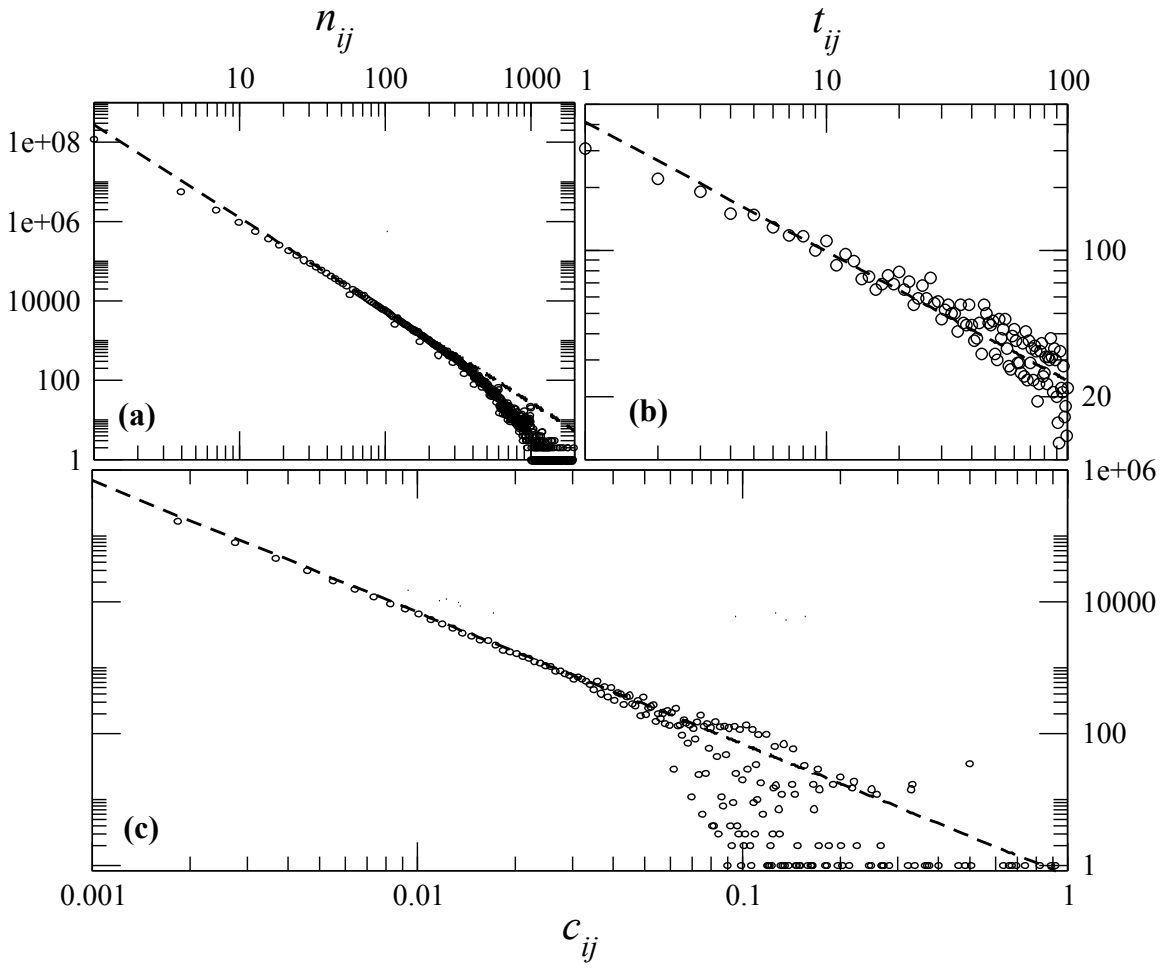

Figure 4-3: (a) Distribution of the number of tweets n_{ij} that a user j retweets from another user i. (b) Distribution of the time-span t_{ij} between the instant user i tweeted a tweet and user j retweeted it. (c) Distribution of the values of the weights c_{ij} as defined in Equation (4.3).

To measure the retweeting rate, we introduce the quantity

$$P_{ij} \coloneqq \text{set of tweets authored by } i \text{ and shared by } j.$$

The retweeting ratio at which user j retweets user i is given by $1/n_{ij}$, where $n_{ij} = |P_{ij}|$ is the cardinality of P_{ij}.

In Figure 4-3a we see the distribution of P_{ij} from the Twitter dataset. From this plot, it is clear that for the majority of user-pairs, there is either a rare exchange of information or no exchange at all. While there are other forms of interaction, such as private messages, comments, or quoted-retweets, we define the weighting function exclusively in accordance with retweets. We argue that this restriction is reasonable

due to the "purity" of retweets, i.e., the lack of opportunity to comment on the shared content. Retweets inherently prohibit the negation of the initial statement and can thus imply agreement [171–173]. Admittedly, this statement is not universal because situation-, users- or target group- properties indirectly provide context, but the coherence results from a retweet's nature and seems, as such, conceivable.

To measure the reaction time, we first introduce the timestamp at which user i publishes a tweet m, represented henceforth as $t_i(m)$, and the time at which user j shares i's given tweet, represented as $t_j(m)$. Thus, the reaction time of user j to i's tweet m is the time difference between a tweet m authored by i and shared by j is given as

$$t_{ij}(m) \coloneqq t_j(m) - t_i(m)\,. \tag{4.1}$$

This time interval is used to define the average reaction time over all tweets that were shared between i and j

$$T_{ij} \coloneqq \frac{1}{n_{ij}} \sum_{m=1}^{n_{ij}} t_{ij}(m)\,. \tag{4.2}$$

In Figure 4-3b we show the distribution of the reaction time differences over all tweet-retweet pairs $t_{ij}(m)$ in seconds, for the entire dataset.

The average reaction for retweeting happens typically within the first seconds. Moreover, similarly to n_{ij}, $t_{ij}(m)$ seems to follow approximately a power-law. A power-law fit (dashed lines) yield exponents of $-7/3$ (see Figure 4-3a) and $-5/3$ (see Figure 4-3b) respectively.

We claim that, up to some extent, the reaction time reflects the level of connectivity the retweeter j has with respect to the tweet author i. Indeed, we assume that the average reaction time implicitly represents a gauge of activity. Users who are more active react more rapidly to each other's content. Moreover, by the very nature of things, someone who approves the same attitude and is particularly interested in someone else's content will not hesitate or need to be convinced and, thus, tends to react instantly. Furthermore, Twitter's option to follow users, i.e., if user j follows user i, user j receives i's tweets exclusively via his or her newsfeed, allows active users to react instantly. User j activating Twitter's build-in notification feature can even extend the following mechanism. For each of i's tweets, j then receives not only a newsfeed update but also a push message, allowing j to react even more rapidly. It is important to note that a follower relation does not necessary imply interaction and rapid interaction in particular. It is entirely possible to follow but never retweet.

Suppose the connection between two users is consistent over the entire history of

their interactions, characterized by short reaction times, i.e., the willingness to share the other's statement. If a user consistently responds quickly to another's messages by sharing them without commenting, we deem that user to be particularly active. In this scope, we introduce the term *connectivity* to describe the stronger or weaker tendency to share Twitter content. Having defined both the retweeting rate and the reaction time, we can now introduce the a property which measures the connectivity strength between two users, i and j, namely

$$c_{ij}(t) \coloneqq \frac{1}{\sum_{k=1}^{N} (n_{ik} + n_{kj})} \frac{n_{ij}}{T_{ij}} \,. \tag{4.3}$$

The weight c_{ij} accounts for the number of tweets that user j shared from user i with the corresponding reaction time, and it increases inversely to the frequency and reaction time. Notice that we divide the fraction n_{ij}/T_{ij} by two sums over the total number N of users in the dataset. One sum, $\sum_{k=1}^{N} n_{ik}$, is the number of tweets published by i and shared by any user, representing a sort of popularity of user i. The other sum, $\sum_{k=1}^{N} n_{kj}$, is the number of tweets published by any user and shared by j, representing a sort of willingness to share content of user j. In this way, the connectivity weight c_{ij} is based on the assumption that (i) it increases with the total number n_{ij} of tweets from user i to user j, normalized by the additive effect of i' popularity and j's willingness to share, and (ii) it decreases with the average reaction time T_{ij}, i.e. the longer user j takes to retweet user i the weaker their connectivity is.

In Figure 4-3c we show the distribution of the weights c_{ij} which also follows approximately a power-law. Notice that sharing a tweet is a "directional" activity: a fan of a Pop-star can have a very strong interaction with his or her idol, while, on the other hand, the pop-star has a weak connection with his or her fan.

4.2.2 Beyond Twitter's follower-network: Building an empirical social network with weighted interactions

Based on the dataset presented in Section 4.1 and on the weight function introduced in Section 4.2.1, we derive a network underlying the Twitter social network, as sketched in Figure 4-1. As illustrated in Figure 4-4b the derived network has one main connected component and several smaller isolated "islands." In particular, we performed a connected component analysis, which reveals a major component with a size of about 30 million nodes and 120 million edges. There are about 150 thousand components

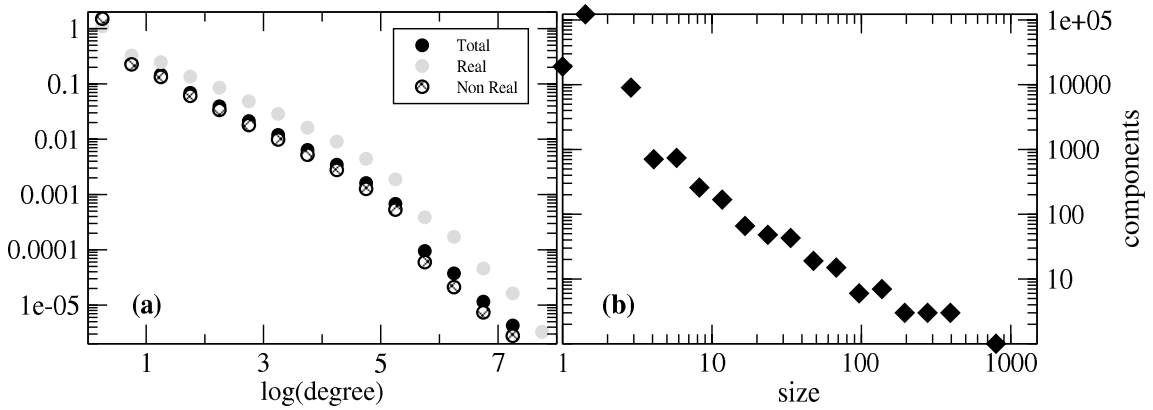

Figure 4-4: (a) Comparison between the degree distribution of the main connected component of the derived network from the Twitter dataset and the one of the known network of followers. The existence of an edge in the Twitter follower network is associated with the c_{ij}-score > 0 of the edge in our derived weighted network. (b) Size distribution of the connected components not included in the main component. All these components have sizes not larger than 1000 nodes, which justifies ignoring these components and focusing the topological analysis on the main component which has approximately 30 million nodes (see text).

overall, with about 400 of them containing more than 10 nodes, 24 containing more than 100 nodes, and just two with more than 1000 nodes. The connected component analysis was performed after converting the directed network into an undirected network. Therefore, a component isolated from the main component is indeed isolated, i.e., users have not shared content with users outside the component, nor has their content been shared by users outside the component. Henceforth, we only consider the main connected component, filtering out all other smaller components. This is justifiable as all other components are negligible.

In this chapter, we provide a framework that allows for large-scale modelling of dynamic processes solely based on Twitter's interaction data. To that end, we now compare our derived network with the follower network accessible through Twitter's API. Twitter gives its users the ability to follow any other user, meaning to subscribe to his or her content and, moreover, to receive notifications if requested. The content written and shared by those who are followed appears in the subscriber's newsfeed. To understand whether sharing content coincides with the active decision to subscribe to another user's content, we checked for each directed edge with $c_{ij} > 0$ whether a corresponding edge exists in Twitter's follower network. Results are shown in Figure 4-4a and indicate that the existence of edges obtained with the proposed approach coincides with the existence of follower connections on Twitter.

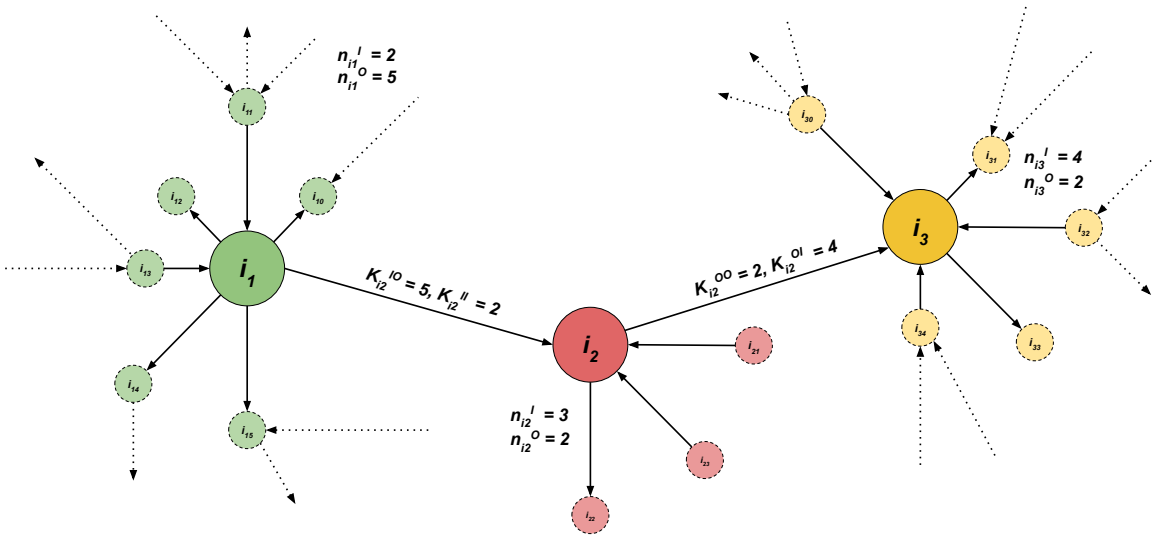

Figure 4-5: Visualization of the topological properties to be investigated in Section 4.3. The Figure does not reflect any structural properties of the derived network itself and is for illustration only. Three adjacent nodes i_1, i_2 and i_3 are shown. The neighborhood of each node is highlighted with the color of the corresponding node, and the node itself is labeled with the respective n_i^O and n_i^I (See Equation 4.4). It should be noted that the node i_2 in the center is connected with an outgoing edge to i_3 and an incoming edge to i_1. Summing the incoming and outgoing edges of i_1's neighbors gives the value for the corresponding $K_{i_2}^{II}$ or $K_{i_2}^{IO}$ (See Equation 4.8). The same applies to the node i_3 but with the outgoing edge from i_2. Therefore, we receive $K_{i_2}^{OI}$ or $K_{i_2}^{OO}$.

This result is not surprising because users who interact with each other by retweeting tend to decide to follow each other. However, there is also the possibility of sharing content from accounts that one does not follow, for example, via third-party websites; even so, this seems not to happen frequently. Twitter's notification and recommender algorithms also presumably contribute to the fact that content is shared more often by followed users. In addition, Twitter makes suggestions for potential followers based on shared tweets, followers, and follows. If a piece of content is shared by someone who is not followed, it is quite conceivable that a follower edge will be added afterwards, based on this recommendation.

4.3 Topological analysis of the Twitter interaction network

In this section, we present a description of the main topological features of the derived network. While the analysis is specific to the German Twitter dataset, it can be extended straightforwardly to other Twitter datasets or similar data with similar data density. We divide the analysis into three parts. First, we describe the first neighborhood's topological properties; later, we will investigate the properties characterizing the nodes' second neighborhoods. A definition for the first and second neighbourhood is given in Section 4.3.1 and Section 4.3.2 accordingly. Finally, we address the entire network's global measures, such as the average shortest path length and betweenness centrality. The aim is to examine the activity, level of connectivity, and impact of our dataset users. A user's influence, activity, level of connectivity, and impact is compared to that of its neighbors, so that statements of the form "The more influential a user i, the less active are those that i influences" can be derived. Figure 4-5 illustrates the spatial meaning of the properties considered in this section.

4.3.1 Assessing the first node neighborhood: *"activity"* and *"impact"* of each user

We first label the set of nodes that have an outgoing edge to node i as $\mathcal{N}_i^I$, and the set of nodes have an incoming edge from node i as $\mathcal{N}_i^O$. The total set of nodes, either in one way or the other, is then given by $\mathcal{N}_i^A = \mathcal{N}_i^I \cup \mathcal{N}_i^O$. The number of nodes in $\mathcal{N}_i^I$, $\mathcal{N}_i^O$ and $\mathcal{N}_i^A$, which we call in-degree, out-degree, and degree respectively, are given by

$$n_i^I \coloneqq |\mathcal{N}_i^I|\,, \tag{4.4a}$$

$$n_i^O \coloneqq |\mathcal{N}_i^O|\,, \tag{4.4b}$$

$$n_i^A \coloneqq |\mathcal{N}_i^I \cup \mathcal{N}_i^O|\,. \tag{4.4c}$$

In order to clarify the interpretation, we name the degrees as follows:

$$n_i^I \coloneqq \text{size of } i\text{'s influenc}\mathbf{ing} \text{ neighbourhood}$$

and

$$n_i^O \coloneqq \text{size of } i\text{'s influenc}\mathbf{ed} \text{ neighbourhood}\,.$$

Influencing in this context only points to the number of different users that i retweeted during its entire lifetime (record in the examined dataset). Another way to put it is that the tie strength c_{ij} is not taken into account. We want to indicate that this function is also helpful as a measure of opinion diversity. Users who share the content of many other users tend to form opinions based on this diversity and are therefore more robust towards content without truth. However, this is only valid with restrictions. The nature of our study does not allow a judgment on so-called filter bubbles [174] or echo chambers [175]. If a user shares many users' content, but all those from whom the content is shared are only linked to each other, n_i^I is not a suitable indicator for the diversity of opinions.

The same applies to the influenced neighborhood n_i^O which is the number of different users that retweeted i at least once throughout their entire lifetime, i.e., with respect to all data points in our dataset. Again, the term Influenced is not a measure of the depth to which i's content diffuses into the social network. Here, all those who share i's content could exist completely isolated from the rest of the network, talking only to themselves.

The weighted counterparts of both these properties (n_i^O, n_i^I), which we represent by w_i^I and w_i^O, account for the weighted degree respectively for all incoming and outgoing neighbours of i, are defined as

$$w_i^I := \sum_{m \in \mathcal{N}_i^I} c_{mi}\,, \tag{4.5a}$$

$$w_i^O := \sum_{m \in \mathcal{N}_i^O} c_{im}\,, \tag{4.5b}$$

$$w_i^A := w_i^I + w_i^O\,. \tag{4.5c}$$

They have an important meaning for the topological analysis, namely:

$$w_i^I := \text{activity of user } i\,,$$

$$w_i^O := \text{impact of user } i\,.$$

Our definition of activity in this context does not consider the diversity of the sources. A user who retweets a particular user k often and with fast reaction time can be as active as a user who shares content from many users but does so infrequently. If a user is quick at sharing and thus consuming content and shares, moreover, often or from many other users, the user seems active. For this reason, we believe that active is an

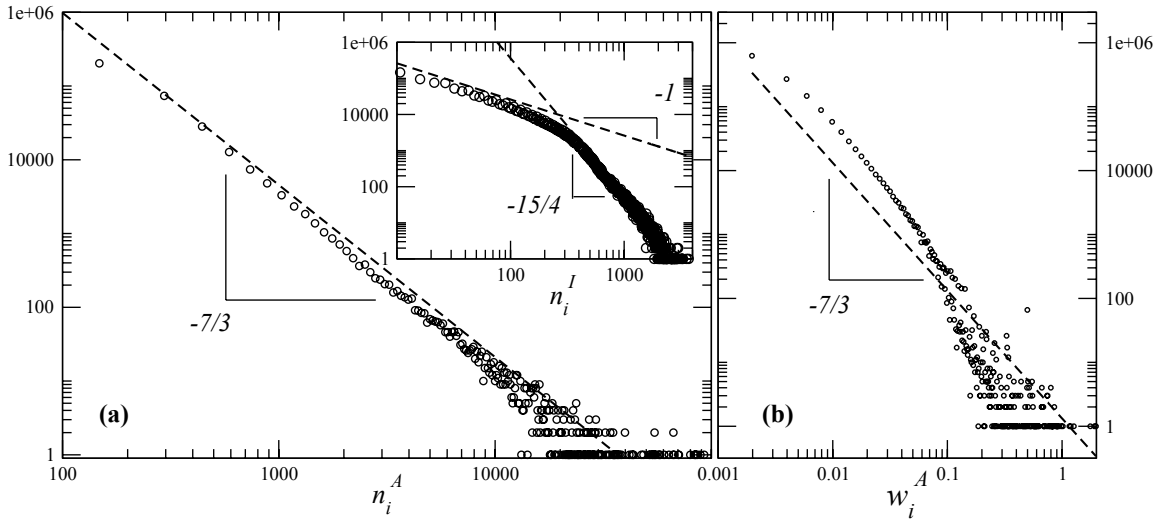

Figure 4-6: (a) Distribution of the total number n_i^A of neighbors of a node i. In the inset we plot the distribution of the size n_i^I of the influencing neighborhood of each user i. Note: the distribution of the size n_i^O of the influenced neighborhood by each user i follows a power-law similar to the total number of neighbors. (b) Distribution of the total weighted degree w_i^A, which sums up the activity and the impact of each user.

appropriate term here. Figure 4-6a shows the distribution of n_i^A (see Equation 4.4c) on a log-log plot. We observe a clear power-law distribution. The interpretation of this behavior is that most users neither influence a wide variety of different users, nor are they influenced by a wide variety. The overlay in the same plot depicts the distribution of n_i^I (see Equation 4.4a) also on a log-log scale. Up to $n_i^I < 130$ the distribution follows x^{-1} and later for $n_i^I > 130$ $x^{-15/4}$ meaning the number of users influenced by at least 130 users decreases faster than the number of users influenced by less than 130 users.

Users i with $n_i^I > 130$ are rare. They are *influenced* by many different users, although this number alone does not imply that they are exposed to diverse opinions. Many such users likely make use of Twitter in a professional or semi-professional capacity.

Figure 4-6b shows the distribution of w_i^A (see Equation 4.5c). Since w_i^A is the sum of w_i^O or a users activity and w_i^I or a users impact, one observes that up to $w_i^A < 0.1$, users tend to show a positive correlation between their activity and impact.

Figure 4-7a shows the average size of a user's influenced neighborhood (n_i^O) over all users having an influencing neighborhood (n_i^I) of the same size in a log-log plot. The plot is visibly divided into three regions. Region I shows that the in-degree increases linearly with the out-degree for users i having $n_i^I < 100$. In other words, the number

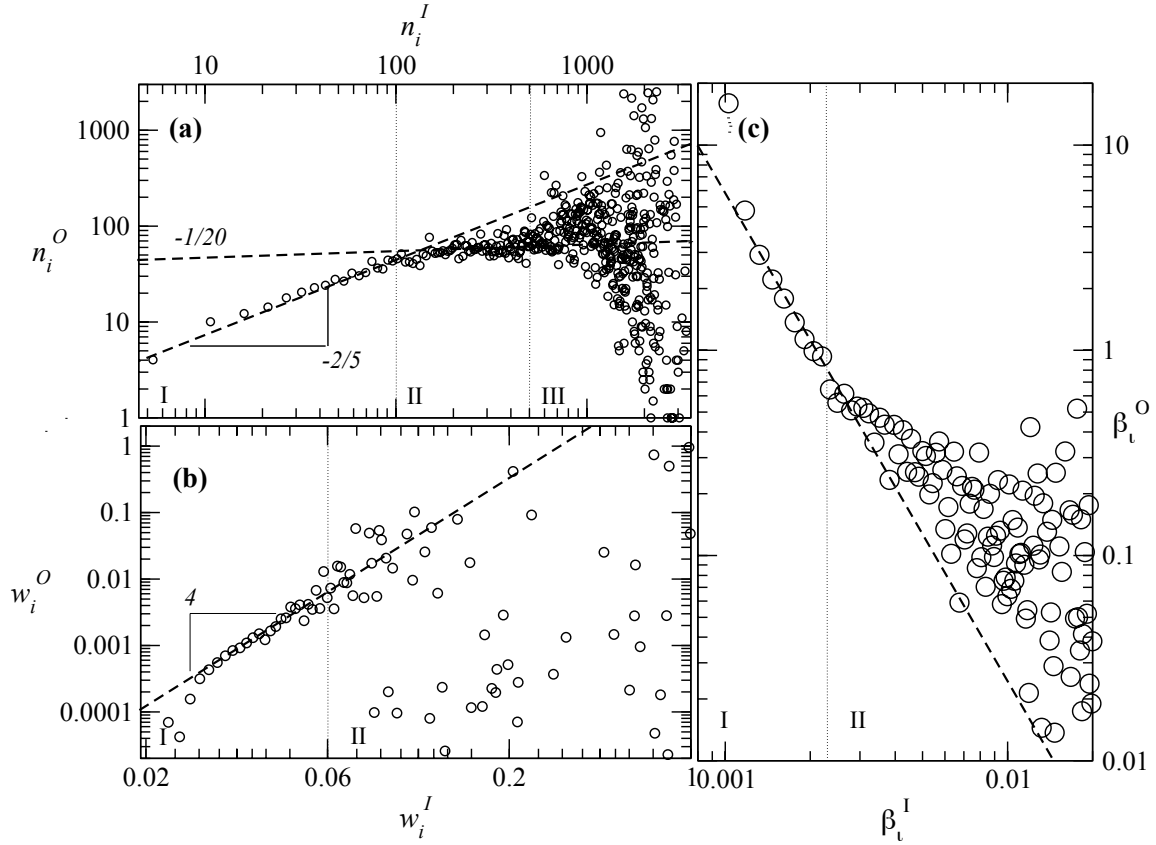

Figure 4-7: Comparing activity with impact and influencing neighborhoods with influenced neighborhoods: (a) $n_i^I \times n_i^O$: those who influenced the most are those who are influenced the least; (b) $w_i^I \times w_i^O$: activity and impact are not strongly correlated (c) $\beta_i^I = w_i^I/n_i^I \times \beta_i^O = w_i^O/n_i^O$: how much is the "influenced" level of i correlated with its respective "influencing" level?

of people who influence a certain user i increases with the number of users that are influenced by i. The assumption that this region contains the group of occasional or "normal" users seems reasonable. Region II still shows a linear increase. In contrast to Region I however, the smaller slope indicates that users who are influenced by more than 300 but less than 700 people influence fewer users than they are influenced by. Region III differs from the other two in that it is divided into the influenced but not influential users in the lower right corner and the highly influential users in the upper right. Naturally, this group is the most impactful for the spread of information, i.e., though with a noisy value of n_i^O all the users in this region are characterized by a large n_i^I.

Figure 4-7b shows the relationship between activity and impact rather than the neighborhoods. Like Figure 4-7a, it is a log-log plot between w_i^O and w_i^I. This plot

is divided into two regions. Region I reflects the results of Figure 4-6a and shows a linear correlation between a user's activity and influence. This indicates that up to a certain degree one can become more influential by being more active. However, this is true until reaching a certain degree of activity. On the other hand, in Region II, no structure can be discerned, meaning there is no strong correlation between a user's activity and impact.

To filter out the influence of the size of (influencing or influenced) neighborhoods, we introduce additionally two other measures, namely

$$\beta_i^I := \frac{w_i^I}{n_i^I}, \tag{4.6a}$$

$$\beta_i^O := \frac{w_i^O}{n_i^O}, \tag{4.6b}$$

which we interpret as the average impact (resp. activity) of user i to (resp. from) each one of his outgoing (resp. incoming) neighbors. In Figure 4-7c we show the average impact as a function of the average activity, uncovering an approximate inverse relation between both properties. The accounts that, on average, influence many people are themselves influenced by fewer people. Such individuals are often referred to as opinion leaders or trendsetters. In Region I, one can observe that the users that are, on average, not intensely influenced by their neighbors influence others more. This behaviour seems plausible since being influenced intensively by many neighbors requires time to consume and process the content, and the majority of users are not very active (see Figure 4-7). Users in Region I seem to maintain few but very strong connections. It should be taken into account that the presented weighting function normalizes over the connections to the respective neighbors. So users who often share content from a few other users achieve high averages. Region II shows that this trend continues with increased variance.

4.3.2 The second neighborhood: correlation between the activity and impact of a node with the activity and impact of its neighborhood

Based on the preceding definitions, it is now possible to define additional properties which connect the first two neighborhoods of a user i. In the following, we investigate how influencing neighbourhood n_i^I behaves to its influencing neighborhood by dividing the sum of i's incoming neighbours' in-degrees by i's in-degree (see K_i^{II} in

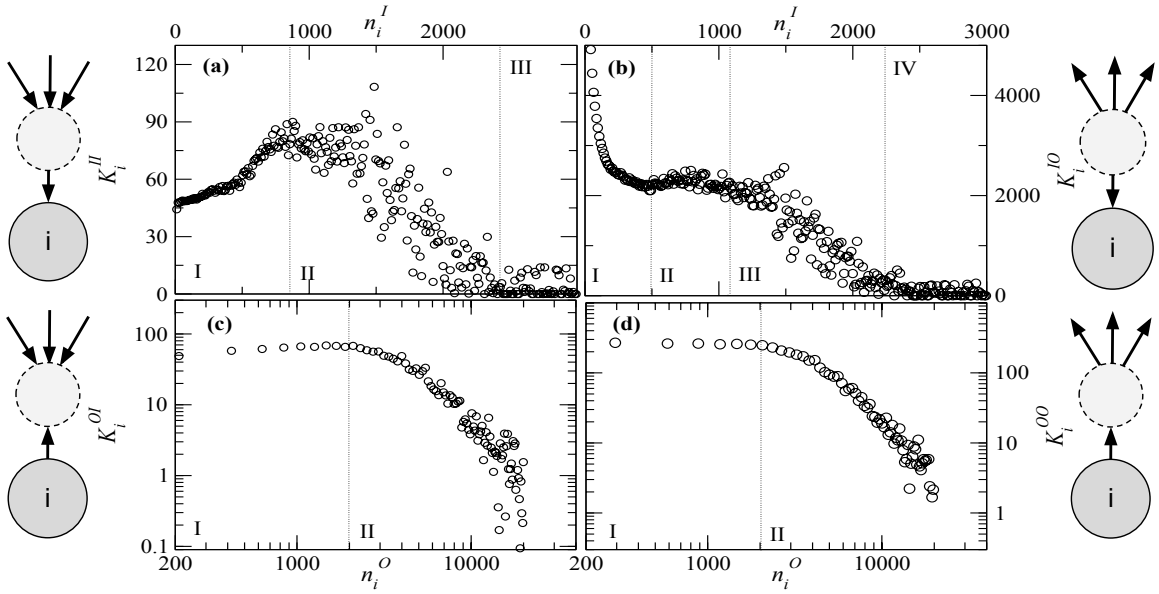

Figure 4-8: Relation between first and second neighborhoods, identifying regions with different user behaviors. (a) $K_i^{(II)} \times n_i^{(I)}$: The degree of correlation between the size of the influencing neighborhood of neighbors of a certain user i with the size of their influencing neighborhood? There is a critical size beyond which disassortativity is observed. (b) $K_i^{(IO)} \times n_i^{(I)}$: The degree of correlation between the size of the influenced neighborhood of user i with the size of its neighbors influencing neighborhood. Here one observes three regions: (I) one with few influencing neighbors and a lot of influenced neighbors (the "stars" of Twitter); (II) one "normal" region, where there is a typical value of the influenced neighborhood; and (III) a region with many influencing neighbors, who basically do not influence anyone. (c) $K_i^{(OI)} \times n_i^{(O)}$: The degree of correlation to which my neighbors are influenced by others. Here one observes complete disassortativity. (d) $K_i^{(OO)} \times n_i^{(O)}$: The degree of correlation between he size of the influenced neighborhood of neighbors of user i with the size of their influenced neighborhood. Similar behavior as in (c).

Equation 4.7a). Analogous to K_i^{II} we define K_i^{IO}, K_i^{OO} and K_i^{OI} as follows.

$$K_i^{II} \coloneqq \frac{1}{n_i^I} \sum_{m \in \mathcal{N}_i^I} n_m^I \,, \tag{4.7a}$$

$$K_i^{IO} \coloneqq \frac{1}{n_i^I} \sum_{m \in \mathcal{N}_i^I} n_m^O \,, \tag{4.7b}$$

$$K_i^{OO} \coloneqq \frac{1}{n_i^O} \sum_{m \in \mathcal{N}_i^O} n_m^O \,, \tag{4.7c}$$

$$K_i^{OI} \coloneqq \frac{1}{n_i^O} \sum_{m \in \mathcal{N}_i^O} n_m^I \,. \tag{4.7d}$$

Here, (I) indicates the influencing and (O) the influenced neighborhood size of i's influencing (I) or influenced (O) neighbors.

Figure 4-8a shows $K_i^{(II)} \times n_i^{(I)}$ or the average size of the neighbourhood that influences each of the users that influences i. In the first region, one can observe that for every user that i retweets the average number of users that i's influencer retweet grows linear. This is true for influencing neighborhood sizes up to 800 influencer. In Region II, a drop in the curve can be observed. With each new influence of i, only users who are less influenced are added. The more one is influenced, the more difficult it is to be influenced by people who are influenced by as many or even more users. Finally, the curve in the third region is constant. Users who are being influenced by more than 2000 other users gain relatively few users who are very influenced.

Figure 4-8b depicts $K_i^{(IO)} \times n_i^{(I)}$ or the average number of users that were influenced by a user that influenced i. This plot can also be divided into four regions. In Region I, up to an in-degree of about 300, users are mainly influenced by influential users. One possible interpretation is that users who are not very active usually follow the stars and share their content, if at all. In Region I, it is also noticeable that the extremely large number of users (compared with Figure 4-8a) who have only shared the content of a few users share the content of users who have an extremely high level of influence. The plot follows a power law in the first region. An appropriate label for users in Region I would be "The Fans". In Region II, the curve flattens out and forms a plateau between users influenced by at least 300 but not more than 1000 other users. For users in this group, they are largely linked to users who are influential to the same extent that they are influenced. This seems especially interesting when comparing with Region I in Figure 4-8a. here the opposite behavior is indicated for the average number of influencers per influencer of i. Therefore, the number of those who influence someone who influences i falls (in the same region), while the number of those who are influenced by those who influence i stagnates with the degree of influence. Region IV in (b) and Region III in (a0 show again the same behavior.

Figure 4-8c shows the correlation between the out-degree of a user i and the average in-degree of all those who are influenced by i. Or, in other words, the mapping shows for each user k who is influenced by i, from how many other users k is influenced (on average). This illustration can also be divided into two regions. In Region I, we observe the group of users who influenced less than 3000 users, and in the second region, we observe the group of users with more than 3000 users influenced.

Figure 4-8d shows the correlation between the out-degree of user i and the average out-degree of all those influenced by i. In other words, the plot shows for each user

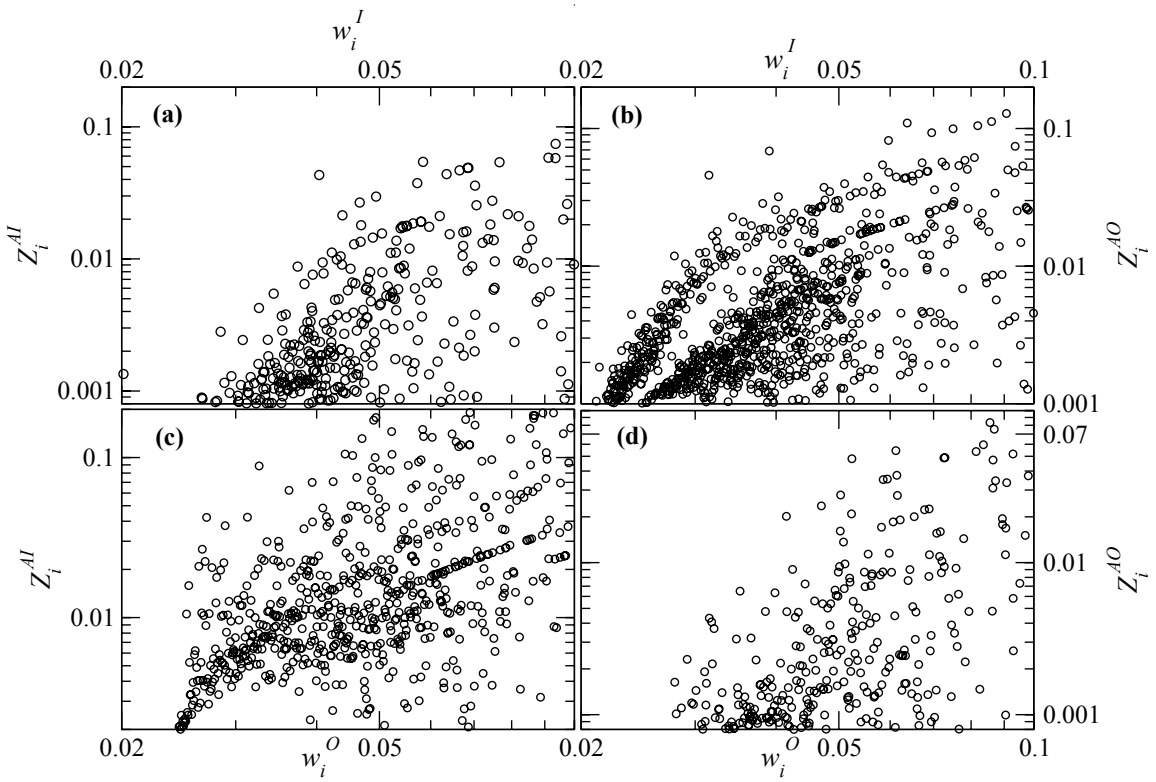

Figure 4-9: Relation between the first and second weighted neighborhoods, identifying regions with different user behaviors. (a) $Z_i^{(AI)} \times w_i^{(I)}$: How the average activity of a node from i's entire neighborhood correlates with i's activity. No significant correlation is observed. (b) $Z_i^{(AO)} \times w_i^{(I)}$: How the average impact of a node from i's entire neighborhood correlates with i's activity. A more pronounced positive correlation is observed, and seemingly the set of users seems to be split into two distinct groups. (c) $Z_i^{(AI)} \times w_i^{(O)}$: How the average activity of a node from i's entire neighborhood correlates with i's impact. A slightly positive correlation is observed. (d) $Z_i^{(AO)} \times w_i^{(O)}$: How the average impact of a node from i's entire neighborhood correlates with i's impact, showing a result similar to (a).

k who is influenced by i, how many other users influence k on average. A similar behavior to Figure 4-8c is observed: two regions are also identified. In the first region, we find the user group that influences up to 2000 different users. Here, the average number of users that a neighbor influences sticks to about a hundred users, indicating that this group's users tend to have influenced similar size neighborhoods. Moreover, users with influenced neighborhoods larger than $n_i^O \simeq 2000$ show disassortativity, typical of famous individuals (stars): the larger their influenced neighborhood, the smaller the influenced neighborhood of their neighbors is.

One additional question is how do the activity and the impact of a user's neighborhood correlate with its own activity and impact. We investigate how the impact of a

user's neighborhood correlates with its own activity and impact using the quantities

$$Z_i^{AI} := \frac{1}{n_i^A} \sum_{m \in \mathcal{N}_i^I \cup \mathcal{N}_i^O} w_m^I, \tag{4.8a}$$

$$Z_i^{AO} := \frac{1}{n_i^A} \sum_{m \in \mathcal{N}_i^I \cup \mathcal{N}_i^O} w_m^O. \tag{4.8b}$$

Figure 4-9a shows how the sum of incoming edge weights, i.e. i's activity, correlates with the average incoming edge weight of i's neighbors or neighbor's activity. Here, we explicitly consider all neighbors, i.e., those who have shared i's content as well as those whose content has been shared by i. Although the figure does not show a strong correlation, it can still be observed that its neighbor's activity exceeds its own for none of i's activity levels. This means that, in general, less active users are influenced by or influence less active users.

Figure 4-9b shows how the sum of incoming edge weights, i.e. i's activity, correlates with the average outgoing edge weight of their neighbors. In other words, how i's activity correlates with the average impact the users he/she influences or is influenced by. Compared to Figure 4-9a, Figure 4-9b indicates a structure. As the level of activity increases, the neighbor's impact grows proportional. Moreover, we can identify two distinct structures for which we do not have an interpretation yet.

Figure 4-9c shows how the sum of outgoing edge weights, i.e. i's impact, correlates with the average incoming edge weight of i's neighbors. In other words, this plot explains how i's impact correlates with the activity of those who either retweeted i or who got retweeted by i. Although (c) is much more blurry than (b) and not quite as separated as (a), a structure emerges. With an increasing impact, i tends to get more active neighbors. However, there is a threshold. All is with an impact $w_i^O > 0.5$ tend to gain proportionally less active neighbors with increasing influence. Notice that a substructure of two distinct regions, similar to Figure 4-9b, is observed, though in this case, it is less clear.

Finally, Figure 4-9d shows that the average impact of a node from a user's entire neighbourhood correlates only weakly with the user's impact, similar to what is observed in Figure 4-9a.

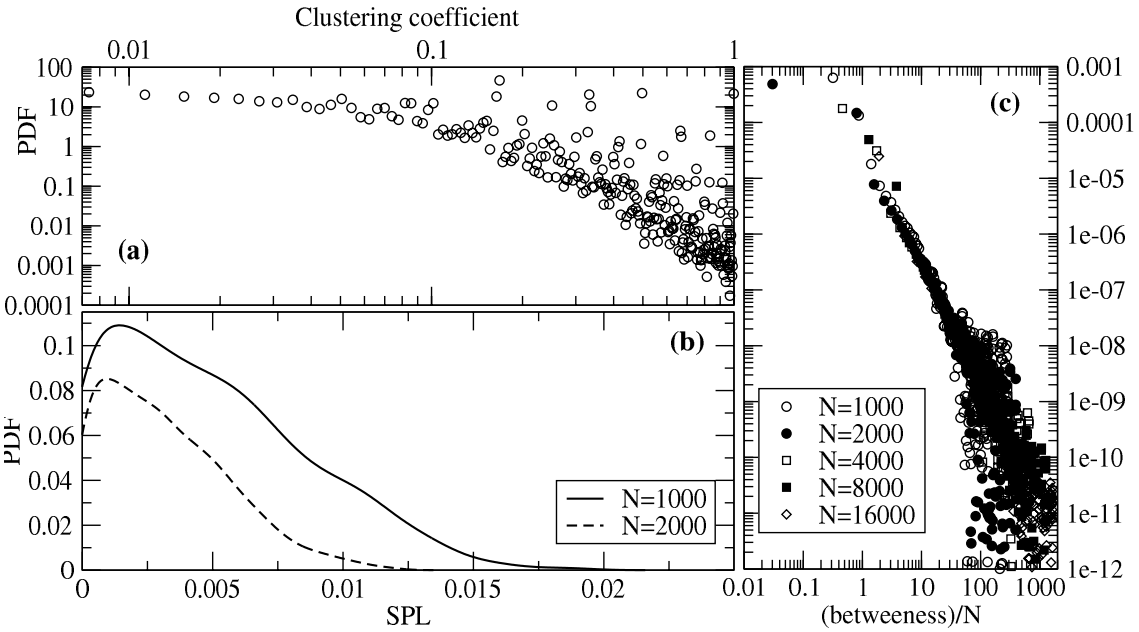

Figure 4-10: (a) Clustering coefficient spectrum of the derived network, (b) the respective shortest path length (SPL) spectrum and (c) betweenness centrality. While SPL, weighted by the values of c_{ij}, shows a unimodal spectrum, a mode at low values, and then an approximately linear decrease, both the clustering coefficient and the betweenness show a power-law decay. For performance reasons, only smaller sets of N nodes sampled from the main component were selected. As one observes, similar behavior is observed for different sizes of the sampled set of nodes, evidencing that results in this Figure are representative of the entire network. We chose a random node from the largest component to create the subgraphs and performed a (directed) breadth first search from this node.

4.3.3 Clustering coefficient and global measures of the network topology

As the final step of our topological analysis of the interaction network behind Twitter, we address three additional quantities. One is the local clustering coefficient while the other two are global measures: the shortest path length, ℓ_{ij}, joining two users i and j, and the betweenness centrality b_i of each user i.

The clustering coefficient, $C_l(i)$, of a user i is the fraction of existing edges among its neighbors among all possible edges they may have. If one observes m of such edges, then

$$C_l(i) = \frac{2m}{n_i^A(1 - n_i^A)}\,, \tag{4.9}$$

where incoming and outgoing connections were considered as single undirected edges.

As shown in Figure 4-10a, one observes a slow decay in the number of nodes with a low clustering coefficient, between 1% and 10%, followed by an approximate power-law decay for higher values of $C_l(i)$.

The shortest path length ℓ_{ij} is defined as the shortest weighted path joining two different users, i and j. As shown in Fig. 4-10b, the spectrum of different ℓ_{ij} is broad and approximately unimodal.

The betweenness measures the level of "importance" of a node in the network by computing the number of shortest paths crossing it. The distribution of node betweenness is plotted in Figure 4-10c, showing power-law behavior again.

Note that, because of the network's size and the resulting runtime for computing the shortest paths, we perform the same procedure as in Figure 4-1. In other words, we sub-sample the component depicted in Figure 4-1, showing that the size of the sub-component is large enough for anaccurate estimate of this quantities: for each of the samples taken, we increase the number of nodes in that sample. Both Figures 4-10b and 4-10c show that this method works because there is convergence in the respective plots, and the result retains its functional shape.

Chapter 5

5G-Corona Connection Event

Contents

In the Chapter 3 and 5, we dealt with the question from which data sources we can obtain large amounts of data of sufficient quality and density based on which it is possible to investigate DWs. Finally, we decided to work with Twitter data and introduced the FACT system, which allows us to build the proverbial haystack (see Section 3.2) from which we now want to extract the DWs.

Extracting a DW, however, confronts us with another challenge, namely, the automatic extraction of data and its assignment to the digital wildfire we plan to extract. In 2020 and 2021, we received more than 15 billion tweets and retweets, about 1 billion user profiles, and more than 100 billion Twitter follower relationships that connect

these user profiles. This enormous amount of data cannot be inspected or analyzed manually. Thus, it is essential to develop methods that are capable of extracting data that belongs to a specific DW without human intervention. The first step towards developing methods for the automatic extraction of data that belongs to a DW must be to understand DWs in general.

In this chapter, we introduce and investigate a DW, namely, the Corona and 5G Connection in which the claim that 5G wireless technology is related to the COVID-19 outbreak is stated. The alleged connection between 5G and Corona differs little from other counterfactual statements which are common on the internet. However, with the attacks on cell towers, it became a DW and had a far greater effect. Thus, we aim to study what happened in this specific case, and how it differs from other misinformation and conspiracy theories. Furthermore, we perform a manual examination of the data and give backgrounds regarding the timeline of the event, its main narrative, as well as the real-world consequences that occurred as a result of the event. To the best of our knowledge, we are the first to capture a complex DW in its entirety, even if only for a single OSN.

On January 21, 2020, the first tweet linking the COVID-19 outbreak in Wuhan, China, to the 5G wireless technology appeared on Twitter, stating:

> *"China is 5G now & working toward 6G. Wireless radiation is an immunosuppressor. Conincidence?"*

The tweet got little reaction, but in the following days, a series of similar tweets appeared. Their numbers grew steadily and about ten weeks later, in early April, a series of arson attacks hit wireless network equipment, mostly cell towers, in the UK as well as in Ireland, the Netherlands, Cyprus, and New Zealand.

It should be noted that the role of Facebook appears to have been significant[1]. However, we are unable to analyse this data due to the reasons described in Section 3.1. Instead, we focus on analyzing the connection between the prevalence and tone of tweets and articles on the one hand and the real-world events on the other. Our goal is to investigate the four questions listed below:

1. How did the 5G-Corona misinformation event start and where did it come from in January 2020?

2. How did it grow from relative obscurity to a widely discussed topic in late March 2020?

[1]Sky News: `https://bit.ly/3yKqMSM`

3. What is the connection to the serious real-world consequences observed in April 2020 and beyond?

4. Which general observations can we make from the structure of the specific event?

For the remainder of this thesis, we use the term *5G-Corona misinformation event* to refer to the entirety of all communication that falsely links 5G and COVID-19, i.e. all tweets, Facebook posts, news articles and videos. In addition, we consider all real-world events such as arson attacks that are presumably a direct consequence of this misinformation to be part of the event. Naturally, a causality cannot be proven in most cases, although it is reasonable to assume such a causality in some cases.

5.1 Scope of investigation

To limit our investigation to a manageable scope, we focus on events and communications in the first half of the year 2020. We only consider publicly available communications (mostly Twitter and Youtube, but not closed groups on Facebook), and we only count real-world events with a clear connection to these communications.

While arson attacks on telecommunications equipment have happened before, for example, in Germany between 2013 and late 2019 [2,3,4,5] which may have been motivated by opposition to wireless communications, we only investigate events that target 5G telecommunication equipment and that have happened during the COVID-19 pandemic as potentially related to the 5G-Corona DW.

5.2 The most Significant Events on a Timeline

It is not clear how the *5G-Corona misinformation event* started, even though multiple conflicting theories have been reported in the media. Our Twitter data collection, which is discussed in detail in Chapter 3, shows no tweets that connect 5G and COVID-19 before January 21, 2020. However, the early tweets do not suggest that the event originated on Twitter. Different origins have been reported in the media[6],

[2]Berliner Zeitung: `https://bit.ly/3baSlub`
[3]Süddeutsch Zeitung: `https://bit.ly/3eZoJ45`
[4]The Register: `https://bit.ly/3tv3Vqa`
[5]General Anzeiger: `https://bit.ly/3b9YbvT`
[6]Wired: `https://bit.ly/366jkUD`

although the idea must have existed before the listed events. There is substantial evidence that the idea grew out of an existing opposition [176] to 5G technology[7].

During the remainder of January, 685 such tweets and 1,081 retweets appeared, along with several videos on Youtube proposing similar ideas. Several of those videos appeared in channels that promote other misinformation and have comparatively large numbers of subscribers (i.e. between 50,000 and 300,000). In the meantime, the daily number of tweets on that topic grew slowly during January and most of February 2020.

With the growing number of COVID-19 cases in Europe, media attention grew sharply in March, peaking between March 20 and March 25. Shortly thereafter, multiple videos from UK-based sources promoting the 5G-Corona narrative appeared on Youtube and other video platforms, and were then spread widely via Twitter. Between March 25 and April 2, the number of tweets on the topic grew fourfold. Immediately thereafter, a series of arson attacks on mobile network infrastructure in the UK and other countries began and continued for several weeks.

Subsequently, on April 5, Youtube announced a ban on videos that claim a 5G-Corona connection and removed some of the most viewed such videos[8]. On April 22, Twitter announced that it would ban tweets and users that call for attacks on 5G infrastructure[9]. The company later added fact checking links to tweets mentioning 5G and Corona. Since then, the number of tweets on the topic has declined, although in January 2021 the topic resurfaced in South Africa.

5.3 Real-World Consequences

Attacks that happened during the COVID-19 pandemic as part of the 5G-Corona DW include arson and attacks or harassment of telecommunication technicians. On the weekend of Friday, April 3, 2020, at least ten arson attacks happened in the UK, New Zealand, and the Netherlands, and at least an additional 20 one week later, predominantly in the UK. A series of six such cases followed in Canada two weeks later. The total rose to 77 in the UK alone by May 7 [10] and 121 by July 2 [11]. Another four arson attacks happened on July 3 in Cyprus[12].

[7]FullFact: `https://bit.ly/2V1rNGB`
[8]Guardian: `https://bit.ly/3to3UVf`
[9]BBC: `https://bbc.in/3barrm1`
[10]Wired: `https://bit.ly/3esH9Lt`
[11]CNET: `https://cnet.co/3bsHzQj`
[12]APNEWS: `https://bit.ly/3nWkfPL`

Meanwhile, technicians that are perceived to be installing 5G infrastructure were harassed or attacked, with 273 reported incidents in the UK[13], most of them minor. A major event in this regard was the kidnapping of eight technicians in Peru[14] on June 10. Similar to attacks in the UK, the perpetrators clearly stated the perceived threat of 5G as the motivation for their actions. These statements, in consideration of the fact that the incident happened shortly after the 5G-Corona misinformation attained wider recognition, suggest a causal relationship between the two. Therefore, we speak of a DW here, i.e. we assume that the attacks are consequences of the misinformation, even though it is impossible to prove this while the perpetrators are not known.

While the wide reach the 5G-Corona misinformation attained in April 2020 seems to have vanished, it has been resurfacing occasionally around the world, and some attacks have occurred, including one in South Africa in early 2021[15] and a suspected case of arson in Canada by the end of March 2021[16]. A list of real-world events connected to the 5G-Corona DW can be found in Table 2 in the Appendix.

5.4 Manual Examination of Tweets related to the 5G Corona Connection

In order to investigate where the idea that 5G and the COVID-19 virus are connected originated, it is necessary to study the content of the messages. As part of this investigation, we labeled more than 10,000 relevant tweets manually[17]. This gave a thorough insight into the common ideas and underlying misinformation narratives in these tweets, which were then used to create the labeled datasets we published in [40, 41]. With the help of this dataset, it is possible to develop an automated detection method described in Section 6.3 (Chapter 6).

5.4.1 Early Twitter Posts

First, we investigate the earliest tweets that mention it. A total of these 104 COVID-19 related tweets authored by 75 accounts were published between January 21 and January 25, 2020. Among those, we identified 38 accounts that were insinuating

[13]cnet: `https://cnet.co/3oZmx0Z`
[14]France24: `https://bit.ly/3eu7LvO`
[15]Connecting Africa: `https://bit.ly/2Uk1TNQ`
[16]City News: `https://bit.ly/3yKlcjb`
[17]The dataset is published as part of the MediaEval Challange 2020 [37]

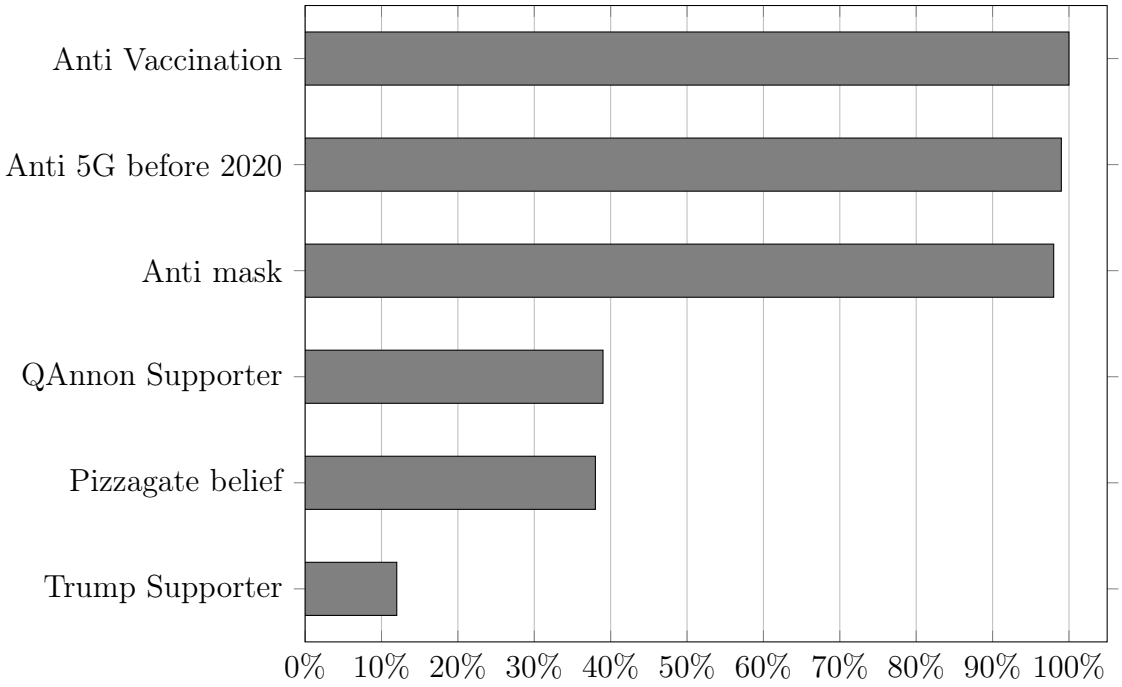

Figure 5-1: Ideas expressed by the first 37 accounts that spread 5G-Corona misinformation in January 2020.

a type of 5G-Corona connection. 28 of these presented themselves as belonging to individuals. One was a satirical website which we removed from the sample. 19 accounts reported their location as US, 9 as UK, AUS, NZ, or CAN, and 9 reported other European or Asian countries. We classified these 37 accounts with respect to six features, the first set of features being support for US president Donald Trump, the QAnon conspiracy theory, and the Pizzagate and Wayfair conspiracy theories [177, 178]. (We combined the Pizzagate and Wayfair narratives since they are very similar, with the latter being located in Europe.)

The second set of features contains statements against vaccination, against wearing face masks during the COVID-19 pandemic, and statements on health risks of 5G prior to 2020. Figure 5-1 shows the results, scaled by the number of followers of the accounts. Since the latter three features in Figure 5-1 are political in nature, it is expected that they have less global appeal than the first three which are health related. Among the US accounts, 11 out of 19 showed support for Trump (same for QAnon), while 27 out of 37 accounts voiced anti-vaccination and anti-5G statements. Additionally, such accounts had far more followers. Thus, there is a strong indication that the very early conversation of the 5G-Corona connection was dominated by people who oppose 5G rather than the political (far) right.

5.4.2 Defining Misinformation Narratives

A multitude of underlying narratives suggest a type of connection between 5G wireless technology and the COVID-19 pandemic, some of which are mutually exclusive. They share the belief that 5G is harmful in some way and that the technology should not be installed or existing installations should be dismantled. Therefore, we group all such narratives as 5G-Corona misinformation, and they all constitute a possible basis for attacks on mobile communication infrastructure.

Due to the nature of the misinformation, which contains wildly contrafactual statements and highly unlikely conspiracy theories, it is neither feasible nor desirable to classify them exhaustively. Instead, we give a rough overview over the main directions we encountered:

1. The *Immunosuppressor* narrative suggests that radiation from 5G antennas weakens the immune system, thereby making people highly susceptible to an otherwise harmless virus.

2. The *Cover-up* narrative states that radiation of 5G equipment is lethal, and that deaths attributed to the COVID-19 pandemic are actually caused by 5G. The pandemic is alleged to have been invented to hide this.

3. The *Mind Control Conspiracy* assumes that 5G allows some form of mind control, and that the coronavirus or a vaccine against it plants a receiver within humans.

4. Finally, there are some truly *Esoteric Conspiracy Theories* that see 5G and COVID-19 as means to prevent humans from reaching their full potential or higher selves. Narratives within this category diverge.

While these ideas may sound outlandish and contradict all established scientific consensus, it is clear that belief in these ideas could drive people to violent action. Our goal is to study how such ideas became widespread and how they may have incited people to action.

In addition to the opposition against 5G, the tweets also show another angle. In the context of 5G, the use of equipment manufactured by the Chinese companies Huawei and ZTE has been widely discussed as a potential cybersecurity risk in Western countries [179]. As a result, some countries have excluded Chinese vendors from supplying 5G infrastructure. In the US, the Trump administration passed the *Secure*

5G and Beyond Act[18] which was introduced on March 27, 2019 and signed on March 23, 2020. The UK later followed suit[19], reversing an earlier decision from January 28, 2020[20]. As a consequence, part of the opposition to 5G in the US stems from an anti-China sentiment of the political far right. One website from that spectrum[21] tweeted on April 17, 2020:

"If you like Corona in your country, you'll love Huawei in your home".

Here the connection between 5G and COVID-19 is the fact that both come from China without any direct interaction between them. Their narrative is essentially that COVID-19 proves that the Chinese Leadership is not trustworthy, and consequently Chinese technology should not be used. An early tweet from January 26, 2020 by a pro-Republican account states: "We can end the adoption of Chinese 5G spy technology in the West by convincing people it causes #coronarvirus outbreaks." It should be noted that this account has only 29 followers and that the tweet generated no reactions and is therefore unlikely to have created such a strategy.

5.5 Other News Sources

In this section, we investigate the role that other online sources have played in the 5G-Corona misinformation event. We thus study the role video platforms such as YouTube, as well as traditional online news websites play. The latter are accessed through GDELT. Doing so allows an automated large-scale analysis. On the other hand, manual analysis was employed for the videos. The videos show clear links to 5G opposition which had existed before January 2020.

5.5.1 Opposition to 5G before COVID-19 on YouTube and other Video Platforms

Concerns about the safety of wireless devices have existed for a long time, and even though the scientific consensus clearly finds low-powered personal devices to be safe for human use [180], rumors to the contrary seem to persist in the population. The 5G standard, which was introduced in late 2018, has drawn particularly vocal criticism which in some places has delayed or stopped 5G adoption[22]. Fringe YouTube channels

[18]Secure 5G an Byond Act: `https://bit.ly/3tAuFG2`
[19]CNN: `https://cnn.it/3f4mVa8`
[20]Deutsche Welle: `https://bit.ly/3o5rQLK`
[21]America First: `https://bit.ly/3y0r4V6`
[22]Financial Times: `https://on.ft.com/2SRujhZ`

provide a large amount of material discussing the alleged dangers of 5G. For instance, the Swiss KlagemauerTV[23] (more than 100,000 subscribers) released more than 50 videos discussing the alleged dangers of 5G in 2018-2019 (with about 10,000 views per video). The topic was also covered by the Russian channel RT[24] in early 2019.

Furthermore, a network of activists[25] organized protests against the adoption of 5G in 2019, i.e. before the COVID-19 pandemic. In January 2020, the group repeatedly called for protests against 5G, including a "Global Day of Protest" on January 25, with the call itself appearing shortly before the first tweets linking 5G and COVID-19 appeared. The British fact-checking organization FullFact suggest that the 5G-Corona connection grew out of this movement[26]. This suggestion is supported by the fact that the early tweets propose the *Immunosuppressor* and the *Cover-up* narratives, which are consistent with the idea that 5G is directly harmful. However, there are also numerous early tweets that point in a different direction. A *Wired* article[27] identifies a Belgian newspaper article from January 22, 2020 as the source of the 5G-Corona misinformation, but there are tweets that predate this article.

5.5.2 Videos promoting 5G-Corona Misinformation

Among the early tweets, only one had a relatively large number of 500 retweets, indicating the potential to reach a large audience. It came from an account that has almost 100,000 followers and was associated with a YouTube channel named *"Amazing Polly"* that was promoting *Mind Control Conspiracy* theories. The YouTube channel was removed in October 2020, but the videos continue to be available on BitChute and a dedicated website of the same name. While the narrative differs from other sources that oppose 5G, and the conversation is not linked to similar conversations on Twitter, considering the reach of this source it is likely that it was instrumental for the misinformation to spread in the early stages. Others who had opposed 5G earlier quickly followed with their own videos[28].

Following the restrictions on public life imposed by governments in Europe in late March 2020 to slow down infection rates, a Youtube video uploaded on March 28 quickly gained popularity among English speaking conspiracy theorists. The person on the recording claims to be the former

[23]KlagemauerTV: `https://bit.ly/3dQhLyv`
[24]Youtube RT: `https://bit.ly/3jK2jrD`
[25]Stop5gInternational: `https://bit.ly/2TJvErn`
[26]Fullfact: `https://bit.ly/3yt8K6t`
[27]Wired: `https://bit.ly/3yt8K6t`
[28]Youtube: `https://bit.ly/36iaZxu`

"[...] head of the largest business unit at Vodafone [...] between 2013 to 2015 [...]"

and back then responsible for the implementation of IoT (Internet of Things) and 5G technologies. The Guardian[29] later identified the speaker as evangelical pastor Jonathon James, who, according to The Guardian's sources at Vodafone, was hired for a sales position in the company in 2014 and left less than a year later.

On the recording, for over 30 minutes, the pastor spins a narrative starting with how the frequencies emitted by cell towers with 5G technology give people radiation poisoning making the human body produce the SARS-CoV-2 virus, continuing with how vaccines are dangerous and how a New World Order and Microsoft founder Bill Gates are behind a plan to

"[...] pave the way for the Antichrist [...]".

On April 2, four days after its initial upload, YouTube deleted the video after it had gained a lot of attention. At the same time, it was uploaded again and again to many different YouTube channels, and later to BitChute. We found over 1,000 tweets promoting the video. On the other hand, we found only 40 mentions of David Icke, who is more widely known and was considered an important factor in the attacks on the towers[30].

5.5.3 Commercial Interests

On May 28, 2020, several news sources reported that London Trading Standards was targeting a British company called BIOSHIELD DISTRIBUTION LTD. for selling USB sticks (see Figure 5-2) as a device that "protects against harmful radiation"[31], following a recommendation by a security company that had analyzed the product[32]. The event has had little impact on Twitter, but searching the tweets for the name of the company pointed to a YouTube channel called Connecting Consciousness (CC)[33] that spreads many different fringe ideas commonly classified as esoteric or conspiracy theories, including the idea that 5G is harmful to humans. The Youtube channel has around 50,000 viewers, with several hundred comments on each video.

[29]Guardian: `https://bit.ly/3i5oS9p`
[30]BBC: `https://bbc.in/2SRgwYJ`
[31]BBC: `https://bbc.in/34qRK3W`
[32]PenTestPartners: `https://bit.ly/3hZeArb`
[33]Youtube: `https://bit.ly/3hJnBmi`

Figure 5-2: the 5G Bioshield, which is advertised as "Full Spectrum Radiation Balancing Technology". At the top end of the USB stick, which is supposed to protect against all types of electromagnetic radiation, is clearly visible the crystal to which its special abilities are attributed.

The channel advertised "5G defense sticks" already in October 2019, with the CC website providing a link to a webshop, even though it claimed to be independent from the manufacturers of the device. On January 6, 2020, the channel announced delays in the fulfilment of the orders and on January 16, 2020, it hosted a presentation by the alleged inventors of the device. BIOSHIELD DISTRIBUTION LTD. was incorporated at Companies House on January 20, 2020.

Like similar products of this kind, the device seems to have been promoted at the "5G Apocalypse Event"[34] held in London in September 2019 (one of the two alleged inventors was a speaker at this event). While the CC channel did not suggest a direct connection between 5G and COVID-19, other speakers did so[35] on January 31, 2020.

The "5G Bioshield" was sold for a price of 350 USD through a professional looking website with a webshop[36] that was still functional in July 2020. Clearly, there is a

[34] 5G App. Event: `https://5gapocalypse.london/`
[35] Youtube: `https://www.youtube.com/watch?v=hjVxajUz09s`
[36] Shield webshop: `https://bit.ly/3jLqS7l`

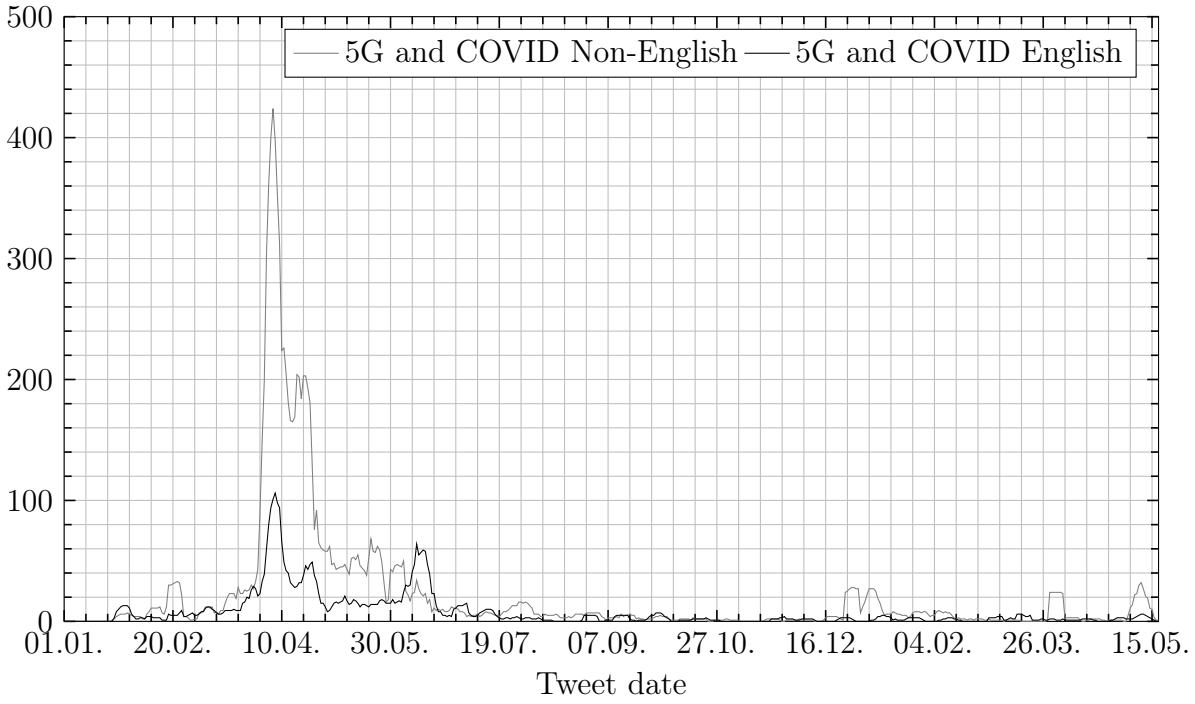

Figure 5-3: Weekly number of articles related to COVID-19 during the period of observation.

substantial commercial interest to spread misinformation for some people, and it is likely that this contributed to the DW. While many channels on YouTube pursue financial gain, it is important to distinguish between direct monetization of content, calling for donations, and the sale of questionable products. The former can easily be controlled by the video platforms, and the content is irrelevant as long as the videos are watched, even if viewers disagree with their content. It suffices that the viewers find it entertaining. For donations, one would expect that a higher level of agreement with the content is required. However, neither of the two depend on a threat narrative. On the other hand, selling so-called "protective equipment" requires creating the idea of a threat. Therefore, such actors have a direct financial motivation to spread threat narratives online.

5.5.4 Tracking the Misinformation Event on GDELT

In order to track 5G-Corona misinformation articles in online mass media sources, we use our high-performance system [35] for the analysis of GDELT. We run an analysis on a subset of news events and their mentions for the same time frame as for

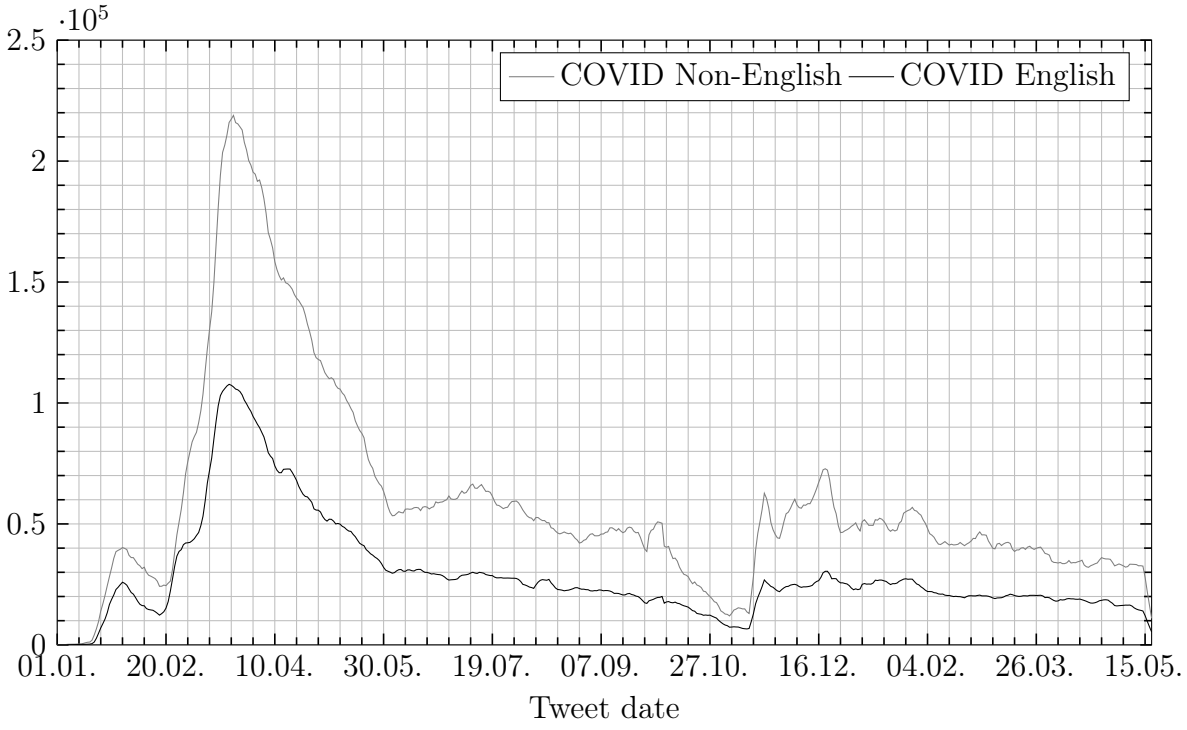

Figure 5-4: Weekly number of articles related to COVID-19 during the period of observation in 2020.

the Twitter data. Results are shown in Figure 5-3. Naturally, the COVID mentions peak in mid-March 2020, when the pandemic started to affect many countries besides China. After that, the number seems to stay relatively constant, interrupted only in early November 2020, i.e. the time of the US presidential election. Interestingly, this affects both English and non-English sources. In contrast, the 5G and COVID mentions peak in early April 2020, the time of the UK arson attacks, as well as in June 2020, when attacks happened in New Zealand. The non-English sources also peak in early April 2020.

Since the GDELT database does not contain news text bodies and contains titles for only a very limited set of news articles, we used an aggressive text mining approach. We converted every news publication record in plain text format, stripping out all non-numerical and non-alphabetical symbols. Next, we truncated all excessive word separators and performed a case-insensitive search for two sets of keywords that correspond to 5G and Corona topics. As a result, using the available keyword sources including full URL links and titles where they were available, we managed to get a relatively high number of relevant news articles i.e. event mentions. While the data mining was successful, we observed that the GDELT event mention database is lack-

ing good per-mention title and keywords coverage. Thus, it has a limited potential for online news sources, data mining, and analysis.

Among the early articles mentioning COVID-19 and 5G, most of the 5G-Corona related news and articles discuss and criticise the related conspiracy theories. However, the very first news article that does not counter-argument but promote direct relations between 5G and Corona was published[37] on February 20, 2020. To check the capabilities of the GDELT-based news chains detection, we performed a manual analysis of the news sources for seven sequential days starting with the very first conspiracy supporting news article. Table 1 in the Appendix depicts the results of the analysis. The interesting discovery here is that only one of 32 articles found via a keyword search for the selected week is actually about the 5G-Corona conspiracy. Other articles are focused on the Corona impact on the 5G technology development and deployment. The next yet more disappointing discovery is that a set of 14 of 32 articles (all have different URL paths) are just copy-paste of the same article published by Reuters. The same can be observed for another set of two and four articles. Also, two articles are missing from their publishing web-resources. All-in-all, this example illustrates the nuances of the GDELT news collection and used data storage and representation schemes. Lacking the full texts and most titles for these news articles makes it almost impossible to perform an effective and fine-grained event mentions search using keyword search and other, more sophisticated text analysis methods that are required for DW detection using traditional online news sources.

5.6 Action and Belief in Conspiracy Theories

While the motivations of the perpetrators of many real-world incidents are not known, the perpetrators of the harassment of workers in the UK[38] and the kidnapping of technicians in Peru[39] clearly stated that the alleged dangers of 5G were the motivation for their actions, and the intentions were clearly stated in some social media channels[40]. Thus, we can assume that 5G-Corona related conspiracy theories played a crucial role in causing real-world harm, which is the characteristic of a *DW*.

Psychological literature investigated the connection between paranoid thought and belief in conspiracy theories, but recent work highlighted differences between the two

[37] veterans today: `https://bit.ly/3wFKqxf`
[38] BBC: `https://bbc.in/3hqckbN`
[39] France24: `https://bit.ly/3wG3YS4`
[40] Sky News: `https://bit.ly/3yKqMSM`

constructs [181]. Endorsement of conspiracy theories is not constant over time and it depends on the situation. Conspiracy theories tend to be associated with major events, times of political instability, and collective threats, such as the 9/11 terror attacks, the death of Princess Diana or the assassination of J. F. Kennedy [182–185]. Having experienced stressful life events during the past six months is connected with belief in conspiracy theories [186]. Such events elicit feelings of uncertainty, and conspiracy theories can give people an explanation for a situation and its ultimate cause and hence reduce confusion. Furthermore, experimentally induced high-anxiety situations were associated with conspiracy thinking in research participants [187]. The COVID-19 pandemic is both a major and a stressful event. Moreover, it is associated with many uncertainties [188, 189] and these factors provide conspiracy theories with ideal conditions to flourish. Thus, a general willingness to endorse conspiracy theories exists. Thus, even the contradictory 5G-Corona narratives, rather than interfering with each other, can combine into a widespread conspiracy belief, such as the 5G-Corona connection observed here.

Recent work has investigated the effects of misinformation on behaviour during the COVID-19 pandemic in different countries [190, 191], as well as the handling of contradictory narratives [192]. One study found that *"5G COVID-19 conspiracy theories was positively correlated with state anger, which in turn, was associated with a greater justification of real-life and hypothetical violence in response to an alleged link between 5G mobile technology and COVID-19"* [193], thus providing strong support for the connection between the misinformation and the arson attacks.

Chapter 6

5G-Corona Connection Analysis

Contents

Having dealt with the qualitative analysis of the Cororona and 5G connection in Chapter 5, we now turn to the quantitative analysis. We perform an extensive analysis of 5G and COVID-19 related Twitter data in order to better understand the misinformation event.

6.1 Quantitative Analysis

In contrast to the qualitative analysis performed in Chapter 5, we now work with the entire dataset. The data collection was performed using the methods proposed in Section 3.3, targeting COVID-19 related keywords such as *Corona*, *Coronavirus*, and *COVID*.

Next, we filtered for those that mention 5G in any conceivable spelling, such as *" 5G "* and *" 5 G "*. We did not remove the whitespaces because removing them produced too many false positives completely unrelated to 5G. We then restored as much of the Twitter threads as possible. Restoring Twitter threads using the FACT

framework is only possible in an upward manner, meaning that for each reply that contains a keyword we can only find the tweets or the reply it replied to. This limitation leaves us with only the threats above the replies we filtered based on keywords. The result is a set of tweets, retweets, and replies that not only contain statuses that mention 5G, but also those to which 5G containing statuses are replies. For simplicity, we again do not distinguish between tweets, quotes, and replies using the term tweet for the remainder of this chapter. The data collection started on January 17, 2020, and the first qualifying tweet was found on January 21. The primary data collection phase ended on May 17, 2020, but we continued into a second phase, ending on March 31, 2021. We divide the results into five overlapping datasets:

1. *Dataset 1* – Comprises all 5G tweets in English from the first phase.

2. *Dataset 2* – A subset of *Dataset 1*. Contains only tweets from accounts that self-report an identifiable location.

3. *Dataset 3* – Comprises all 5G tweets in English from the first phase that contain a Twitter-reported location.

4. *Dataset 4* – Comprises all 5G tweets in English from the first and second phases that contain a Twitter-reported location.

5. *Dataset 5* – Comprises all 5G tweets in English from the first and second phase.

Table 6.1: Number of tweets by type in *Dataset 1*.

Collection begin	January 17, 2020
Collection end	May 15, 2020
Tweets	177498
Retweets	351848
Quotes	24588
Replies	216474
Quotes replies	14410
Quotes retweets	16697
Users	413885
Deleted users	11788

6.1.1 Development over Time

Due to the problems Twitter had with flagging 5G-Corona misinformation tweets[1], we decided not to rely primarily on an automated analysis. Their shortness and the fact that many tweets convey misinformation by insinuation makes automated labeling difficult. For example, a tweet from January states:

> "Wuhan province is the source of an outbreak of 'corona' virus, interestingly enough, it also has around 10,000 operational 5G base stations. Coincidence...?".

In this case, the word "coincidence" turns the factual statement into misinformation because it insinuates a connection that does not exist. Even when labeling manually, in some cases the intention was only understandable by considering the source, e.g. in case of satire.

Table 6.2: Classification of Tweets in Early 2020

Category	February	March
a) Unrelated to 5G-Corona connection	87	88
b) Mention 5G-Corona connection	71	97
c) Belief in 5G-Corona connection	120	117
d) Belief in 5G-Corona conspiracy	106	47
e) Could not determine	15	51
Sum	400	400

We sampled tweets from *Dataset 1* in two groups of 400 each, one from February and one from March 2020. We distinguish between the following five categories:

a) unrelated to the 5G-Corona connection (e.g. "Will COVID-19 affect the 5G rollout?"),

b) mentioning the 5G-Corona connection without any indication that they might consider it to be true (e.g. "Some people believe that 5G causes Corona. People are stupid"),

c) believing that a connection between 5G and corona exists or could exist (e.g. "Wuhan was the first city to deploy 5G. Coincidence ?"),

d) believing that there is a connection between 5G and corona, and that authorities know about this connection and somehow hide it (e.g. "Wuhan being the

[1]the Verge: `https://bit.ly/3yMWG0E`

the epicentre of the New World Order testing ground China is awash with 5G radiation which when it affects a subjects presents the same symptoms as the Coronavirus/common cold except it also kills you"),

e) tweets which we could not categorize due to language barriers or because they did not contain a full sentence.

Results are given in Table 6.2. The numbers support the assumption that the misinformation is initially spread in a relatively closed circle. Over time, it becomes more widespread and outsiders hear about it without believing (Category b). As is to be expected, the fraction of people who believe in a full-blown conspiracy (Category d) goes down. The reason for this is most likely that the belief in conspiracy theory is more driven by the general willingness to assume that authorities would act in this manner, and less by concrete pieces of information (see Section 5.6 for an in depth discussion). On the other hand, as time progresses, the fraction of people willing to entertain the idea that 5G is linked to COVID-19 (Category c) is relatively constant. Since the total number of qualifying tweets goes up, this means that the misinformation was primarily spread by growth of this group. The rising numbers in Category e) between February and March are likely due to the fact that other countries started to discuss the idea as well.

While we cannot determine the exact motivations of the people who committed arson attacks against cell towers, it is interesting to note that these mostly happened after a larger group of people started to apparently believe in the connection. Those who believed in a conspiracy in February 2020 could have acted earlier but did not. We conjecture that the real-world harm of the DW stems more from confused individuals rather than those who hold deeply entrenched conspiracy beliefs. Naturally, this analysis is limited by the fact that we can only study a small fraction of the relevant tweets.

6.1.2 Mapping Tweet Locations

Naturally, it is interesting to determine in which countries the concentration of misinformation tweets is the highest. Thus, we need to obtain the location of tweets, or an approximation thereof. Twitter offers multiple ways of doing so. One is the tweet location, while the other is the self-reported account location. However, such information has to be considered with caution. There is no verification of the self-reported location, and location reporting by Twitter is turned off by default. Furthermore,

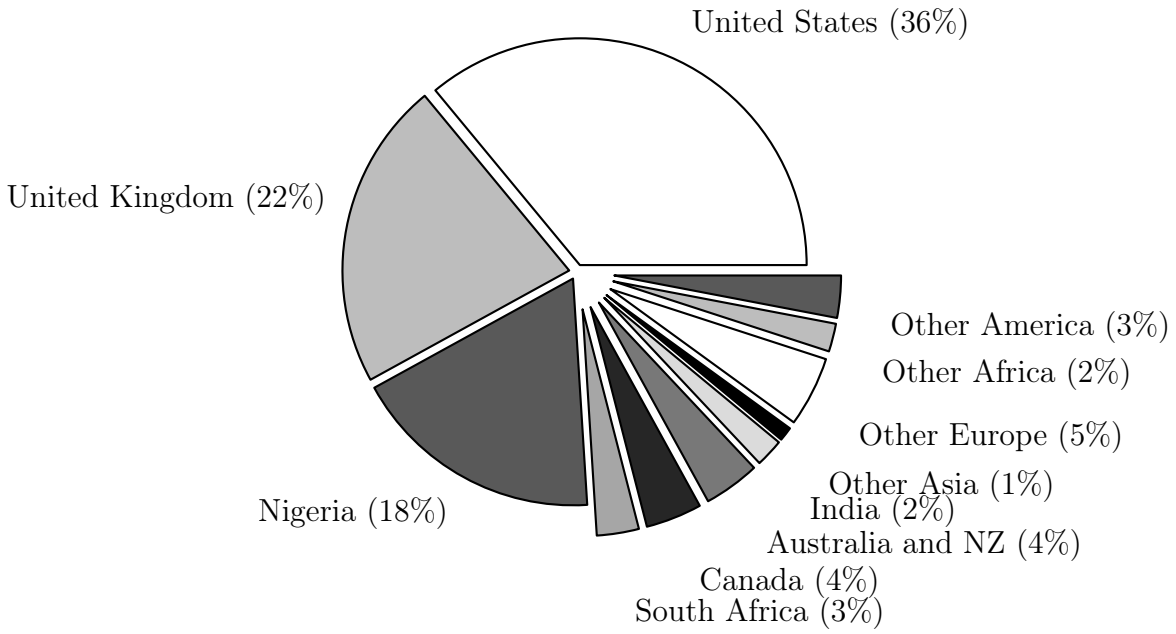

Figure 6-1: Prevalence of countries in Dataset 2: 5G-Corona tweets with self-reported location from January to May 2020.

accounts spreading misinformation might be less likely than others to share their location, which may distort the results.

From the period of January to May 2020, we obtained close to 250,000 tweets with self-reported account locations, out of which about 220,000 could be decoded. The distribution is shown in Figure 6-1. We would expect to find a large number of 5G tweets in the UK where the misinformation had the highest impact, and the large number of UK tweets seems to confirm this. However, without comparing these numbers to the population size and Twitter usage of a country, they contain little information.

For the Twitter-reported locations, which are shown in Figure 6-2, the situation is quite similar, despite the fact that this dataset comprises only 6,500 tweets and only contains English tweets. The most striking difference is between Nigeria, whose fraction among the self-reported locations is almost twice that of the Twitter-reported locations, and Europe without the UK, where it is the other way around. However, for US and UK the fractions are very similar. The same is true for most other listed regions and countries.

Naturally, we would expect to find the highest concentration of such tweets in places where the misinformation was most effective. Thus, in order to obtain such a

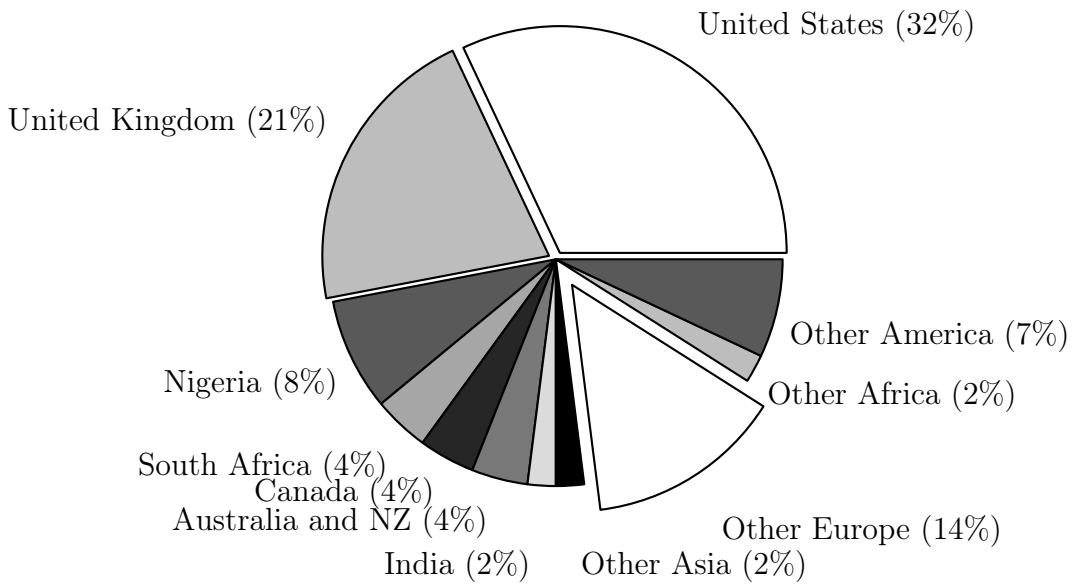

Figure 6-2: Prevalence 5G-Corona tweets by country in Dataset 3: English Corona-related tweets with Twitter-reported location from January to May 2020.

concentration of 5G tweets, we divide their number by the total number of Corona-related tweets in this dataset, which is 10.2 million. Results are shown in Figure 6-3. The rate is given as 5G-Corona tweets per 100,000 Corona-related tweets. The results clearly show a high concentration of 5G and COVID-19 related tweets in Nigeria, UK, and South Africa. At the same time, similar discussions in the US focused more on the politization of the coronavirus and its labeling as a "hoax" by Donald Trump[2] occured with a frequency of 290 per 100,000 Covid-related tweets.

6.1.3 Later Development in the Different Regions

We analyze an extended set of Twitter data and again, the basis of this set is the results of the search API for Corona/COVID-19 related terms from which occurrences of 5G are selected. These data overlap with the reported location data used in the previous section. The only difference is that here we include all available tweets until March 31, 2021. As a result, we obtain 9,376 5G tweets from a set of 16.8 million tweets with reported locations. While this is much smaller than the total number of available 5G tweets, it is sufficient to map trends, and we can assume that the reported locations are reliable.

To make the numbers from different world regions comparable, we again report

[2]NBC: `https://nbcnews.to/3ew087Z`

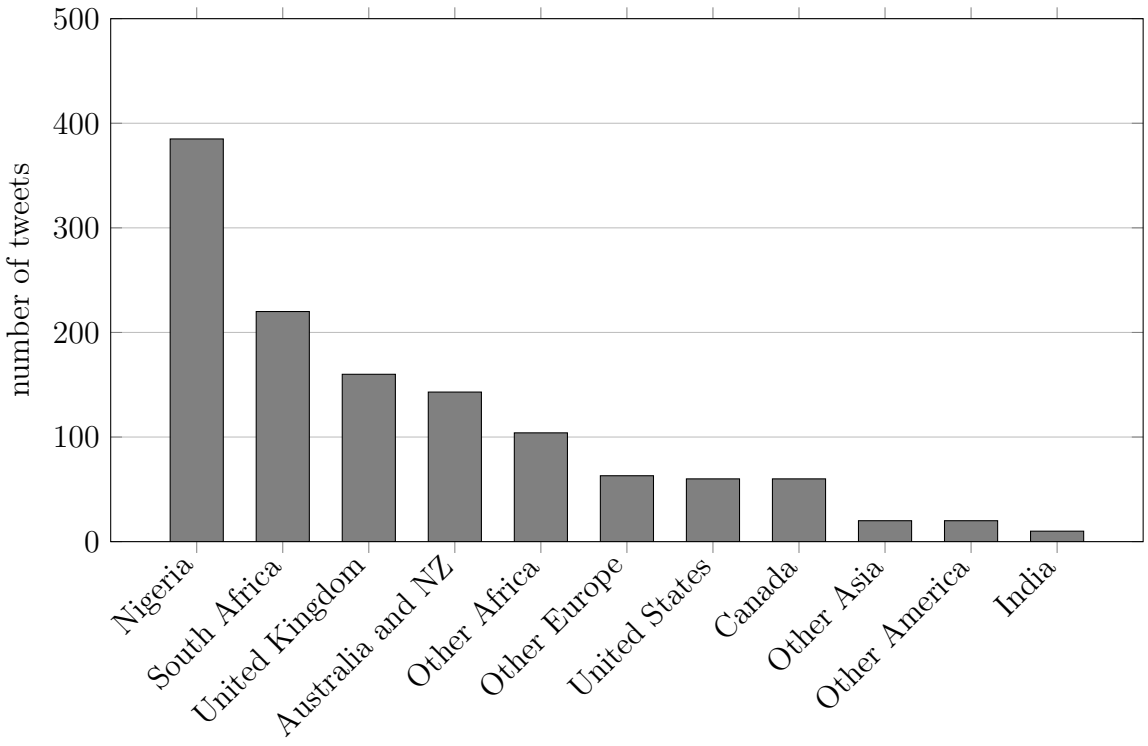

Figure 6-3: Concentration of 5G-Corona tweets in Dataset 3: English Corona-related tweets with Twitter-reported location from January to May 2020.

the number of 5G and Corona-related tweets per 100,000 Corona-related tweets. To improve readability, we split the results into two charts. Figure 6-4 shows our main countries of interest, and Figure 6-5 shows the remaining countries aggregated by continent. Again, the term "Other" (e.g. America) is to be read as e.g. America without the countries that were treated separately, i.e. USA and Canada. The numbers are obtained by summing up to ten major countries per region. Smaller countries were omitted since the number of relevant tweets in this dataset is very small.

The initial peak in April 2020 is dominated by Nigeria, UK, and South Africa, where South Africa is the first country that shows a massive increase in 5G tweets. All other countries, including the UK, seem to follow. While the number of qualifying tweets in the US is high, the massive use of Twitter implies that the fraction of 5G tweets is comparatively low. While it is not possible to prove from these numbers alone, it seems that 5G-Corona misinformation events are not primarily driven by Twitter activity in the US.

Interestingly, Australia and New Zealand also have a high number of 5G tweets

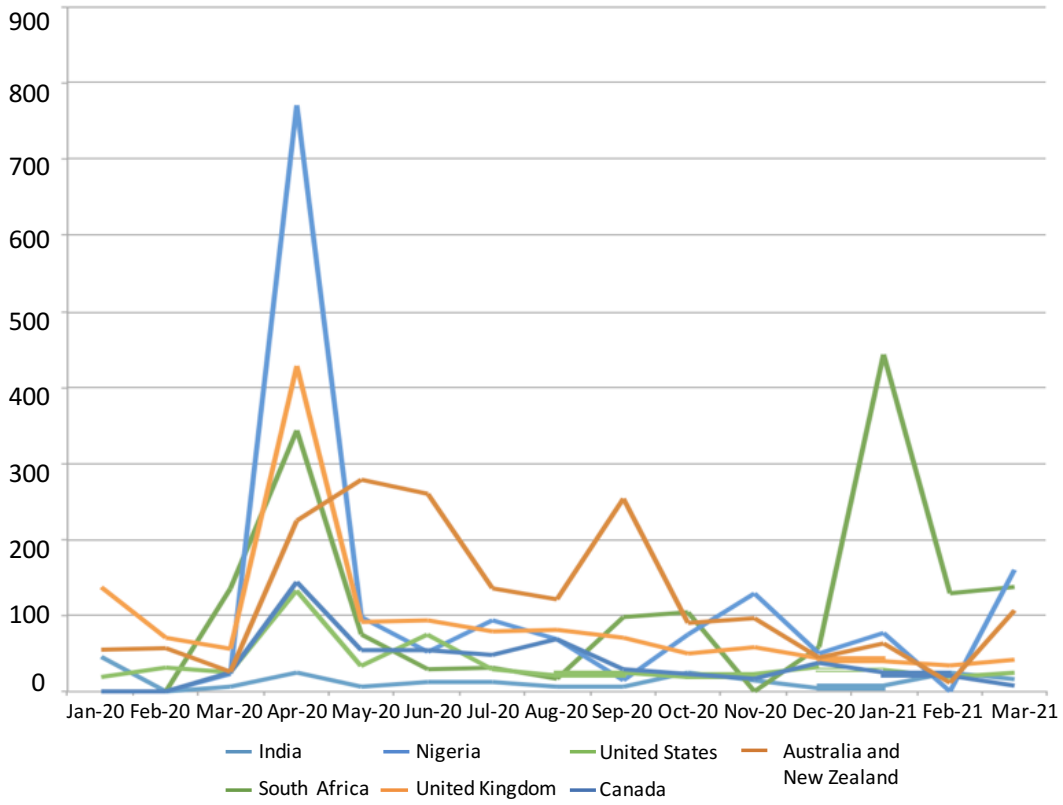

Figure 6-4: Concentration of 5G-Corona tweets over time in selected countries in Dataset 4: English Corona-related tweets with Twitter-reported location from January 2020 to March 2021.

and 14 reported tower attacks in New Zealand alone by mid-May[3], whereas Canada has four reported tower attacks in the beginning of May[4] but an overall low number of 5G tweets. Furthermore, there is a second peak for South Africa in January 2021, which coincides with the resurfacing of 5G-Corona misinformation[5] and a subsequent tower attack[6].

For the remaining countries, which we aggregated by continent, we observe that Africa has a massive increase in 5G tweets in April 2020, but also in February 2021, which again seems to follow the spike in South Africa in January. Among the other continents, Europe has consistently more 5G tweets than America, and Asia has the fewest. It seems that the 5G-Corona misinformation plays no noticeable role there, at least not according to Twitter data. Thus, while the idea may not have originated

[3]RNZ: `https://bit.ly/3vBjlLy`
[4]Montreal News: `https://bit.ly/34FinlP`
[5]Timeslive: `https://bit.ly/3f8tEzz`
[6]Connect Africa: `https://bit.ly/3faAnci`

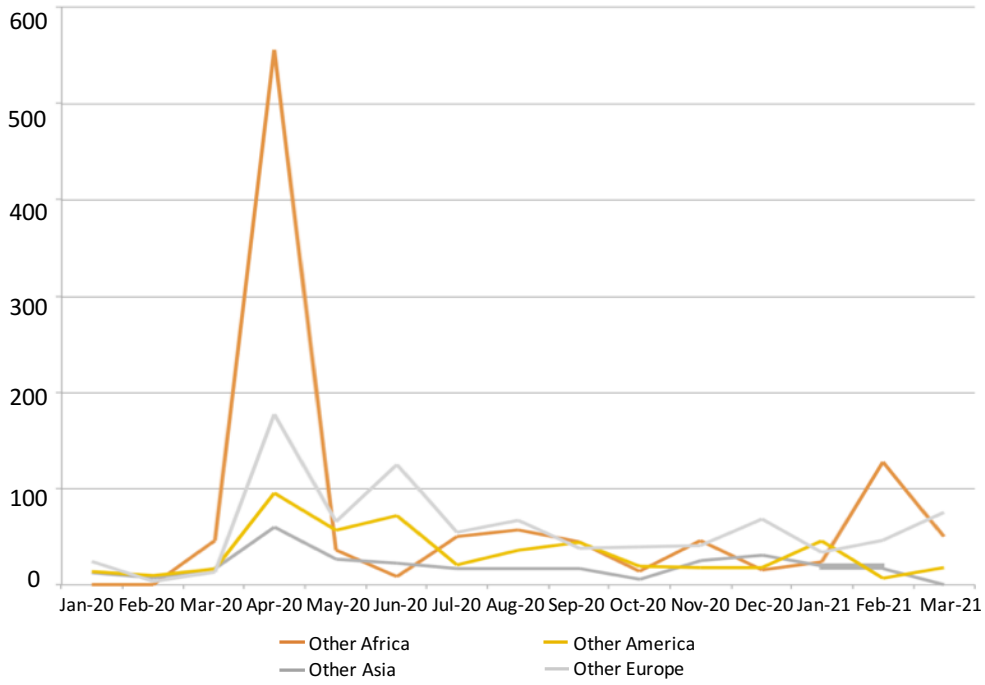

Figure 6-5: Concentration of 5G-Corona tweets over time in aggregated continents in Dataset 4: English Corona-related tweets with Twitter-reported location from January 2020 to March 2021.

in Africa, it seems that it has taken hold there and resurfaced after more than half a year.

6.2 Sentiment analysis

During the qualitative analysis, we found a large number of tweets that used expletives and strong language to state that 5G is not the cause of COVID-19. On the other hand, we noticed that many tweets that promote the misinformation adopted a neutral, scientific-sounding tone. This is also true for the videos we discuss in Chapter 5.5. Consequently, we use automatic sentiment analysis in order to investigate whether this observation can be verified quantitatively.

We use the Valence Aware Dictionary and sentiment Reasoner (VADER) [194] sentiment analysis software. Results are shown in Table 6.3. We split the table to show the analysis for the human annotated tweets separately. Note that retweets, replies, and tweets that promote other conspiracies are not part of the analysis. VADER computes the fraction of *positive*, *negative* and *neutral* words. It also determines an aggregated *composite* score which reflects how positive or negative each individual

Table 6.3: Results of the sentiment analysis.

	Annotated		Automatic		COVID
	5G-C	Other	5G-C	Other	Baseline
Positive	0.063	0.077	0.066	0.076	0.070
Neutral	0.828	0.808	0.830	0.817	0.854
Negative	0.107	0.114	0.102	0.106	0.075
Composite	-0.188	-0.126	-0.151	-0.116	-0.019
Stdev positive	0.072	0.084	0.084	0.077	0.095
Stdev negative	0.097	0.107	0.101	0.098	0.101
Stdev composite	0.499	0.529	0.517	0.519	0.428
Tweets	1876	6572	37360	150280	16567036

word is. These scores are shown in Table 6.3. We also list the number of tweets in each class, as well as the standard deviation of the scores. As a result of the large number of observations, the results are highly significant ($p < 0.01$). The table is split between 5G-Corona (5G-C) misinformation tweets, i.e. those that imply a connection, and tweets that do not (Other). We give the numbers for the manually annotated tweets, as well as for the automatically classified ones. In addition, we provide a baseline for COVID-related Tweets from April 2020 that were not filtered for the 5G keyword.

The numbers partially support our initial assumption. The 5G-Corona tweets use fewer negative words. They also use less positive words and are thus more neutral in tone. However, when considering the composite index, we see that the overall sentiment in the 5G-Corona tweets is more negative. These observations apply to both datasets. Furthermore, all 5G tweets are substantially more negative than the baseline, where positive and negative words are approximately balanced.

To understand this result, consider that the non-misinformation and baseline sets contain positive statements such as:

> *Huawei installed 5G in Chinese hospital in Wuhan to help fight Corona virus.*

(January 24, 2020). On the other hand, while the tone of many misinformation tweets is more neutral, they all deal with infection and death caused by 5G in some fashion. Thus, for a more detailed comparison of the tone, it would be necessary to further subdivide the non-misinformation 5G tweets into those that mention the 5G-Corona connection, and those that are unrelated to it. While we did that in the manual analysis presented in Table 6.2, the data that was annotated for training the auto-

mated system did not contain this distinction, and therefore we cannot perform this analysis here. However, with the current rapid advances in natural language processing [195, 196], it may become possible to perform such an analysis automatically in the near future.

The most important lesson from this analysis is that 5G-Corona misinformation does not primarily target people at a direct emotional level, and a substantial part of it is presented in a non-emotional and seemingly scientific manner. This also means that while automated approaches for recognizing misinformation by looking for highly emotional tweets may sometimes work, they fail in the case of 5G-Corona and similar misinformation events. Such approaches are therefore not suitable as a general detection strategy.

6.3 Automated Conspiracy Detection

Naturally, simply observing the number of times 5G is mentioned is not a sufficient analysis. As we have seen, the largest number of mentions was in April 2020, after the arson attacks occurred. However, it would be interesting to understand which part of the tweets that mention 5G and COVID-19 actually insinuate a causal relationship and thus spread misinformation. Initial sampling suggested that there are several keywords that are typically used in tweets that spread misinformation (e.g. "immunosupressor"), but other tweets that mock or parody such tweets make use of these words as well. Thus, even for human readers, it is not always easy to discern the intention of a tweet.

Consequently, the task is very difficult for automated systems. On the other hand, the large number of tweets makes it impossible to manually evaluate the tweet text, even by sampling. Obtaining daily samples with 95% confidence would require the manual evaluation of more than 50,000 tweets. Thus, it is necessary to look for automated systems that are as accurate as possible. To this end, a dataset containing 10,000 manually annotated English tweets from *Dataset 1* is published as part of the MediaEval Challange 2020 [37]. Tweets were classified as either

- *spreading 5G-Corona misinformation*,
- *spreading other conspiracy theories*,
- *not spreading conspiracy theories*.

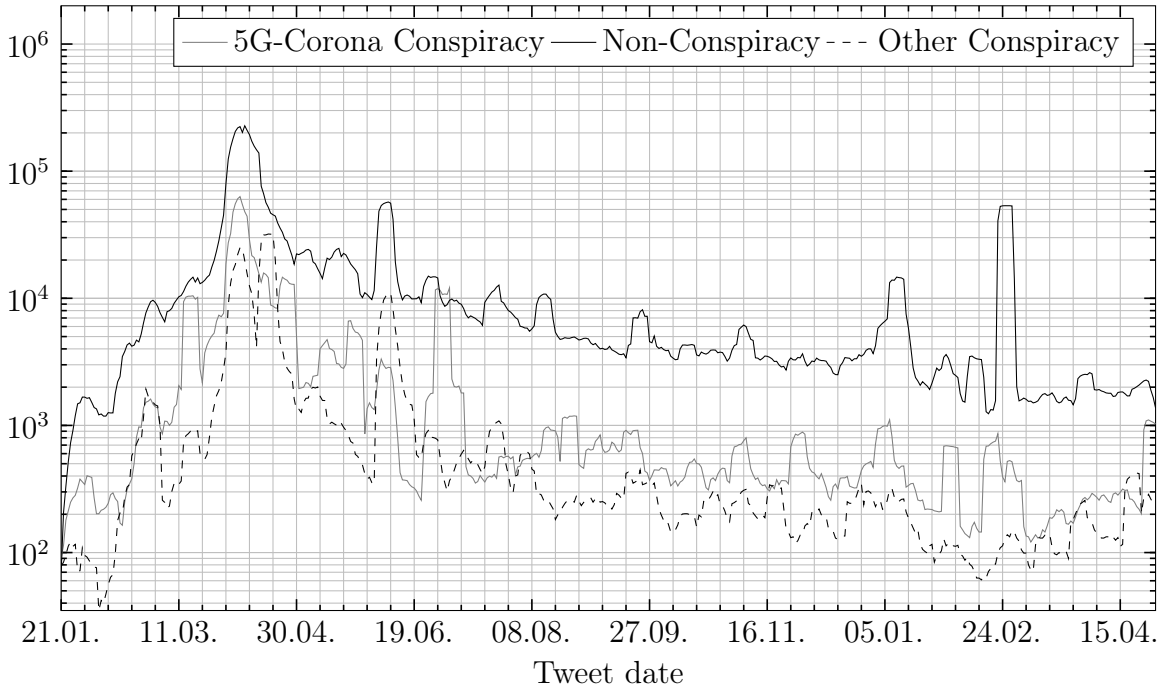

Figure 6-6: Weekly number of tweets mentioning COVID-19 and 5G, classified automatically into misinformation and non-misinformation categories.

Some effort was made to develop and test automatic classifiers for that dataset [39]. We used a Multilayer Perceptron Classifier [197]. The classifiers were found to be fairly accurate, and we use the best classifiers identified there on the tweets in *Dataset 5*. Note that *fairly accurate* means that while the false positive/negative rate is high, the accuracy is good enough to quantify the number of misinformation tweets.

Results are given in Figure 6-6. Clearly, the *non-conspiracy* tweets far outnumber the misinformation. To make the change over time more visible, we also show the fraction of the total that each class represents in Figure 6-7. We clearly observe a spike in 5G-Corona misinformation tweets that peaks on March 14, 2020, shortly before successful misinformation videos were released (see Section 5.5.2). Other spikes shortly predate the later UK attacks in May, and those in Cyprus in June.

From the information presented here, it is quite clear that there is no single tweet dataset that popularized the 5G-Corona conspiracy ideas. We did not find extremely influential misinformation tweets that triggered an significant information cascade directly. Instead, it appears that a constant stream of misinformation tweets pushed the idea forward and, after a tipping point was reached, mass propagation happened on other platforms, which we will discuss in Section 5.5. The results show that long

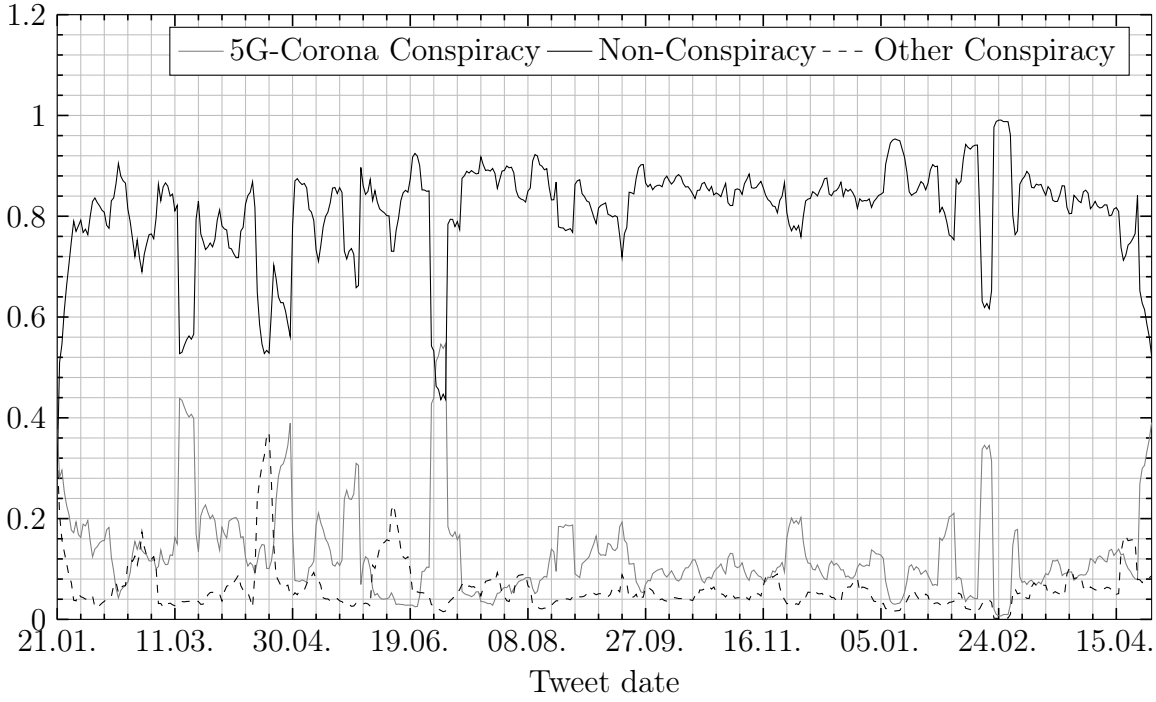

Figure 6-7: Weekly percentage of misinformation and non-misinformation tweets among the tweets mentioning COVID-19 and 5G.

after the initial wave of misinformation disappeared, the topic resurfaced wit a large concentration of tweets every few months. Note that the number of *other conspiracy theory* tweets is relatively low since these are tweets that contain the term 5G. Thus, the figures do not describe the ratio between 5G-Corona misinformation and other COVID-19 related misinformation.

Chapter 7

Conclusion

In this work, we have designed, evaluated, and implemented methods for acquiring large amounts of data from OSNs and Twitter in particular. The entirety of these methods is summarized and published under the umbrella of the FACT framework. FACT implements a crowd-based approach, which multiplies the amount of data that can be retrieved by making use of donated user data contingents that are distributed over a network of proxy servers which mirror Twitter's API for internal purposes. The operation of FACT, for the first time, made it possible to collect volumes of social network data in the order of magnitudes that allows to capture and investigate a DW, namely the Corona and 5G misinformation event in its entirety.

In order to identify the overall amount of tweets attributed to the DW presented, we manually labeled a dataset of 10,000 tweets and extracted their information cascades. Based on this dataset we organised the MediaEval 2020 *FakeNews: Corona virus and 5G conspiracy*[1] task to invite researchers from the fields of Natural Language Processing and Geometric Deep Learning to contribute with ideas and approaches for misinformation classification. Given the MediaEval results as a base for inspiration, including the evaluation of a variety of different classification approaches, we developed the proposed automatic conspiracy detection that allows identifying several hundred thousand 5G-Misinformation tweets out of a dataset with more than eight billion corona-related tweets. Our results show that classification based on information cascades delivers poor results compared to text-based classification methods. Since previous work [198] including similar approaches applied to different data sets show more promising results, we must assume that the number or size of the provided cascades is not sufficient. While text-based classification promises better results in

[1]MediaEval: `https://bit.ly/3CrAYBh`

general, the transformer-based machine learning techniques, including the well-known BERT model [199], stand out with particularly good results. Especially pre-trained variants specialized on corona fake news [200] seem to be very promising.

The analysis of the entire dataset obtained using the automated extraction has led to a number of findings. Thus, it has become apparent that DWs are not born or invented out of nowhere. They rather have their origin in a multitude of already existing ideas that are reinterpreted and recombined in the light of a new situation. In the presented case study, the corona and 5G misinformation event, for example, we find a multitude of different, partially contradictory narratives (see Chapter 5.4.2).

Contrary to general expectations, the existence of contradictory narratives does not limit the extent of dissemination for a given DW. In fact, the opposite appears to be the case, the existence of contradictory narratives leads to increased belief and extended spread of DWs. One of the many examples of such contradictory narratives are the *Immunosuppressor* narrative introduced in Chapter 5.4.2 suggesting that 5G radiation weakens the human immune system and therefore makes people more susceptible to the Corona virus, while the *Mind Control Conspiracy* suggests that 5G technology is used to influence the way people think. Here, even though both of these narratives agree in their core statements, namely, that 5G technology is causally related to the Corona pandemic, they provide contradictory reasons for the cause. Although further research is required to investigate this phenomenon in detail, it seems that the common belief in a superordinate narrative is a mechanism strong enough to encourage people regardless of their political views or beliefs to unite in order to cause severe real-world harm, while explanations for the emergence of such a narrative as well as a congruent embedding into the framework of reality have no importance at all for the the resultant success of a DW.

Furthermore, we could show that the dissemination of DWs is not limited to one medium. Despite the strong presence of 5G-Corona misinformation on Twitter, the rapid spread is the result of an interplay between different social media platforms, where traditional news websites only play a subordinate role and contribute little in the spread of misinformation, while the distribution of videos with misinformation content on video platforms such as YouTube is crucial for increasing the reach of misinformation. This insight seems to be essential for the development of strategies to contain DWs since their detection and containment in an early stage, i.e., before resulting in harmful consequences, must be a joint effort including news agencies and OSNs providers. In this context, it seems that there is still a long way to go even regarding the detection mechanisms implemented by single OSN. A subsequent

analysis of the 5G Corona tweets shows that only a fraction of the tweets including potentially harmful content have been removed by Twitter, while a not inconsiderable proportion contains keywords such as "5G" or "Covid" but are not related to the misinformation event at all were also subject to deletion. Although we have not done any further research in this direction, nor do we have insight into Twitter's internal misinformation detection, the previous example shows that the implemented prevention measures do not work satisfactorily and seem to be based on a simple keyword search.

In previous work, the concept of information cascade has often been equated with a single tweet and its retweets [201]. However, we observe that the number of particularly influential tweets, i.e., those that produce cascades with large node sets, is rather small. There is certainly no single tweet that is solely responsible for the spread of the DW. In fact, the opposite is true, we observe that the information did not only spread through a large number of tweets with comparatively small influence but also in parallel with other social media channels. Therefore, we propose to refrain from the consideration of single information cascades and to understand information cascades as a conglomerate of a multitude of individual messages that may contain even conflicting narratives. One could argue that this should be classified as a set of individual information cascades instead, but all these messages strengthened the core idea of the DW, i.e. that 5G and COVID-19 are linked.

Furthermore, based on the number of interactions between pairs of users and their reaction time, we introduced a function that calculates the *connectivity*. Applied to a dataset of more than one billion tweets and retweets, we were able to reconstruct and study the environment in which the DWs spreads.

The most important conclusion this thesis comes to is that it is not possible to identify a DW just by how it spreads. We used structural, text-based and sentiment analysis to investigate and understand DWs but despite the claims that can be found in the literature, suggesting that the analysis of spreading information alone is sufficient to predictively label spreading information as probably leading to a DW, we could show that this is not the case. Precisely because DWs cannot be detected in their stage of development, it is impossible to understand the spread of a DW in retrospect unless all data is retained in a form that allows its analysis. FACT allows researchers to do this.

Bibliography

[1] M. Settembre, "Towards a hyper-connected world," in *2012 15th International Telecommunications Network Strategy and Planning Symposium (NETWORKS)*, pp. 1–5, IEEE, 2012.

[2] A. Van den Bosch, T. Bogers, and M. De Kunder, "Estimating search engine index size variability: a 9-year longitudinal study," *Scientometrics*, vol. 107, no. 2, pp. 839–856, 2016.

[3] H. Webb, P. Burnap, R. Procter, O. Rana, B. C. Stahl, M. Williams, W. Housley, A. Edwards, and M. Jirotka, "Digital wildfires: Propagation, verification, regulation, and responsible innovation," *ACM Transactions on Information Systems (TOIS)*, vol. 34, no. 3, pp. 1–23, 2016.

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[5] S. Glezos, *The politics of speed: Capitalism, the state and war in an accelerating world*. Routledge, 2013.

[6] C. Spivak, "The fact-checking explosion: In a bitter political landscape marked by rampant allegations of questionable credibility, more and more news outlets are launching truth-squad operations," *American Journalism Review*, vol. 32, no. 4, pp. 38–44, 2010.

[7] M. Schäfer, "Science journalism and fact checking," *Journal of science communication*, vol. 10, no. 4, p. C02, 2011.

[8] W. E. Forum, "Digital wildfires in a hyperconnected world," in *Global Risks 2013 Eighth Edition*, pp. 23–28, 2013.

[9] P. R. Berthon, L. F. Pitt, K. Plangger, and D. Shapiro, "Marketing meets web 2.0, social media, and creative consumers: Implications for international marketing strategy," *Business horizons*, vol. 55, no. 3, pp. 261–271, 2012.

[10] M. M. Davis, J. C. Spohrer, and P. P. Maglio, "Guest editorial: How technology is changing the design and delivery of services," *Operations Management Research*, vol. 4, no. 1-2, pp. 1–5, 2011.

[11] E. C. Pease, "'skyful of lies' and black swans: The new tyranny of shifting information power in crises," *Journalism and Mass Communication Quarterly*, vol. 87, no. 3/4, p. 680, 2010.

[12] O. R. Rodriguez, "Mexico tweets cause massive shootout panic," *Retrieved September*, vol. 20, p. 2013, 2012.

[13] J. Borger, "Israel and hamas deploy twitter feeds in media war," *The Guardian. Accessed January*, vol. 7, p. 2013, 2012.

[14] A. Bruns, S. Harrington, and E. Hurcombe, "<? covid19?>'corona? 5g? or both?': the dynamics of covid-19/5g conspiracy theories on facebook," *Media International Australia*, vol. 177, no. 1, pp. 12–29, 2020.

[15] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, 2011.

[16] C. T. Nguyen, "Echo chambers and epistemic bubbles," *Episteme*, vol. 17, no. 2, p. 141–161, 2020.

[17] S. Flaxman, S. Goel, and J. M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public opinion quarterly*, vol. 80, no. S1, pp. 298–320, 2016.

[18] A. Stevenson, "Facebook admits it was used to incite violence in myanmar," *The New York Times*, vol. 6, 2018.

[19] C. Newton, "The trauma floor," Feb 2019.

[20] W. Merrin and A. Hoskins, "Tweet fast and kill things: digital war," *Digital War*, pp. 1–10, 2020.

[21] Á. Figueira and L. Oliveira, "The current state of fake news: challenges and opportunities," *Procedia Computer Science*, vol. 121, pp. 817–825, 2017.

[22] M. Kreil, "Social bots, fake news und filterblasen."

[23] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

[24] R. Li and A. Suh, "Factors influencing information credibility on social media platforms: Evidence from facebook pages," *Procedia computer science*, vol. 72, pp. 314–328, 2015.

[25] D. Gayo-Avello, P. T. Metaxas, E. Mustafaraj, M. Strohmaier, H. Schoen, P. Gloor, C. Castillo, M. Mendoza, and B. Poblete, "Predicting information credibility in time-sensitive social media," *Internet Research*, 2013.

[26] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, pp. 675–684, 2011.

[27] Z. Wei, J. Chen, W. Gao, B. Li, L. Zhou, Y. He, and K.-F. Wong, "An empirical study on uncertainty identification in social media context," in *Social Media Content Analysis: Natural Language Processing and Beyond*, pp. 79–88, World Scientific, 2018.

[28] Y. Chen, N. K. Conroy, and V. L. Rubin, "News in an online world: The need for an "automatic crap detector"," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[29] V. L. Rubin, N. J. Conroy, and Y. Chen, "Towards news verification: Deception detection methods for news discourse," in *Hawaii International Conference on System Sciences*, pp. 5–8, 2015.

[30] B. Shi and T. Weninger, "Fact checking in heterogeneous information networks," in *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 101–102, 2016.

[31] A. L. Ginsca, A. Popescu, and M. Lupu, "Credibility in information retrieval," *Foundations and Trends in Information Retrieval*, vol. 9, no. 5, pp. 355–475, 2015.

[32] K. Pogorelov, D. T. Schroeder, P. Filkukova, and J. Langguth, "A system for high performance mining on gdelt data," in *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 1101–1111, IEEE, 2020.

[33] D. T. Schroeder, K. Pogorelov, and J. Langguth, "Fact: a framework for analysis and capture of twitter graphs," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 134–141, IEEE, 2019.

[34] L. Burchard, D. T. Schroeder, S. Becker, and J. Langguth, "Resource efficient algorithms for message sampling in online social networks," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–8, IEEE, 2020.

[35] K. Pogorelov, D. T. Schroeder, P. Filkukova, and J. Langguth, "A system for high performance mining on gdelt data," 2020.

[36] J. Langguth, P. Filkuková, S. Brenner, D. T. Schroeder, and K. Pogorelov, "Covid-19 and 5g conspiracy theories: long term observation of a digital wildfire," *International Journal of Data Science and Analytics*, pp. 1–18, 2022.

[37] K. Pogorelov, D. T. Schroeder, L. Burchard, J. Moe, S. Brenner, P. Filkukova, and J. Langguth, "Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020," in *MediaEval 2020 Workshop*, 2020.

[38] K. Pogorelov, D. T. Schroeder13, S. Brenner, and J. Langguth, "Fakenews: Corona virus and conspiracies multimedia analysis task at mediaeval 2021," in *Multimedia Benchmark Workshop*, p. 67, 2021.

[39] D. T. Schroeder, K. Pogorelov, and J. Langguth, "Evaluating standard classifiers for detecting covid-19 related misinformation,"

[40] D. T. Schroeder, F. Schaal, P. Filkukova, K. Pogorelov, and J. Langguth, "Wico graph: A labeled dataset of twitter subgraphs based on conspiracy theory and 5g-corona misinformation tweets.," in *ICAART (2)*, pp. 257–266, 2021.

[41] K. Pogorelov, D. T. Schroeder, P. Filkukova, and J. Langguth, "Wico text: a labeled dataset of conspiracy theory and 5g-corona misinformation tweets," in *ACM International Conference on Information Technology for Social Good*, 2021.

[42] D. T. Schroeder, P. G. Lind, K. Pogorelov, and J. Langguth, "A framework for interaction-based propagation analysis in online social networks,"

[43] D. T. Schroeder, J. Langguth, L. Burchard, K. Pogorelov, and P. G. Lind, "The connectivity network underlying the german's twittersphere: a testbed for investigating information spreading phenomena," *Scientific Reports*, vol. 12, no. 1, pp. 1–13, 2022.

[44] J. Langguth, K. Pogorelov, S. Brenner, P. Filkuková, and D. T. Schroeder, "Don't trust your eyes: Image manipulation in the age of deepfakes," *Frontiers in Communication*, p. 26, 2021.

[45] D. T. Schroeder, K. Styp-Rekowski, F. Schmidt, A. Acker, and O. Kao, "Graph-based feature selection filter utilizing maximal cliques," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 297–302, IEEE, 2019.

[46] L. Burchard, J. Moe, D. T. Schroeder, K. Pogorelov, and J. Langguth, "ipug: Accelerating breadth-first graph traversals using manycore graphcore ipus," in *International Conference on High Performance Computing*, pp. 291–309, Springer, 2021.

[47] M. Gabielkov, A. Ramachandran, A. Chaintreau, and A. Legout, "Social clicks: What and who gets read on twitter?," in *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science*, pp. 179–192, 2016.

[48] R. Hobbs, *Mind over media: Propaganda education for a digital age*. WW Norton & Company, 2020.

[49] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[50] R. Hobbs, C. Seyferth-Zapf, and S. Grafe, "Using virtual exchange to advance media literacy competencies through analysis of contemporary propaganda," *Journal of Media Literacy Education*, vol. 10, no. 2, pp. 152–168, 2018.

[51] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[52] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[53] N. Higdon, *The Anatomy of Fake News: A Critical News Literacy Education*. University of California Press, 2020.

[54] C. Wardle, "Fake news. it's complicated.," May 2017.

[55] H. D. Lasswell, *Propaganda technique in the world war*. Ravenio Books, 1927.

[56] B. Rubin, "Techniques of persuasion: From propaganda to brainwashing," 1965.

[57] G. Maletzke, "Propaganda. eine begriffskritische analyse," *Publizistik*, vol. 17, no. 2, p. 153, 1972.

[58] P. M. Taylor, *Munitions of the mind: A history of propaganda from the ancient world to the present era*. Manchester University Press, 2013.

[59] M. LeBoeuf, "The power of ridicule: An analysis of satire," *Senior Honors Projects*, p. 63, 2007.

[60] R. Jackall, *Propaganda*, vol. 8. NYU Press, 1995.

[61] N. A. Karlova and K. E. Fisher, "A social diffusion model of misinformation and disinformation for understanding human information behaviour," 2013.

[62] C. Fox, "Information and misinformation. an investigation of the notions of information, misinformation, informing, and misinforming," 1983.

[63] R. M. Losee, "A discipline independent definition of information," *Journal of the American Society for information Science*, vol. 48, no. 3, pp. 254–269, 1997.

[64] R. Gregor, "Ladislav bittman," the kgb and soviet disinformation: An insider's view"," *Canadian Slavonic Papers*, vol. 28, no. 4, p. 480, 1986.

[65] D. Fallis, "A conceptual analysis of disinformation," 2009.

[66] M. A. Porter and J. P. Gleeson, "Dynamical systems on networks," *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, vol. 4, 2016.

[67] J. Paßmann, *Die soziale Logik des Likes: Eine Twitter-Ethnografie*. Campus Verlag, 2018.

[68] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "Wtf: The who to follow service at twitter," in *Proceedings of the 22nd international conference on World Wide Web*, pp. 505–514, 2013.

[69] L. Hammer, "Bachelorarbeit 2020: Vermessung der deutschsprachigen twittersphäre," 2020.

[70] C. Torres-Lugo, K.-C. Yang, and F. Menczer, "The manufacture of political echo chambers by follow train abuse on twitter," *arXiv preprint arXiv:2010.13691*, 2020.

[71] A. Almaatouq, A. Alabdulkareem, M. Nouh, E. Shmueli, M. Alsaleh, V. K. Singh, A. Alarifi, A. Alfaris, and A. Pentland, "Twitter: who gets caught? observed trends in social micro-blogging spam," in *Proceedings of the 2014 ACM conference on Web science*, pp. 33–41, 2014.

[72] A. Bruns, B. Moon, F. Münch, and T. Sadkowsky, "The australian twittersphere in 2016: Mapping the follower/followee network," *Social Media+ Society*, vol. 3, no. 4, p. 2056305117748162, 2017.

[73] A. Maireder and J. Ausserhofer, "Political discourses on twitter: networking topics, objects and people," *Twitter and society*, vol. 89, pp. 305–318, 2014.

[74] A. Bruns and G. S. Enli, "The norwegian twittersphere: structure and dynamics," *Nordicom Review*, vol. 39, no. 1, pp. 129–148, 2018.

[75] D. Van Geenen, M. T. Schäfer, T. Boeschoten, E. Hekman, P. Bakker, and J. Moons, "Mining one week of twitter. mapping networked publics in the dutch twittersphere.," 2016.

[76] J. Kelly, V. Barash, K. Alexanyan, B. Etling, R. Faris, U. Gasser, and J. G. Palfrey, "Mapping russian twitter," *Berkman Center Research Publication*, no. 2012-3, 2012.

[77] M. S. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[78] J. D. Montgomery, "Job search and network composition: Implications of the strength-of-weak-ties hypothesis," *American Sociological Review*, pp. 586–596, 1992.

[79] C. Bilton, "Manageable creativity," *International Journal of Cultural Policy*, vol. 16, no. 3, pp. 255–269, 2010.

[80] J. M. McPherson and L. Smith-Lovin, "Women and weak ties: Differences by sex in the size of voluntary organizations," *American Journal of Sociology*, vol. 87, no. 4, pp. 883–904, 1982.

[81] R. L. Coser, "The complexity of roles as a seedbed of individual autonomy," *The idea of social structure: Papers in honor of Robert K. Merton*, pp. 237–263, 1975.

[82] O. Zorzi, "Granovetter (1983): The strength of weak ties: A network theory revisited," in *Schlüsselwerke der Netzwerkforschung*, pp. 243–246, Springer, 2019.

[83] J. Kostka, Y. A. Oswald, and R. Wattenhofer, "Word of mouth: Rumor dissemination in social networks," in *International colloquium on structural information and communication complexity*, pp. 185–196, Springer, 2008.

[84] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics and endemic states in complex networks," *Physical Review E*, vol. 63, no. 6, p. 066117, 2001.

[85] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.

[86] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, pp. 888–893, 2010.

[87] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[88] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing letters*, vol. 12, no. 3, pp. 211–223, 2001.

[89] R. K. Ghosh and G. Bhattacharjee, "Parallel breadth-first search algorithms for trees and graphs," *International Journal of Computer Mathematics*, vol. 15, no. 1-4, pp. 255–268, 1984.

[90] D. Gregor and A. Lumsdaine, "Lifting sequential graph algorithms for distributed-memory parallel computation," *ACM SIGPLAN Notices*, vol. 40, no. 10, pp. 423–437, 2005.

[91] A. Yoo, E. Chow, K. Henderson, W. McLendon, B. Hendrickson, and U. Catalyurek, "A scalable distributed parallel breadth-first search algorithm on bluegene/l," in *SC'05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, pp. 25–25, IEEE, 2005.

[92] D. A. Bader and K. Madduri, "Designing multithreaded algorithms for breadth-first search and st-connectivity on the cray mta-2," in *2006 International Conference on Parallel Processing (ICPP'06)*, pp. 523–530, IEEE, 2006.

[93] R. E. Korf and P. Schultze, "Large-scale parallel breadth-first search," in *AAAI*, vol. 5, pp. 1380–1385, 2005.

[94] P. Harish and P. J. Narayanan, "Accelerating large graph algorithms on the gpu using cuda," in *International conference on high-performance computing*, pp. 197–208, Springer, 2007.

[95] R. C. Murphy, K. B. Wheeler, B. W. Barrett, and J. A. Ang, "Introducing the graph 500," *Cray Users Group (CUG)*, vol. 19, pp. 45–74, 2010.

[96] A. Buluç and K. Madduri, "Parallel breadth-first search on distributed memory systems," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–12, 2011.

[97] F. Checconi and F. Petrini, "Traversing trillions of edges in real time: Graph exploration on large-scale parallel machines," in *2014 IEEE 28th International Parallel and Distributed Processing Symposium*, pp. 425–434, IEEE, 2014.

[98] Z. Chenglong, C. Huawei, W. Guobo, H. Qinfen, Z. Yang, Y. Xiaochun, and F. Dongrui, "Efficient optimization of graph computing on high-throughput computer," *Journal of Computer Research and Development*, vol. 57, no. 6, p. 1152, 2020.

[99] S. Hong, T. Oguntebi, and K. Olukotun, "Efficient parallel graph exploration on multi-core cpu and gpu," in *2011 International Conference on Parallel Architectures and Compilation Techniques*, pp. 78–88, IEEE, 2011.

[100] Y. Yasui, K. Fujisawa, and K. Goto, "Numa-optimized parallel breadth-first search on multicore single-node system," in *2013 IEEE International Conference on Big Data*, pp. 394–402, IEEE, 2013.

[101] A. Gaihre, Z. Wu, F. Yao, and H. Liu, "Xbfs: exploring runtime optimizations for breadth-first search on gpus," in *Proceedings of the 28th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 121–131, 2019.

[102] H. Liu and H. H. Huang, "Enterprise: breadth-first graph traversal on gpus," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–12, 2015.

[103] Y. Wang, A. Davidson, Y. Pan, Y. Wu, A. Riffel, and J. D. Owens, "Gunrock: A high-performance graph processing library on the gpu," in *Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, pp. 1–12, 2016.

[104] C. Yang, A. Buluc, and J. D. Owens, "Graphblast: A high-performance linear algebra-based graph framework on the gpu," 2020.

[105] S. Beamer, K. Asanovic, D. Patterson, S. Beamer, and D. Patterson, "Searching for a parent instead of fighting over children: A fast breadth-first search implementation for graph500," *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2011-117*, 2011.

[106] A. Azad and A. Buluç, "Distributed-memory algorithms for maximum cardinality matching in bipartite graphs," in *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 32–42, IEEE, 2016.

[107] J. Langguth, A. Azad, M. Halappanavar, and F. Manne, "On parallel push–relabel based algorithms for bipartite maximum matching," *Parallel Computing*, vol. 40, no. 7, pp. 289–308, 2014.

[108] J. Langguth, M. M. A. Patwary, and F. Manne, "Parallel algorithms for bipartite matching problems on distributed memory computers," *Parallel Computing*, vol. 37, no. 12, pp. 820–845, 2011.

[109] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[110] H. Feng, J. Tian, H. J. Wang, and M. Li, "Personalized recommendations based on time-weighted overlapping community detection," *Information & Management*, vol. 52, no. 7, pp. 789–800, 2015.

[111] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, 2012.

[112] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[113] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[114] A.-L. Barabási, "Network science," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.

[115] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[116] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On modularity clustering," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 2, pp. 172–188, 2007.

[117] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[118] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Physical review E*, vol. 72, no. 2, p. 027104, 2005.

[119] V. A. Traag, L. Waltman, and N. J. Van Eck, "From louvain to leiden: guaranteeing well-connected communities," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[120] L. Waltman and N. J. Van Eck, "A smart local moving algorithm for large-scale modularity-based community detection," *The European physical journal B*, vol. 86, no. 11, pp. 1–14, 2013.

[121] N. Ozaki, H. Tezuka, and M. Inaba, "A simple acceleration method for the louvain algorithm," *International Journal of Computer and Electrical Engineering*, vol. 8, no. 3, p. 207, 2016.

[122] S.-H. Bae, D. Halperin, J. D. West, M. Rosvall, and B. Howe, "Scalable and efficient flow-based community detection for large-scale graph analysis," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 3, pp. 1–30, 2017.

[123] V. A. Traag, "Faster unfolding of communities: Speeding up the louvain algorithm," *Physical Review E*, vol. 92, no. 3, p. 032801, 2015.

[124] S. M. Van Dongen, *Graph clustering by flow simulation*. PhD thesis, 2000.

[125] V. Satuluri and S. Parthasarathy, "Scalable graph clustering using stochastic flows: applications to community discovery," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 737–746, 2009.

[126] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social networks*, vol. 28, no. 4, pp. 466–484, 2006.

[127] S. Wasserman, K. Faust, *et al.*, "Social network analysis: Methods and applications," 1994.

[128] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[129] A. Bavelas, "Communication patterns in task-oriented groups," *The journal of the acoustical society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[130] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[131] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[132] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," tech. rep., Stanford InfoLab, 1999.

[133] Q. Li, S. Shah, M. Thomas, K. Anderson, X. Liu, A. Nourbakhsh, and R. Fang, "How Much Data Do You Need? Twitter Decahose Data Analysis," July 2016.

[134] "Decahose stream."

[135] K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann, *Twitter and society*, vol. 89. Peter Lang, 2014.

[136] S. Gilbert and D. Gaffney, "The 140kit." `https://bit.ly/33OxlWb`, 2010.

[137] P. Barbera, V. Haunschmid, and D. Kronovet, "streamr: Access to twitter streaming api via r." `https://bit.ly/3uXhxft`, 2018.

[138] J. OBrien, "yourtwapperkeeper: an open version of twapperkeeper.com." `https://bit.ly/3eUVbWt`, 2011.

[139] Y. Roth and R. Johnson, "New developer requirements to protect our platform." `http://bit.ly/2XD5Z3Q`, 2018.

[140] `http://twitonomy.com/`, 2019.

[141] X. Gao and J. Qiu, "Supporting queries and analyses of large-scale social media data with customizable and scalable indexing techniques over nosql databases," in *2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pp. 587–590, May 2014.

[142] E. Borra and B. Rieder, "Programmed method: developing a toolset for capturing and analyzing tweets," *Aslib Journal of Information Management*, vol. 66, no. 3, pp. 262–278, 2014.

[143] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The pushshift reddit dataset," 2020.

[144] A. McCrow-Young, "Approaching instagram data: reflections on accessing, archiving and anonymising visual social media," *Communication Research and Practice*, pp. 1–14, 2020.

[145] M. Schroepfer, "An update on our plans to restrict data access on facebook," *Facebook newsroom*, vol. 4, 2018.

[146] P. Singer, F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier, "Evolution of reddit: from the front page of the internet to a self-referential community?," in *Proceedings of the 23rd international conference on world wide web*, pp. 517–522, 2014.

[147] K. E. Anderson, "Ask me anything: what is reddit?," *Library Hi Tech News*, 2015.

[148] M. Fire and C. Guestrin, "The rise and fall of network stars: Analyzing 2.5 million graphs to reveal how high-degree vertices emerge over time," *Information Processing & Management*, vol. 57, no. 2, p. 102041, 2020.

[149] J. Stern, "In the elevator with reddit CEO steve huffman." The Wall Street Journal: Video.

[150] K. Leetaru and P. A. Schrodt, "Gdelt: Global data on events, location, and tone, 1979–2012," in *ISA annual convention*, vol. 2, pp. 1–49, Citeseer, 2013.

[151] K. Long and L. Hanks, "reddit wiki API." Library Catalog: github.com.

[152] L. Burchard, D. T. Schroeder, K. Pogorelov, S. Becker, E. Dietrich, P. Filkukova, and J. Langguth, "A scalable system for bundling online social network mining research," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–6, IEEE, 2020.

[153] A. Huber, "Observing reddit's interaction network - an actor-based approach for large scale network analysis on reddit," in *Master thesis University Oslo*, 2021.

[154] H. Rana, "Simulating diffusion of fake news on online social networks," in *Master thesis University Oslo*, 2021.

[155] I. MongoDB, "Mongodb," *URL https://www. mongodb. com/. Cited on (2014)*, vol. 9, 2014.

[156] J. Webber, "A programmatic introduction to neo4j," in *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*, pp. 217–218, 2012.

[157] A. Vukotic, N. Watt, T. Abedrabbo, D. Fox, and J. Partner, *Neo4j in action*, vol. 22. Manning Shelter Island, 2015.

[158] R. Van Bruggen, *Learning Neo4j*. Packt Publishing Ltd, 2014.

[159] D. Fernandes and J. Bernardino, "Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb.," in *DATA*, pp. 373–380, 2018.

[160] K. Chodorow, *Scaling MongoDB: Sharding, Cluster Setup, and Administration*. " O'Reilly Media, Inc.", 2011.

[161] L. Liu, L. Wang, W. Wu, H. Jia, and Y. Zhang, "A novel hybrid-jump-based sampling method for complex social networks," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 241–249, 2019.

[162] D. Sayce, "The number of tweets per day in 2020," Dec 2020.

[163] X. Han, X. Gu, and S. Peng, "Analysis of tweet form's effect on users' engagement on twitter," *Cogent Business & Management*, 2019.

[164] J. Pfeffer, K. Mayer, and F. Morstatter, "Tampering with twitter's sample API," *EPJ Data Science*, vol. 7, no. 1, pp. 1–21, 2018. Number: 1 Publisher: SpringerOpen.

[165] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[166] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who says what to whom on twitter," in *Proceedings of the 20th international conference on World wide web*, WWW '11, pp. 705–714, Association for Computing Machinery, 2011.

[167] P. Mathews, L. Mitchell, G. Nguyen, and N. Bean, "The nature and origin of heavy tails in retweet activity," in *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, pp. 1493–1498, International World Wide Web Conferences Steering Committee, 2017.

[168] D. R. Bild, Y. Liu, R. P. Dick, Z. M. Mao, and D. S. Wallach, "Aggregate characterization of user behavior in twitter and analysis of the retweet graph," *ACM Transactions on Internet Technology*, vol. 15, no. 1, pp. 1–24, 2015.

[169] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *2010 IEEE Second International Conference on Social Computing*, pp. 177–184, 2010.

[170] D. Michail, J. Kinable, B. Naveh, and J. V. Sichi, "Jgrapht—a java library for graph data structures and algorithms," *ACM Transactions on Mathematical Software (TOMS)*, vol. 46, no. 2, pp. 1–29, 2020.

[171] P. T. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn, "Do retweets indicate interest, trust, agreement?," *arXiv preprint arXiv:1411.3555*, 2014.

[172] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *2010 IEEE second international conference on social computing*, pp. 177–184, IEEE, 2010.

[173] R. Recuero, R. Araujo, and G. Zago, "How does social capital affect retweets?," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[174] F. Zuiderveen Borgesius, D. Trilling, J. Möller, B. Bodó, C. H. De Vreese, and N. Helberger, "Should we worry about filter bubbles?," *Internet Policy Review. Journal on Internet Regulation*, vol. 5, no. 1, 2016.

[175] A. Bruns, "Echo chamber? what echo chamber? reviewing the evidence," in *6th Biennial Future of Journalism Conference (FOJ17)*, 2017.

[176] W. Ahmed, J. Downing, M. Tuters, and P. Knight, "Four experts investigate how the 5g coronavirus conspiracy theory began," *The Conversation*, 2020.

[177] P. Metaxas and S. T. Finn, "The infamous# pizzagate conspiracy theory: Insight from a twittertrails investigation," 2017.

[178] M. Spring, "Wayfair: The false conspiracy about a furniture firm and child trafficking," Jul 2020.

[179] O. Lysne, *The Huawei and Snowden Questions: Can Electronic Equipment from Untrusted Vendors be Verified? Can an Untrusted Vendor Build Trust Into Electronic Equipment?* Springer Nature, 2018.

[180] R. Baan, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, F. Islami, L. Galichet, and K. Straif, "Carcinogenicity of radiofrequency electromagnetic fields," *The lancet oncology*, vol. 12, no. 7, pp. 624–626, 2011.

[181] R. Imhoff and P. Lamberty, "How paranoid are conspiracy believers? toward a more fine-grained understanding of the connect and disconnect between paranoia and belief in conspiracy theories," *European Journal of Social Psychology*, vol. 48, no. 7, pp. 909–926, 2018.

[182] K. M. Douglas and R. M. Sutton, "The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of princess diana," *The Journal of social psychology*, vol. 148, no. 2, pp. 210–222, 2008.

[183] M. Grzesiak-Feldman, "Conspiracy thinking and state-trait anxiety in young polish adults," *Psychological reports*, vol. 100, no. 1, pp. 199–202, 2007.

[184] J. W. McHoskey, "Case closed? on the john f. kennedy assassination: Biased assimilation of evidence and attitude polarization," *Basic and Applied Social Psychology*, vol. 17, no. 3, pp. 395–409, 1995.

[185] D. Sharp, "Advances in conspiracy theory.," *Lancet (London, England)*, vol. 372, no. 9647, pp. 1371–1372, 2008.

[186] V. Swami, A. Furnham, N. Smyth, L. Weis, A. Lay, and A. Clow, "Putting the stress on conspiracy theories: Examining associations between psychological stress, anxiety, and belief in conspiracy theories," *Personality and Individual Differences*, vol. 99, pp. 72–76, 2016.

[187] M. Grzesiak-Feldman, "The effect of high-anxiety situations on conspiracy thinking," *Current Psychology*, vol. 32, no. 1, pp. 100–118, 2013.

[188] B. Pfefferbaum and C. S. North, "Mental health and the covid-19 pandemic," *New England Journal of Medicine*, 2020.

[189] R. P. Rajkumar, "Covid-19 and mental health: A review of the existing literature," *Asian journal of psychiatry*, p. 102066, 2020.

[190] D. Freeman, F. Waite, L. Rosebrock, A. Petit, C. Causier, A. East, L. Jenner, A.-L. Teale, L. Carr, S. Mulhall, *et al.*, "Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in england," *Psychological Medicine*, pp. 1–13, 2020.

[191] P. Filkukova, P. Ayton, K. Rand, and J. Langguth, "What should i trust? individual differences in attitudes to conflicting information and misinformation on covid-19," *Frontiers in Psychology*, 2021.

[192] R. Imhoff and P. Lamberty, "A bioweapon or a hoax? the link between distinct conspiracy beliefs about the coronavirus disease (covid-19) outbreak and pandemic behavior," *Social Psychological and Personality Science*, vol. 11, no. 8, pp. 1110–1118, 2020.

[193] D. Jolley and J. L. Paterson, "Pylons ablaze: Examining the role of 5g covid-19 conspiracy beliefs and support for violence," *British journal of social psychology*, vol. 59, no. 3, pp. 628–640, 2020.

[194] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, 2014.

[195] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[196] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," 2020.

[197] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[198] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint arXiv:1902.06673*, 2019.

[199] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[200] M. Müller, M. Salathé, and P. E. Kummervold, "Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter," *arXiv preprint arXiv:2005.07503*, 2020.

[201] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65–74, 2011.

Table 1: The collection of 5G-Corona news articles found in the GDELT database for the period from February 20 to 27 (identical titles represent articles that are essentially copies of other articles).

Date	Supports 5G-Corona conspiracy	Number of referred events	Publisher	Title
2020-02-20	No	1	lightreading.com	Coronavirus cuts into 5G standards work
2020-02-20	No	12	reuters.com	Huawei says no impact on 5G supply from coronavirus
2020-02-20	No	12	investing.com	Huawei says no impact on 5G supply from coronavirus
2020-02-20	No	11	marketscreener.com	Huawei says no impact on 5G supply from coronavirus
2020-02-20	No	14	reuters.com	Huawei says no impact on 5G supply from coronavirus
2020-02-20	No	11	oann.com	*# URL not found*
2020-02-20	No	3	cnn.com	Apple needs a 5G iPhone now more than ever
2020-02-20	**YES**	1	veteranstoday.com	UPDATE On Coronavirus, 5G, ELF, Apirin, Mismanagement — the Perfect Storm
2020-02-20	No	11	phonearena.com	Production of Huawei's 5G networking equipment unaffected by the coronavirus
2020-02-20	No	11	cnbc.com	UPDATE 1-Huawei says no impact on 5G supply from coronavirus
2020-02-21	No	13	abs-cbn.com	Huawei says no impact on 5G supply from coronavirus
2020-02-21	No	11	indiatimes.com	Huawei says no impact on 5G supply from coronavirus
2020-02-21	No	6	reuters.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	reuters.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	reuters.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	channelnewsasia.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	oann.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	reuters.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	4	businesstimes.com.sg	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	4	msn.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	4	indiatimes.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	4	indiatimes.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	6	marketscreener.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	2	theepochtimes.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	1	marketwatch.com	Opinion: The 5G rollout is already behind, and coronavirus could slow it even more
2020-02-21	No	6	msn.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	2	ndtv.com	China's ambitious 5G push heading into slow lane due to coronavirus disruptions
2020-02-21	No	1	lightreading.com	*# URL not found*
2020-02-22	*Nothing for the date*			
2020-02-23	No	1	nyoooz.com	China's ambitious 5G push slows due to coronavirus disruptions
2020-02-24	No	7	sputniknews.com	India Gets 'First 5G Smartphone' From China Amid Delayed Trials Due to Coronavirus
2020-02-24	No	4	computerweekly.com	Coronavirus, global economic slowdown to cap 5G smartphone sales in 2020
2020-02-25	No	3	zdnet.com	Jabil cuts outlook as coronavirus curbs manufacturing for tech giants, 5G ecosystem
2020-02-26	*Nothing for the date*			
2020-02-27	*Nothing for the date*			

Table 2: A collection of news articles we used to keep track of worldwide real-world consequences linked to the 5G-Corona misinformation event (arson attacks and harassment of telecommunication technicians). Column #e contains the number of events with an exact location mentioned in the article, #m all mentioned events in the article. Although the original table was much larger, this selection contains only those articles with precise location information. Links start with https://bit.ly/

Date	#e	#m	Country	Location	Media/Source	Title	Journalist/editor	Link
2020-01-02	1		DE	Bonn	General Anzeiger	Polizei vermutet Brandstiftung nach Feuer an ...	Silke Elbern	3p3rrtL
2020-03-27	1	14	NZ	Waiharara	MSN	Suspected cell tower arsons prompt call for witnesses	Liu Chen	2TsBzRl
2020-04-02	1	30	IE	Belfast	New York Times	Burning Cell Towers, Out of Baseless Fear ...	Adam Satariano et al.	34v92g0
2020-04-02	1		UK	Birmingham	The Sun	HOLY SMOKE Idiots 'BURNING 5G masts' ...	Holly Christodoulou	34sYnCH
2020-04-03	2	30	UK	Liverpool	Independent	Coronavirus 5G conspiracy theory ...	Adam Satariano et al.	3fS0VPX
2020-04-04	1	4	NL	Beesd	Telegraaf	Wéér incident bij mast: verzet 5G wordt militant	Martin Nuver	3fy4vQm
2020-04-05	1	14	NZ	Manurewa	MSN	Suspected cell tower arsons prompt ...	Liu Chen	3vtShhc
2020-04-05	1		NL	Liessel	Omroepbrabant	Leuzen tegen 5G bij brandende telefoonmast in Liessel	Femke de Jong	34Hcrc1
2020-04-05	1		NL	Rotterdam	Telegraaf	Wéér incident bij mast: verzet 5G wordt militant	Martin Nuver	3wNbGdr
2020-04-05	1		NL	Deurne	DMG	Brand in GSM-mast Liessel waarschijnlijk ...	Ivo Boudewijns	2R6S1G6
2020-04-09	1		NL	Nuenen	Omroepbrabant	Brand in telefoonmast in Nuenen, mogelijk aangestoken	Peter de Bekker	3p3vF4Y
2020-04-10	1		NL	Groningen	Telegraaf	Wéér incident bij mast: verzet 5G wordt militant	Martin Nuver	34HcyEt
2020-04-10	1		NL	Oudenbosch	NOS	Waarom worden door heel ... ?	ANP	2RTf2wx
2020-04-11	1	20	UK	Birmingham	Telegraph	Birmingham Nightingale phone mast ...	Matthew Field	3vKNQip
2020-04-11	1		UK	West Yorkshire	Telegraph	Birmingham Nightingale phone ...	Matthew Field	3i2ubGE
2020-04-11	1	several	CY	Limassol	Ars Technica	How a 5G coronavirus ...	Nic Fildes et al.	3fuCjOc
2020-04-11	0	22	UK	Several locations	Businessinsider	Vandals set 50 cellphone masts in the ...	Isobel Asher Hamilton	3cl4e1p
2020-04-11	1		NL	Veldhoven in Brabant	NOS	Waarom worden door heel Nederland zendmasten ... ?	ANP	3c4jdfP
2020-04-11	2		IE	Co Donegal	Irish Times	Gardaí suspect masts set on fire deliberately in Co Donegal	Stephen Maguire	3fwdPV5
2020-04-13	1	several	NL	Amsterdam (Almere)	Ars Technica	How a 5G coronavirus conspiracy spread across Europe	Nic Fildes et al.	3i16NJq
2020-04-14	1	20	UK	Dagenham, Essex	BBC	Coronavirus: 20 suspected phone mast attacks over Easter	Leo Kelion	2SFdMgE
2020-04-17	1	77	UK	Huddersfield	Wired	The 5G coronavirus conspiracy ...	James Tamperton	34uiP6c
2020-04-19	1		DE	Bonn	General Anzeiger	Gibt es in Bonn eine Brandanschlagserie auf Funkmasten?	Ayla Jacob	2RSoNeg
2020-04-22	1		DE	Wilhelmshaven	Police press	POL-WHV: 30-Jähriger durchtrennt ...		3fWykcb
2020-04-28	1	several	NZ	Papatoetoe	MSN	Suspected cell tower arsons prompt call for witnesses	Liu Chen	3fVCuBc
2020-05-01	1		CA	Montreal	CTV News	Cell tower set on fire north of Montreal	Katelyn Thomas	34qTkCz
2020-05-04	1		CA	Piedmont	CTV News	Two more cell towers went up in flames north of Montreal	Katelyn Thomas	3uxRJWf
2020-05-04	1		CA	Prévost	CTV News	Two more cell towers went up in flames north of Montreal	Katelyn Thomas	3c2tEQV
2020-05-05	1		CA	Laval	CTV News	Another cell tower in Quebec – the fourth ...	Katelyn Thomas	3uBWzSh
2020-05-12	1		NZ	Māngere	MSN	Suspected cell tower arsons prompt call for witnesses	Liu Chen	3c1Tftr
2020-05-15	1		NZ	Ōtāhuhu	MSN	Suspected cell tower arsons prompt call for witnesses	Liu Chen	3p8r80X
2020-05-15	1		NZ	Favona	MSN	Suspected cell tower arsons prompt call for witnesses	Liu Chen	3i7ZT5n
2020-05-18	1		AU	Melbourne	news.com.au	Cranbourne West phone tower fire: Counter-terrorism ...	Jack Paynter	3fwCQzh
2020-08-06	3	20	NZ	South Auckland	TVNZ	Cell phone towers burned in latest 'senseless' arson ...	Luke Appleby	3vC0TCG
2020-06-10	8		PE	Provinz Acobamba	RPP	Huancavelica: Secuestran a 8 ingenieros que arreglaban ...	Redacción	2RTRoA2
2020-07-01	1		CY	Limassol	KNEWS	More 5G targets torched overnight	unknown journalist	3uBWSfT
2020-08-10	1		UK	Essex	Essex Live	Police fear Chelmsford 5G mast 'will fall over' ...	unknown journalist	3vwXrZO
2020-09-10	1	90	UK	Bradford	Daily Mail	Arsonists destroy another 5G mast fueled by false ...	Clare McCarthy	3wIETG6
2020-09-23	1		UK	Bradford	Telegraph & Argus	Police appeal for information as 5G mast set on ...	unknown journalist	3vzMw1t
2020-11-05	1		UK	Bierley	Telegraph & Argus	Firefighters called to 5G mast blaze in Bierley	Felicity Macnamara	3yOPUYk
2021-01-05	4		ZA	Durban, KwaZulu-Natal	Daily Maverick	KZN cellphone towers torched as 5G ...	Rebecca Pitt	2RTfM4N
2021-01-31	3		UK	Chelmsford, Essex	Telegraph	Conspiracy theorists could be behind ...	Martin Evans	3p2bQLc
2021-03-31	1		CA	Scarborough	Toronto Citynews	Police say cell phone tower fire in ...	"News Staff"	3p3thuF

Table 3: Twitter lists including the initial accounts from which the data was collected.

Twitter List	
kunigkeit/pod-parteien-politiker	wahl_beobachter/mdl-sachsen-anhalt
VeHoltz/bundespolitik	wahl_beobachter/mdl-baden-w-rttemberg
MarcusSchwarze/fakes	wahl_beobachter/mdl-nrw
wahl_beobachter/botschaften	wahl_beobachter/mdl-saarland
wahl_beobachter/kandidaten-europawahl	wahl_beobachter/mdl-schleswig-holstein
wahl_beobachter/bundesministerien	wahl_beobachter/mdl-mecklenburg-vorpommen
wahl_beobachter/bundestagsfraktionen	wahl_beobachter/abgeordnetenhaus-agh
wahl_beobachter/mdb-bundestag	wahl_beobachter/bundesregierung
wahl_beobachter/mdl-bayern	wahl_beobachter/politikwissenschaftler
wahl_beobachter/mdl-hessen	wahl_beobachter/ministeriums-twitterati
wahl_beobachter/mdl-niedersachsen	wahl_beobachter/alle-25-parteien-ep2014
wahl_beobachter/mdbb-bremen	wahl_beobachter/deutsche-mep-2019-2024
wahl_beobachter/mdl-brandenburg	wahl_beobachter/open-government-hamburg
wahl_beobachter/mdl-sachsen	wahl_beobachter/mdhb-hamburg
wahl_beobachter/mdl-rheinland-pfalz1	AfD/verifizierte-accounts
wahl_beobachter/mdl-th-ringen	AfD/bundestagsabgeordnete