Detection of Conspiracy Theories on Telegram

Leveraging Graph Theory and Natural Language Processing for Influential Channel and Content Analysis

SEN2331: CoSEM Master Thesis

Neal Chang-Sing-Pang

Delft University of Technology

TU Delft

Detection of Conspiracy Theories on Telegram

Leveraging Graph Theory and Natural Language Processing for Influential Channel and Content Analysis

by

Neal Chang-Sing-Pang

Student Name	Student Number
Neal Chang-Sing-Pang	4687590

Chair:	Prof.dr. M.E. (Martijn) Warnier
First supervisor:	Dr.ir. P.W. (Petra) Heijnen
Second supervisor:	Dr. S. (Savvas) Zannettou
Faculty:	Faculty of Technology, Policy & Management, Delft

Cover: Flat Earth with Illuminati created with OpenAI's DALL-E

TU Delft

Preface

As I present this master thesis, *"Detection of Conspiracy Theories on Telegram"*, my years in Delft as a student at the University of Technology Delft have come to an end. Reflecting on my journey at the Faculty of Technology, Policy and Management, I can genuinely say that this thesis serves as the cherry on top of the cake, marking a significant personal milestone. From the day I started the CoSEM program, I viewed the thesis as a daunting challenge, something to be dreaded. However, as I navigated through the past half-year, the process proved to be more fulfilling than I had anticipated. Time flew by quickly, largely because I had the opportunity to work on a topic that genuinely piqued my interest and because of the strong camaraderie within my friend group. We studied together, faced the challenges of our theses together, and in doing so, made the journey all the more rewarding.

I would like to extend my deepest gratitude to my three supervisors, Petra Heijnen, Savvas Zannettou and Martijn Warnier, for their invaluable input, time, and flexibility throughout this process. Their guidance and feedback were crucial in shaping this thesis, and their support has been instrumental in my growth as a researcher. I am truly grateful for the opportunity to have worked under their supervision.

Lastly, I would like to express my heartfelt thanks to my friends and family, who have supported me unconditionally throughout these years. Their encouragement and belief in me have been a constant source of motivation. I have put a great deal of energy, effort, and even fun into this thesis, and I hope that you, the reader, will enjoy reading it as much as I enjoyed doing it.

Neal Chang-Sing-Pang
Delft, August 2024

Summary

In the digital age, the dissemination of conspiracy theories through social media platforms poses significant challenges to societal stability and public trust. Telegram, known for its encryption and privacy features, has emerged as a key platform for the dissemination of such misinformation. This study aims to develop a robust model to detect conspiracy theories on Telegram at an early stage, combining graph theory and natural language processing (NLP) techniques. The focus is on identifying influential channels and classifying conspiracy-related content to understand and mitigate the spread of misinformation. The main research question guiding this study is: *How can conspiracy theories be identified on Telegram?*

The research methodology encompasses several critical steps. Firstly, Telegram channels related to the Great Reset conspiracy theory were identified using custom web crawlers and manual validation. A dataset of 747,224 messages from 48 channels, spanning from November 1, 2021, to January 31, 2024, was collected. The network structure of Telegram channels was then modeled as a directed weighted graph, where each node represents a channel, and edges represent forwarded messages. Centrality measures such as weighted degree centralities (both in and out), betweenness centrality, and viral message centrality were used to compute an influence score for each channel.

To detect conspiracy-related messages, m-BERT, a transformer-based model, was fine-tuned on a manually labeled dataset of 1,728 messages, of which 16% were labeled as conspiracy-related. The fine-tuned model achieved an accuracy of 84.1% and an F1 score of 0.768, indicating its effectiveness in classifying conspiracy-related content. BERTopic was employed to analyze and identify distinct conspiracy-related topics within the classified messages. The model identified 17 topics, all of which were linked to known conspiracy theories using the OpenAI API.

To validate the model, a fictive conspiracy theory called ”The Verdant Shadow Conspiracy” was created. This theory suggests that houseplants are not merely decorative or good for improving air quality, but are actually sophisticated surveillance devices genetically engineered by a covert organization known as ”Greenwatch” to monitor human activities worldwide. The OpenAI API was used to simulate 100 Dutch Telegram messages about this fictive conspiracy, which were then re-added to the dataset. Out of these messages, 10 were labeled as conspiracy-related. BERTopic was applied to the enriched dataset, resulting in 18 topics, including the detection of the fictive conspiracy topic. The fictive conspiracy was correctly identified as a new conspiracy, demonstrating the model's capability to detect previously unknown conspiracy theories.

The analysis identified the top 10 influential channels based on their influence scores, with a clear top 3 significantly impacting the network. The network visualization highlighted these top channels. The fine-tuned m-BERT model's performance metrics indicated strong classification capabilities. BERTopic's application to the most influential channel's messages revealed a relatively low percentage (2.11%) of messages classified as conspiracy-related, suggesting the model's high precision or the specific nature of the dataset. All identified topics were linked to known conspiracy theories, highlighting the classifier's accuracy.

The findings underscore the critical role of influential channels in spreading conspiracy theories on Telegram. The combined use of graph theory and NLP provides a comprehensive approach to the detection of conspiracy theories. However, the study also faced limitations such as the dataset's specificity and the strict classification threshold, which may need adjustment for broader applicability.

Future research should focus on expanding the dataset to include a wider range of channels and messages to capture a broader spectrum of conspiracy theories. Adjusting the classification threshold to balance precision and recall could potentially identify more diverse conspiracy theories. Additionally, integrating additional social media platforms could provide a more holistic understanding of misinformation spread. This study provides a foundation for further exploration and development of effective

tools to combat the spread of conspiracy theories, contributing to the enhancement of public trust and societal stability.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Conspiracy theories

In today's digitally interconnected world, the propagation of misinformation, particularly the circulation of conspiracy theories, poses a profound societal challenge. Within the Netherlands, this issue has emerged as a notable concern, intertwining with discussions about societal security and stability. As highlighted in a report by the Dutch intelligence department AIVD, conspiracy theories regarding an "evil elite" wielding power in the Netherlands are identified as posing a serious long-term threat to the nation's security (Algemene Inlichtingen- en Veiligheidsdienst, 2023).

A conspiracy theory is "the belief that certain events or situations are secretly manipulated behind the scenes by powerful forces with negative intent" (European Commission, 2023). Conspiracy theories are often used to undermine trust in institutions and to hinder the spread of information (Sunstein & Vermeule, 2008) and are also used as instruments of political propaganda (Cassam, 2019). These theories often suggest hidden agendas, cover-ups, or operations that are orchestrated to deceive the public. While some conspiracy theories have factual bases, they frequently lack evidence and are generally considered speculative and unfounded by mainstream scientific communities. Some well-known examples of conspiracy theories are:

- The moon landing hoax: this theory claims that the Apollo moon landings were staged by NASA in a Hollywood film studio (Lewandowsky, Oberauer, & Gignac, 2013, 5).
- Flat earth theory: this theory claims that the earth is flat rather than a sphere (Paolillo, 2018, 12).
- COVID-19 conspiracy theories: Various conspiracies have emerged surrounding the COVID-19 pandemic. Some examples are claims that the virus is an intentionally released bio weapon, that vaccines contain tracking microchips, or that the pandemic is a hoax orchestrated for political control (Douglas, 2021, 2).
- The idea that the Netherlands/world is ruled by a small, evil elite that wants to oppress, enslave, and even kill the population (Algemene Inlichtingen- en Veiligheidsdienst, 2023).

This topic can be seen as a typical CoSEM problem. A typical CoSEM problem is situated in a complex socio-technical environment. It includes issues covering both public and private domains, reflecting the versatile nature of modern technological challenges. The problem of the spread of conspiracy theories is a socio-technical challenge, merging societal dynamics with technological platforms. Conspiracy theories, often driven by distrust and speculation, have found a platform for propagation across various social media platforms (Napolitano & Reuter, 2021). Conspiracy theories can be hostile to all kinds of institutions that are important to the democratic rule of law (Algemene Inlichtingen- en Veiligheidsdienst, 2023). Understanding the dynamics of the spread on social media platforms like Twitter/X, Facebook, and Telegram is therefore important to the democratic rule of law.

Within the public domain, the spread of conspiracy theories challenges societal trust, influences public discussion, and necessitates policy and governance interventions from governments and public insti-

tutions. Simultaneously, within the private domain, tech companies are entrusted with content moderation responsibilities, necessitating ethical considerations and corporate accountability in managing information propagation.

1.2. Social Media

Social media platforms such as Twitter/X, Facebook, and Telegram have significantly increased the spread and impact of conspiracy theories in today's society. For a lot of people, social media have become the main source for news and updates (Ratkiewicz et al., 2011). People who get their news from social media and use social media frequently, believe more often in conspiracy theories (Enders et al., 2023). These platforms allow people to share information quickly and widely, often without proper checks for accuracy. On Twitter/X, the short and immediate nature of tweets helps conspiracy theories spread fast, especially when they are retweeted or liked by many users. Facebook's algorithm tends to push engaging content to the top, which can include sensational conspiracy theories, leading to echo chambers where users only see information that confirms their existing beliefs (Cinelli et al., 2022). Telegram, known for its encrypted messaging, provides a space for people to share conspiracy theories without fear of being monitored or censored.

The influence of social media is so big because these platforms use algorithms designed to keep users engaged, often by showing them content that is controversial or emotionally charged. This type of content tends to get more likes, shares, and comments, which means it gets seen by even more people (Ren, Dimant, & Schweitzer, 2023). Social media also creates communities where people can connect over shared interests, including conspiracy theories. These groups can make conspiracy theories seem more credible and accepted, as people reinforce each other's beliefs (González-Padilla & Tortolero-Blanco, 2020). This widespread and fast spread of misinformation makes it hard to have informed public discussions and can lead to divisions in society.

Telegram has become a key platform for spreading conspiracy theories, primarily because of its unique features that differ from other social media sites like Twitter/X and Facebook (Gerster, Kuchta, Hammer, & Schwieter, 2022). Unlike these platforms, Telegram allows users to create large public channels and groups with little content moderation, which means information can spread quickly and without much oversight. This is particularly important when it comes to conspiracy theories, as the lack of control allows these ideas to spread more freely. Telegram also offers strong privacy and security features, encouraging users to share content without fear of censorship. As a result, Telegram has become a popular space for those who want to share and discuss conspiracy theories, making it a critical platform to study. By focusing on Telegram, this research aims to develop effective methods for detecting conspiracy theories within this platform, addressing the challenges unique to monitoring and identifying such content in encrypted and less regulated digital environments.

1.3. Research Question

The main research question guiding this study is:

'How can Conspiracy Theories be identified on Telegram?'

This question addresses the challenge of detecting potentially harmful narratives early in their dissemination process. The objective of this research is to create a model for the detection of conspiracy theories on Telegram. This involves understanding the structural and functional characteristics of Telegram that facilitate the spread of conspiracy theories, identifying influential users or channels, and analyzing messages if they are talking about conspiracy theories.

The significance of this study lies in its potential to enhance social media monitoring, improve public safety, and ensure information integrity. By developing methods to detect conspiracy theories, stakeholders such as policymakers, researchers, and social media platforms can better understand and address the spread of misinformation.

1.4. Report structure

The structure of this thesis proposal is as followed: Chapter 2 provides a generic background on Telegram, conspiracy theories, and the role of social media in spreading these theories. Chapter 3 presents

the literature review, identifying the knowledge gap related to detecting conspiracy theories on Telegram and introducing the sub-research questions. This chapter covers relevant literature on graph theory, machine learning for text classification, and topic modeling, highlighting gaps and research needs specific to Telegram. Chapter 4 describes the operationalization of the research, relating the background information to the research questions. Chapter 5 details the research design and methodology, including data collection methods, graph theory application, machine learning classification, and topic modeling. Chapter 6 presents the results of the study, including descriptive statistics, graph analysis, machine learning classification results, and topic modeling results. Chapter 7 integrates the findings, compares them with existing literature, discusses implications, and addresses study limitations. Finally, Chapter 8 summarizes key findings, contributions, and provides recommendations for future research.

2

Background

This chapter reviews the existing studies and information relevant to the early identification of conspiracy theories on Telegram. The review is divided into three main sections. First, it looks at Telegram's features and why it's popular, focusing on how it supports large group communications and ensures user privacy. Second, it examines how graph theory is used in social media to find influential users and understand how social networks are structured. Third, it discusses different methods used to detect conspiracy theories, including the challenges and current techniques. These sections provide the background needed to understand the tools and methods used in this study, helping to explain how a model can be developed to detect conspiracy theories on Telegram.

2.1. Telegram

Telegram is a cloud-based instant messaging service that provides end-to-end encryption, ensuring users' privacy and security when sending text messages, multimedia files, and conducting voice and video calls (Telegram, n.d.). The platform was founded by Pavel Durov and launched in 2013 (Khanday et al., 2023). It distinguishes itself through a strong emphasis on user privacy, advanced security features, and a wide array of functionalities that cater to both individual and group communications. It acts as a hybrid between social media and instant messengers (Khaund, Shaik, & Agarwal, 2020). This makes Twitter/X and Whatsapp the most similar alternatives. Table 2.1 shows how these platforms differ.

Table 2.1: An overview of the characteristics of Whatsapp, Telegram and Twitter.

Characteristic	**Whatsapp**	**Telegram**	**Twitter**
Release Date	January 2009	August 2013	March 2006
User Base	Over 2 billion (2021)	Over 500 million (2021)	Over 330 million monthly active users (2021)
Clients	iOS, Android, Windows, macOS, Web	iOS, Android, Windows, macOS, Linux, Web	iOS, Android, Windows, macOS, Web
Registration Method	Phone number	Phone number	Email or phone number
Options for Public Chats	Limited (status updates)	Extensive (public groups and channels)	Extensive (tweets and hashtags)
Maximum Members in Public Chats	256 (Group chats)	200,000 (Group chats); Unlimited (Channels)	Unlimited followers; List capacity: 5,000 members
API for Data Collection	Limited and paid (Business API)	Extensive and free (Bot API, Telegram API)	Extensive but paid (Twitter API)
Message Forwarding Option	Yes	Yes	No direct forwarding, but retweets are possible
Encryption	Client-client	Server-client (private and group chats) Client-client (secret chats)	Server-client Client-client (private chats)

Telegram is designed to offer a fast and secure messaging experience. It operates on a decentralized infrastructure, which means that messages are stored on multiple servers around the world, providing users with seamless access to their messages from any device. This design also contributes to the platform's robustness and resistance to censorship (Telegram, n.d.).

Key features of Telegram include:

- End-to-End Encryption: This ensures that messages can only be read by the intended recipients, protecting user data from potential interception.

- Cloud-Based Storage: Users can access their messages and media from any device without the need for local backups.
- Self-Destructing Messages: For heightened privacy, users can set messages to self-destruct after a certain period.
- Cross-Platform Support: Telegram is available on various platforms, including iOS, Android, Windows, macOS, and Linux.

Two of the most powerful features of Telegram are its groups and channels, which facilitate large-scale communication and information distribution.

- **Groups**: Telegram groups allow users to create spaces where multiple participants can interact and share content. Groups can be either public or private. Groups can support up to 200,000 members, making them ideal for community discussions, collaborative projects, and interest-based interactions. Within groups, members can share text messages, photos, videos, and documents, and even conduct polls to gather opinions from the group.
- **Channels**: Channels are used primarily for broadcasting messages to a large audience. Unlike groups, only the channel's administrators can post content, while subscribers can view the posts. Channels can have an unlimited number of subscribers, making them a powerful tool for mass communication. Channels are often used by businesses, news organizations, and public figures to disseminate information and updates to a wide audience. Content posted on channels can be forwarded by subscribers to other users or groups, further increasing the reach of the information. Schäfer and Choi (2023) suggests that channels rather than groups have a much greater impact on creating and spreading information to other Telegram users.

Telegram offers unique features that set it apart from other platforms like WhatsApp and Twitter/X, particularly through its group and channel functionalities. In terms of group messaging, Telegram supports groups with up to 200,000 members, far exceeding WhatsApp's limit of 256 members per group. This scalability makes Telegram ideal for large communities and collaborative projects. On the other hand, Telegram channels are more comparable to Twitter/X, serving as a tool for broadcasting messages to an unlimited number of subscribers. Like Twitter/X's tweets and retweets, Telegram channels allow for wide spread of information, but with the added advantage of message forwarding within the Telegram network. All platforms excel in their respective areas, but Telegram's combination of robust group capabilities and versatile channel broadcasting provides a unique combination that serves to both private group interactions and public information distribution. Table 2.2 shows how the different platforms differ in functionalities.

Table 2.2: An overview of the functionalities in Whatsapp, Telegram and Twitter.

Whatsapp	Telegram	Twitter
Send message to user	Send message to user	Direct message
Send message to group	Send message to user	–
–	Send message to group	Tweet
–	Send message to channel	Read tweets
Forward	Read public message	–
–	User message forwarded to user/group	Direct message
–	Channel message forwarded to user/group	Re-tweet
–	Channel message forwarded to another channel	–
Message reply (private and group chats)	User message forwarded to channel	Message reply (direct and group chats)
–	Message reply (private and group chats)	Comment
–	–	Like
Join group	–	–
	Join group	Follow/unfollow an account
	Subscribe/leave a channel	

2.2. Conspiracy Theories

Conspiracy theories are the belief that certain events or situations are secretly manipulated behind the scenes by powerful forces with negative intent. (European Commission, 2023). They are often presented as a logical explanation of events or situations which are difficult to understand and bring a false sense of control and agency. Conspiracy theories often start as a suspicion. The theories ask

who is benefiting from the event or situation and thus identify the conspirators. Any ‘evidence’ is then linked to fit the theory. Once conspiracy theories have emerged, they can grow quickly. They are hard to debunk because any person who tries to debunk them, is seen as being part of the conspiracy. Conspiracy theories can span a wide range of topics, including politics, health, science, and history, and they often thrive in times of uncertainty and crisis when people seek to make sense of complex and troubling events. Conspiracy theories often target or discriminate against an entire group perceived as the enemy behind a real or imagined threat. They polarise society and fuel violent extremism. While most people who spread conspiracy theories genuinely believe in them, others deploy them cynically to achieve these effects.

Several well-known conspiracy theories have gained prominence in recent years. For example, the ”Great Reset” theory posits that global elites are using the COVID-19 pandemic to reorganize society and take control of the global economy (Robinson, Sardarizadeh, Goodman, Giles, & Williams, 2021). The Great Reset theory is a conspiracy theory that emerged prominently during the COVID-19 pandemic. It is currently one of the most popular conspiracy theories. The theory claims that a global elite, including influential political leaders, business magnates, and international organizations such as the World Economic Forum (WEF), are orchestrating a large-scale plan to restructure and control the global economy and society. The term ”Great Reset” itself originates from the World Economic Forum’s initiative, introduced by its founder Klaus Schwab in June 2020, which calls for rebuilding the global economy sustainably and inclusively in the wake of the pandemic (Schwab, n.d.). The theory often intertwines with other conspiracies, such as those involving vaccine misinformation, the supposed creation of a ”New World Order,” and the manipulation of financial systems for elite gain. It is also the conspiracy theory specifically mentioned by the Dutch intelligence department AIVD as dangerous (Algemene Inlichtingen- en Veiligheidsdienst, 2023). The theory is maybe even most popular in the Netherlands because of Dutch right-wing populist Theory Baudet (Dlewis, 2024).

Another widely circulated theory is the ”anti-vaccine” conspiracy, which claims that vaccines are harmful and part of a plan by pharmaceutical companies and governments to control the population (Douglas, 2021, 2). The spread of such theories can have significant real-world consequences, including undermining public trust in institutions, spreading misinformation, and even inciting violence.

2.3. Telegram and Conspiracy Theories

Telegram has also become a significant platform for the spread of conspiracy theories. Several factors contribute to this:

- Encryption and Privacy: The platform’s strong encryption and privacy policies make it an attractive space for individuals and groups who wish to discuss controversial or fringe ideas without fear of being monitored. This has naturally included communities that propagate conspiracy theories, who find Telegram’s secure environment conducive for their activities (Wischerath et al., 2024).
- Large Group and Channel Capabilities: Telegram supports massive group chats and channels, enabling the rapid spread of information to large audiences. Conspiracy theorists can leverage these features to reach and mobilize followers quickly. The ability to forward messages seamlessly also aids in the viral spread of conspiracy content (Wischerath et al., 2024).
- Lack of Moderation: Compared to platforms like Facebook and Twitter/X, Telegram has relatively lax content moderation policies. Telegram takes down illegal public content within the app, but notes itself that this does not apply to local restrictions on freedom of speech (Telegram, n.d.). This leniency allows for the unchecked spread of misinformation and conspiracy theories.
- Echo Chambers and Confirmation Bias: Telegram’s group and channel structure can create echo chambers where users are exposed predominantly to information that reinforces their existing beliefs (Cinelli et al., 2022). This selective exposure can strengthen confirmation bias, making users more susceptible to accepting conspiracy theories as truth.

There has been done some research on conspiracy theories on Telegram. Simon, Welbers, C. Kroon, and Trilling (2022) used a network approach to understand information flows within the Dutch Telegramsphere. For this they researched which major communities were part of the Dutch Telegramsphere. Hoseini, Melo, Benevenuto, Feldmann, and Zannettou (2023) performed an analysis on the spread of

the QAnon theory through Telegram. Additionally, they used a BERT-based topic modelling (M. Grootendorst, 2022) approach to analyse QAnon messages over multiple languages.

3

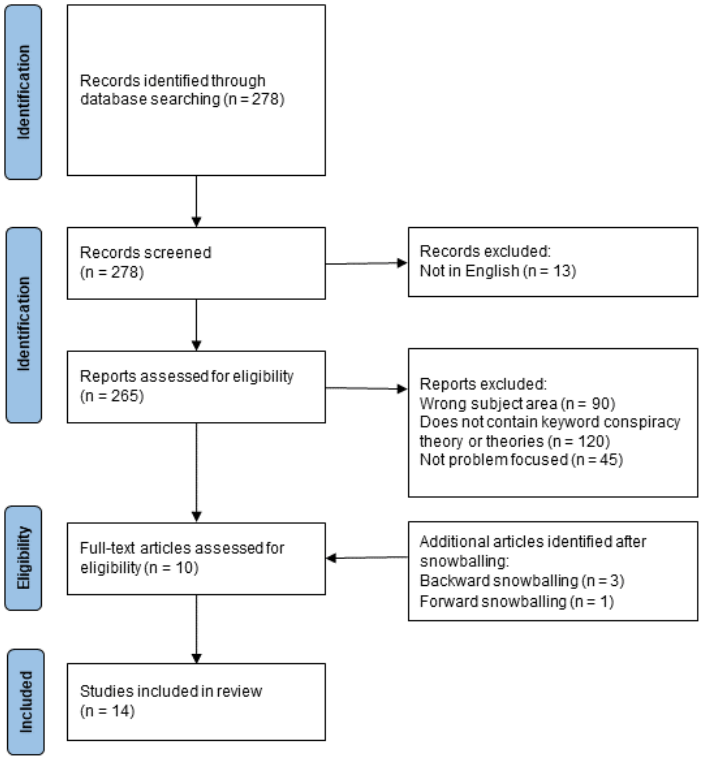
Literature Review

This chapter provides an extensive review of the current academic research on the dissemination of conspiracy theories across social media platforms, with a particular focus on identifying knowledge gaps. The review aims to set the stage for the research by highlighting key findings from previous studies and identifying areas where further investigation is needed. It examines the methodologies and technologies used to detect and analyze conspiracy theories, including the application of graph theory and machine learning techniques. This comprehensive review helps to frame the research questions and informs the methodological choices for the study. The chapter concludes with an identification of the research gap that this study aims to fill and the sub-questions that aim to guide the main research question.

3.1. Knowledge Gap

A literature review was undertaken to examine state-of-the-art academic research on the topic of the spread of conspiracy theories on social media. To conduct a comprehensive literature review, the Scopus database, ACM Digital Library, and IEEE Xplore Digital Library were selected as the primary sources of relevant literature. Scopus is a reputable multidisciplinary database that covers a wide range of scholarly publications. However, many computer science conferences are not indexed by Scopus, which necessitates the inclusion of the ACM Digital Library and IEEE Xplore Digital Library. These additional sources are important for accessing research on conspiracy theories published in computer science conferences, ensuring a thorough and well-rounded review of the literature. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach was employed to ensure a systematic and transparent method of literature selection (Page et al., 2021). A flow diagram of this PRISMA approach can be found in Figure 3.1. The string used for the first stage was *'spread AND conspiracy AND (theory OR theories) AND (social AND media)'*, which returned 278 publications.

Figure 3.1: PRISMA flow diagram of study identification and selection



Following the identification of articles based on the keywords, a series of exclusion steps were applied to narrow down the selection. The exclusion steps were implemented in a sequential manner to ensure the relevance and quality of the included articles. First, articles not in English were excluded, which reduced the number to 265. Next, articles outside the relevant subject areas were excluded. The subject areas were limited to only Social Sciences and Computer Science. The subsequent step involved the exclusion of articles based on specific keywords to further refine the selection. Only articles that had *'conspiracy theory'* or *'conspiracy theories'* as a keyword were included, ensuring a direct relevance to conspiracy theories. Following the exclusion based on keywords, an abstract examination was conducted to evaluate the relevance and alignment of the remaining articles. This brought the selection to 10 publications. Lastly, using the backward and forward snowball technique, the number was increased to a final selection of 14 publications. The final selection of 14 articles is in Table 3.1.

Table 3.1: Final selection of literature

Authors	Title	Year
Theocharis et al.	Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries	2023
Shao et al.	A Dynamic Analysis of Conspiratorial Narratives on Twitter During the Pandemic	2023
Moffitt et al.	Connecting the domains: an investigation of internet domains found in Covid-19 conspiracy tweets	2023
Bodaghi et al.	A Literature Review on Detecting, Verifying, and Mitigating Online Misinformation	2023
Heft & Buehling	Measuring the diffusion of conspiracy theories in digital information ecologies	2022
Walther & McCoy	US Extremism on Telegram: Fueling Disinformation, Conspiracy Theories, and Accelerationism	2021
Moffitt et al.	Hunting Conspiracy Theories During the COVID-19 Pandemic	2021
Shahsavari et al.	Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news	2020
Bruns et al.	'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook	2020
Gruzd & Mai	Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter	2020
Enders et al.	The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation	2023
Kokkera et al.	A Study on the Conspiracy Theory Propagation Network on Twitter	2023
Khanday et al.	Hybrid Approach for Detecting Propagandistic Community and Core Node on Social Networks	2023
Ahmed et al.	COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data	2020

To gain a comprehensive understanding of the current academic research on the spread of conspiracy theories on social media, the selected articles were subjected to detailed analysis. The analysis aimed to identify key findings, themes, and patterns across the literature, providing insights into the governance approaches employed during this technology transition. The results of the analysis are shown in Table 3.2.

Table 3.2: Themes and key findings

Theme	Key findings	References
Detection	Detection techniques are either based on content features, or on user & network features	Bodaghi et al., Khanday et al., Heft & Buehling, Kokkera et al., Ahmed et al.
	Detecting CT early is important in order to mitigate their spread	Shahsavari et al., Bruns et al.
Spread	Social media serve as widespread channels for propagating conspiracy theories and misinformation	Bodaghi et al., Heft & Buehling, Enders et al.
	Frequent social media users are more likely to agree with conspiracy theories and misinformation	Bodaghi et al., Enders et al.
	Emotionally charged conspiratorial narratives are more likely to breed other conspiratorial narratives	Shao et al., Moffitt et al.
	Twitter has a negative effect on conspiracy beliefs—as opposed to all other platforms under examination which are found to have a positive effect.	Theocharis et al.
	Claims that are driven by politics and supported by strong convictions and not science are much harder to mitigate	Gruzd & Mai
Background	The COVID-19 pandemic is the most prominent source of disinformation and conspiracy theories	Walther & McCoy
	Conspiracy theory groups are typically (far)right-winged	Walther & McCoy, Theocharis et al.

The literature review on the detection and propagation of conspiracy theories and misinformation across diverse social media platforms has given some key insights. Detecting conspiracy theories involves a multifaceted approach (Arezo Bodaghi, Schmitt, Watine, & Fung, 2023). Arezo Bodaghi et al. (2023) mention that there are three effective techniques at detecting misinformation:

- Content-based features: using the content like text, images, and videos.
- User & Network features: detecting misinformation by analyzing the role of users in the dissemination of misinformation. All features relating to users and propogation structures fall under this.
- Hybrid techniques: this considers both the content as social features and diffusion structure.

Arezo Bodaghi et al. (2023) also mention papers per technique. Content-based detection are always classifiers trained using Natural Language Processing. However, of all types of misinformation the least amount of attention is paid towards conspiracy-related content. Medina Serrano, Papakyriakopoulos, and Hegelich (2020) used the user comments to identify COVID-19 misinformation videos on Youtube using transformer based models BERT (Devlin, Chang, Lee, Google, & Language, 2019) and RoBERTa (Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, Stoyanov, & Allen, 2019). Cheema (2021) also use transformer based models for detecting conspiracy-related content. They use Vision and Language BERT (ViLBERT) for tweets with images. Additionally, they fine tuned BERT in order to extract textual features.

There were only a few Social and Structure-Based Strategies for misinformation mentioned by Arezo Bodaghi et al. (2023). These were all aimed at rumor detection and that is why it will no be further discussed.

Arezo Bodaghi et al. (2023) mention one Hybrid-Based strategy for detecting conspiracy related content. This paper by S. Shahsavari, Holur, Wang, Tangherlini, and Roychowdhury (2020), uses the assumption that conspiracy theories often link various preexisting domains of human activity through

speculation, typically portrayed as being grounded in the theorist's access to ”hidden knowledge”. They anticipate that the narrative frameworks they construct will exhibit clusters of nodes and edges corresponding to different domains. Given that these clusters are densely connected within themselves and only sparsely connected to other clusters, they then applied community detection algorithms to identify them. For instance, the ”public health” domain showed dense connections among sub-nodes like ”doctors” and ”hospitals,” while having fewer links to the ”telecommunications” domain, which will, in turn, had dense connections among sub-nodes such as ”5G” and ”cell towers.” This crossing across different communities supposedly mirrors the conspiracy theorist's cross-domain exploration in their efforts to create a conspiracy theory. They then validated their theory by connecting the communities to conspiracy theories found in 4Chan.

However, Arezo Bodaghi et al. (2023) did not mention some notable research that has been done on detecting conspiracy theories on Twitter. Kokkera, Chen, Malapati, Demchenko, and Fung (2023) propose a novel scoring algorithm to automatically search for likely conspiracy theory posts based on their connectives. The algorithm uses seed posts that were considered conspiracy spreading. It then uses a score that indicates the likelihood that a tweet is conspiracy related or a user is a conspiracy believer. However, the score is, apart from the amount of new tweets and retweets, heavily based on hashtagging and likes. The latter two are both possible on Telegram, but Telegram users typically barely do this. That is why this algorithm is not deemed applicable for Telegram.

Arezo Bodaghi et al. (2023) mentioned quite some transformer base models for content-based detection techniques. Especially BERT has been used the most, but since the paper another notable transformer based model has been published. CT-BERT is pre-trained on a large amount of COVID-19 related Twitter messages (Zaghouani, Vladimir, & Ruiz, 2023). CT-BERT is specifically designed to be used on particularly social media content, and can be utilized for various natural language processing tasks such as classification, question-answering, and chatbots.

Social media platforms, acknowledged as channels for spreading narratives, reveal a nuanced relationship between platform usage and conspiratorial belief acceptance. Social media serve as widespread channels for propagating conspiracy theories and misinformation by exposing large numbers of individuals to fringe concepts and ultimately finding credulous consumers of information (A. Bodaghi, Schmitt, Watine, & Fung, 2023; Heft & Buehling, 2022). Notably, frequent users demonstrate a higher tendency to align with conspiracy theoriesand misinformation, signifying the impact of prolonged exposure to online content (Enders et al., 2023). The emotional charge within conspiratorial narratives significantly influences their virality, promoting the spread of similar narratives within social media (Moffitt, King, & Carley, 2021; Shao, Hazel Kwon, Walker, & Li, 2023). Intriguingly, Twitter/X stands out by showing a negative impact on conspiracy beliefs—a distinctive contrast to other platforms, which commonly show a positive association with conspiracy theories acceptance (Theocharis et al., 2023). This is a very surprising finding, because a lot of the current research on the spread of conspiracy theories has been focused on Twitter/X.

The challenge of mitigating conspiracy theories is worsened by claims driven more by political motivations and passionate convictions rather than empirical evidence, posing hurdles for intervention strategies (Gruzd & Mai, 2020). Undeniably, the COVID-19 pandemic acted as a catalyst for disinformation and conspiracy theories, dominating discussions and creating numerous false narratives across online platforms (Walther & McCoy, 2021). This shows the criticality of timely intervention. Detecting conspiracy theories is therefore very important in order to mitigate the pervasive spread of conspiracy theories and misinformation within online communities (Bruns, Harrington, & Hurcombe, 2020; S. Shahsavari et al., 2020). Moreover, the review highlights the dominating ideological leanings associated with conspiracy theory groups, predominantly identified as (far)right-winged (Walther & McCoy, 2021), highlighting the ideological foundations that often shape and propagate conspiratorial narratives across digital landscapes (Theocharis et al., 2023).

3.2. Research Questions

A substantial body of research has been dedicated to investigating social media platforms like Twitter/X, particularly regarding the dissemination of misinformation, with a significant focus on COVID-19-related content. However, limited attention has been directed towards studying Telegram, a platform that is

increasingly becoming a breeding ground for diverse conspiracy theories, especially those leaning towards right-wing ideologies. Prior research often employed graph theory and network analysis for conspiracy detection, which presents an opportunity to explore the realm of conspiracy theory detection specifically within Telegram channels. This gap in existing literature underscores the need to investigate the detection methodologies uniquely applicable to Telegram, potentially utilizing graph theory and network analysis techniques, as discussed further in the subsequent chapter dedicated to methodology. This research gap has led to the following research question:

*'**How can Conspiracy Theories be identified (early) on Telegram?**'*

The following four sub-questions were formulated to further lead this research.

Sub-Question 1: "How can graph theory be applied to model the network structure of Telegram channels?"

This sub-question aims to create an initial understanding of how graph theory can be utilized to represent the network structure of Telegram channels. A review will be conducted to model Telegram networks, focusing on nodes and edges representing channels and their interactions. This analysis is primarily related to user and network features, as it examines the structural connections and interactions within the Telegram network.

Sub-Question 2: "Which graph metrics can be used in identifying influential nodes within Telegram networks?"

Building on the insights from the first sub-question, this sub-question will explore specific graph metrics that can be applied to Telegram networks to identify influential nodes. Metrics such as degree centrality, betweenness centrality, and others will be analyzed to determine their effectiveness in highlighting key users or channels that play significant roles in spreading conspiracy theories. Like sub-question 1, this sub-question focuses on user and network features by identifying and analyzing the roles of influential users within the network.

Sub-Question 3: "How can Telegram messages be analyzed to determine if they are conspiracy-related?"

This sub-question focuses on the methods and techniques that can be used to analyze Telegram messages for conspiracy-related content. The research will involve selecting appropriate models or algorithms, preparing the dataset, and adapting these methods to the specific context of detecting conspiracy theories. This sub-question is primarily related to content-based features, as it examines the textual and multimedia content of the messages to determine if they are related to conspiracy theories.

Sub-Question 4: "What new and existing conspiracy theories can be identified on Telegram?"

The final sub-question aims to identify and categorize both new and existing conspiracy theories present in the classified messages. This sub-question will uncover the underlying themes and narratives within the messages, giving insight on the topics of the conspiracy theories circulating on Telegram.

Together, these sub-questions address both user and network features (sub-questions 1 and 2) and content-based features (sub-question 3), integrating them into a hybrid technique for the comprehensive identification of conspiracy theories on Telegram. Sub-question 4 tries to go one step further, by analyzing the conspiracy theories that were identified.

4

Operationalization

This chapter sets the stage for the practical implementation of the research design and methodology. The chapter details how the theoretical concepts and research questions outlined in the previous chapters can be used to come to the methodology used in this research. The methodology chapter and the operationalization chapter serve distinct but complementary roles in this research. The operationalization chapter focuses on the rationale behind the choices made throughout the research process. It delves into the reasoning for selecting specific methods, metrics, and tools, explaining how these decisions align with the research objectives and theoretical foundations. This chapter is where the justifications for methodological approaches are thoroughly explored, ensuring that each choice is grounded in the research's overall goals.

4.1. Research Design

The methodological approach of this study is a quantitative approach to provide a comprehensive analysis of the research question: **"How can conspiracy theories be identified on Telegram?"** Quantitative methods, such as graph theory metrics and machine learning classifiers, offer measurable insights into the structure and content of Telegram messages. Topic modeling then provides context of the conspiracy theories. By integrating these methods, the study aims to address the research question from multiple angles. Detecting conspiracy theories involves a multifaceted approach (Arezo Bodaghi et al., 2023), focusing either on content-based features, examining textual elements and narrative structures, or on user and network-based attributes (Ahmed, Vidal-Alaball, Downing, & López Seguí, 2020; Khanday et al., 2023; Kokkera et al., 2023), emphasizing user behavior and interaction patterns. This research will use a hybrid technique, which considers both the content as the social attributes.

Graph theory is utilized to model the network structure of Telegram channels, as it allows for the identification of influential channels within the network. This is important for understanding how information, and specifically conspiracy theories, disseminate through the platform. Machine learning, particularly fine-tuning transformer-based models, is employed to classify messages as conspiracy-related or not. This method leverages natural language processing techniques to handle the large volume of Telegram messages. Topic modeling, using BERTopic, is used to identify and categorize the underlying themes within the conspiracy related messages. This combination of methods ensures that both the structural and content aspects of the data are thoroughly analyzed.

4.2. Graph theory

4.2.1. Graph theory in social media

Graph theory, a branch of mathematics focusing on the study of graphs, is extensively used in analyzing social media networks. In this context, a graph consists of nodes (representing users) and edges (representing connections or interactions between users). By leveraging graph theory, researchers and analysts can uncover patterns, identify influential users, and understand the structural properties of social networks.

Social media platforms like Facebook, Twitter/X, and Telegram generate vast amounts of data that can be effectively analyzed using graph theory. Here's how graph theory is applied in these contexts:

- Network Structure: Graph theory can help in understanding the overall structure of social media networks. By analyzing the connections between users, we can determine if a network is centralized or decentralized, identify clusters or communities, and observe how information flows within the network (Newman, 2018). For example, a decentralized network might exhibit a more resilient structure against targeted attacks or misinformation spread, whereas a centralized network might be more efficient in disseminating information from a central authority (Pósfai & Barabási, 2016).
- Influential Users: One of the key applications of graph theory is identifying influential users. Metrics such as degree centrality, betweenness centrality, and eigenvector centrality help pinpoint which users have the most connections, act as bridges between different parts of the network, or are connected to other influential users (Bonacich, 2007; Freeman, 1978). For instance, on Twitter/X, users with high betweenness centrality can control the spread of information by acting as intermediaries between different user groups (Kwak, Lee, Park, & Moon, 2010).
- Community Detection: Graph theory techniques like modularity optimization and spectral clustering can identify communities within social media networks. These communities are groups of users who interact more frequently with each other than with the rest of the network. Which can act as an echo chamber, where users only see information that confirms their existing beliefs (Cinelli et al., 2022). Understanding these communities can provide insights into shared interests, behaviors, and the spread of information or misinformation.

Social networks are often modeled as weighted directed graph networks to capture the varying strengths of relationships and interactions between users. In these graphs, nodes represent users, and edges represent the connections or interactions between them. The weights of the edges can signify different aspects of these interactions, such as the frequency of communication (Shadi Shahsavari, Holur, Wang, Tangherlini, & Roychowdhury, 2020), the duration of interactions (XLin, Shang, & Liu, 2014), or the level of trust between users (Ullah & Lee, 2017). For example, in a messaging platform like Telegram, the weight of an edge could represent the number of messages exchanged between two users or the frequency of their interactions. Similarly, weights on nodes can represent user activity levels, such as the number of posts or the overall engagement within the network. In a directed graph, the direction of an edge indicates the flow of interaction from one user to another, such as who initiates a message or who follows whom.

4.2.2. Graph theory applied to Telegram

Graph theory has been mainly applied to Telegram to research and understand the dynamics of online communities. On Telegram, users interact within groups and channels, forming complex networks that can be effectively modeled using graph theory. Researchers use these models to study how communities form, evolve, and interact on the platform.

The nodes of a graph can be users, groups or channels. By representing users or channels as nodes and their interactions as edges, graph theory helps identify clusters or communities where users engage more frequently with each other than with the rest of the network. Examples of applications are researching clone and fake channels (Morgia, Mei, Mongardini, & Wu, 2021), network analysis to gain insights into political discourse and public opinion (Khaund et al., 2020), and studying right-wing and conspiratorial communities (Peeters & Willaert, 2022). All of this research focuses on channels. Channels are often used as central hubs for information distribution, where a single entity (or a small group of administrators) controls the content. This centralized nature makes channels ideal for studying the flow of information, as they serve as focal points from which content is disseminated to users or other channels. This is particularly important in conspiracy theory networks, where certain channels may play a key role in spreading specific narratives. It is a lot harder to fetch groups to research since these groups are private and a lot harder to find than channels. This makes it difficult to obtain a comprehensive dataset for analysis. (Wischerath et al., 2024) have done one of the few researches on a conspiracy-oriented group chat during COVID-19. However, this research is also solely a network analysis and not aimed at node level. With users, groups or channels as nodes, there are two types of relationships possible as an edge:

- Forwarding: An user or a channel can forward an original message (sent by another user or channel) to another user, group or channel. Forwarding in Telegram is similar to the retweet in Twitter/X.
- Mentioning: Users and channels can mention another user or channel in a message. A mention starts with the @ character.

Using mentions as edges in Telegram is quite limited due to how the platform is structured. Mentions are only possible within the same group among users or between different channels. You can't have mentions between different groups, as Telegram doesn't allow cross-group interactions like it does for channels. This limitation means that using mentions as edges in a network analysis is not very effective if you are trying to model broader interactions across various entities on Telegram. While mentions can help analyze how users interact within a single group or across channels, they are not useful for understanding how different groups interact with each other. This reduces their value in creating a complete network analysis. Additionally, Nobari, Reshadatmand, and Neshati (2017) note that typically mentioning is extremely sparse and a mention network consists of several separated connected component which indicates the lower tendency of members to mention. Such a network structure can limit the effectiveness of the analysis, as it may not accurately reflect the true pathways through which information spreads. Consequently, most research aimed at Telegram model a graph the same. Each node represents a **channel** and each edge denoted as $u \rightarrow v$ represents a message published in channel u which mentions the name of channel v, in other words each edge represents a **forwarded** message.

This answers **Sub-Question 1: "How can graph theory be applied to model the network structure of Telegram channels?"** Therefore, it is chosen to model the network structure of Telegram as followed: Each node represents a **channel** u and each edge denoted as $u \rightarrow v$ represents a message published in channel u which mentions the name of channel v. An edge has a weight corresponding to the number of messages published in channel u which mentions the name of channel v within a specified time interval. This results in a weighted direct graph.

Figure 4.1: Telegram network example

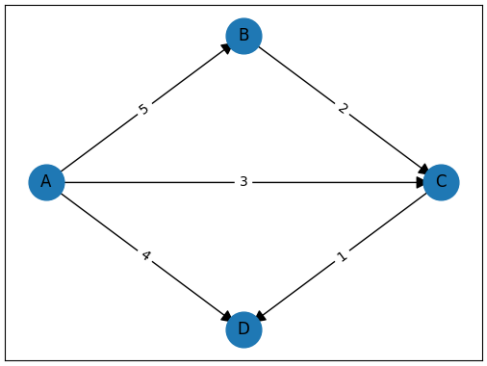


Figure 4.1 shows an example of a simple telegram network structure. For instance, consider the edge from node A to node B with a weight of 5. This weight signifies that channel A has forwarded or mentioned messages from channel B a total of five times. Such an edge weight captures the frequency and strength of the interaction between these two channels, indicating how often information from channel B is propagated through channel A. This can highlight channels that play a crucial role in disseminating content from other channels.

4.2.3. Graph theory for finding influential users

Identifying influential users on social media is crucial, particularly in the context of conspiracy theories, because these users can greatly influence and spread stories within the network. Influential users often act as hubs for information distribution; their posts and interactions reach a broader audience quickly and efficiently. In the context of conspiracy theories, these users can play an important role in spreading misinformation and reinforcing false beliefs (Aïmeur, Amri, & Gilles Brassard, 2023; Shu

et al., 2020). So by identifying influential users, we can monitor their activity to detect early signs of conspiracy theories gaining attention.

In order to effectively identify influential users in the context of spreading conspiracy theories, it is important to use a comprehensive set of metrics that can capture different dimensions of influence within the network. The metrics need to be versatile enough to measure not only how central a user is within their immediate circle of connections but also how pivotal they are in the broader network structure. Furthermore, it is important to account for the nature and impact of the content these users share.

When using the standard node centralities to predict which node will be the top influencer, no single centrality reliably ranks nodes well across networks. Research from Bucur (2020) has shown that combining two or more centralities are instead consistently predictive. This approach integrates **local** centrality metrics, which focus on the size of a node's immediate neighborhood, with **global** centrality metrics that consider a node's influence within the entire network. Examples of a local centrality are the *degree*, *neighbourhood* (the sum of the degrees of direct neighbours), and *two-hop neighbourhood* (the sum of the degrees of neighbours exactly two hops away). Examples of a global centrality are the *closeness* centrality, *betweenness* centrality and *PageRank* centrality.

Local centrality measures

Criteria for Selecting Local Metrics

When choosing metrics to analyze influence or connectivity within a network, especially in a context like social media, certain criteria are important to ensure that the metrics accurately reflect the underlying structure and behavior of the network. For local metrics, which focus on the immediate environment of a channel in a social network, the first criterion is *sensitivity to direct connections*. The metric should be able to capture how well a node is connected to its immediate neighbors, as these direct interactions are often crucial in understanding a node's immediate influence. This is particularly important in networks where the strength and frequency of interactions can vary widely among nodes.

Consideration of connection strength is another key criterion, especially in weighted networks where not all connections are equal. In such networks, it is important that the metric accounts for the weight of each connection, providing a more nuanced view of a node's influence based on the intensity or frequency of interactions with its neighbors.

Directional sensitivity is crucial in directed networks, where the direction of interactions matters. A metric should distinguish between outgoing and incoming connections, offering insights into whether a node is primarily a source of information or a recipient. This distinction is essential in many social media networks where the roles of nodes can be asymmetric.

Finally, *computational efficiency* is a practical consideration, particularly for large-scale networks. Metrics that are computationally simple yet effective are preferred, as they allow for real-time analysis and can be applied to large datasets without excessive computational resources.

Degree centrality

The standard degree centrality states that the degree of a focal node is the number of adjacency's in a network (Freeman, 1978). Let G be a graph with a set of nodes $N(G)$ and a set of edges $E(G)$. The degree centrality $C_D(i)$ for a node i in the graph G is defined as:

$$C_D(i) = \sum_{j \in N(G)} x_{ij} \tag{4.1}$$

where:

- $x_{ij} = 1$ if the edge $(i, j) \in E(G)$, and $x_{ij} = 0$
- otherwise $i, j \in N(G)$.

The degree centrality can be applied to weighted networks with a few changes: Let G be a graph with a set of nodes $N(G)$ and a set of edges $E(G)$. The degree centrality can be extended to weighted networks as follows:

$$C_D^w(i) = \sum_{j \in N(G)} w_{ij} \tag{4.2}$$

where:

- w is the weighted adjacency matrix of the graph G,
- $w_{ij} > 0$ if the edge $(i, j) \in E(G)$, representing the weight of the edge between node i and node j,
- $i, j \in N(G)$.

Opsahl, Agneessens, and Skvoretz (2010) propose a weighted degree centrality for directed networks that distinguishes between the activity and popularity of a node. The **weighted out-degree or activity centrality** $C_{D-out}^{w\alpha}(n)$ measures the influence of node n based on the outgoing connections, while the **weighted in-degree or popularity centrality** $C_{D-in}^{w\alpha}(n)$ measures its influence based on the incoming connections.

Let G be a directed graph, $N(G)$ the set of nodes, and $E(G)$ the set of edges. For each node $n \in N(G)$:

- The **weighted out-degree centrality** is defined as:

$$C_{D-out}^{w\alpha}(n) = k_n^{out} \cdot \left(\frac{s_n^{out}}{k_n^{out}} \right)^{\alpha} \tag{4.3}$$

  Where:

$$k_n^{out} = \sum_{m \in N(G)} x_{nm}, \quad s_n^{out} = \sum_{m \in N(G)} w_{nm} \tag{4.4}$$

- The **weighted in-degree centrality** is defined as:

$$C_{D-in}^{w\alpha}(n) = k_n^{in} \cdot \left(\frac{s_n^{in}}{k_n^{in}} \right)^{\alpha} \tag{4.5}$$

  Where:

$$k_n^{in} = \sum_{m \in N(G)} x_{mn}, \quad s_n^{in} = \sum_{m \in N(G)} w_{mn} \tag{4.6}$$

In these equations:

- k_n^{out} is the out-degree of node n, representing the number of edges originating from node n.
- k_n^{in} is the in-degree of node n, representing the number of edges directed towards node n.
- $x_{nm} = 1$ if there is a directed edge from node n to node m, and 0 otherwise.
- s_n^{out} is the total weight of outgoing edges from node n, summing the weights of all edges that originate from n.
- s_n^{in} is the total weight of incoming edges to node n, summing the weights of all edges directed towards n.
- w_{nm} is the weight of the edge from node n to node m.

Opsahl et al. (2010) also use a tuning parameter α, which determines the relative importance of the number of ties compared to tie weights. $\alpha \in [0, 1]$ indicates weak ties and $\alpha > 1$ indicates strong ties. The parameter α is used to adjust the emphasis on the total weight of the edges relative to their degree:

- If α is set to 1, the metric becomes a simple weighted degree centrality, giving equal importance to the number of adjacent edges and their weights.
- If α is greater than 1, it places more emphasis on the total weight of the edges. This means that a node with fewer but heavier weighted connections will score higher.
- If α is less than 1, it places more emphasis on the number of the edges. This means that a node with many connections, regardless of their weights, will score higher.

For example, consider node A and node B of figure 4.1. For node A, the out-degree (k_A^{out}) is 3, as there are three outgoing edges from A to nodes B, C, and D. The total outgoing weight (s_A^{out}) is the sum of the weights of these edges, resulting in $s_A^{out} = 5 + 3 + 4 = 12$.

For node B, the out-degree (k_B^{out}) is 1, as there is only one outgoing edge from B to node C. The total outgoing weight (s_B^{out}) is the weight of this single edge, resulting in $s_B^{out} = 2$.

For $\alpha = 1$:

- Node A:

$$C_{D-out}^{w1}(A) = 3 \cdot \left(\frac{12}{3}\right)^1 = 3 \cdot 4 = 12$$

- Node B:

$$C_{D-out}^{w1}(B) = 1 \cdot \left(\frac{2}{1}\right)^1 = 1 \cdot 2 = 2$$

For $\alpha = 2$:

- Node A:

$$C_{D-out}^{w2}(A) = 3 \cdot \left(\frac{12}{3}\right)^2 = 3 \cdot 16 = 48$$

- Node B:

$$C_{D-out}^{w2}(B) = 1 \cdot \left(\frac{2}{1}\right)^2 = 1 \cdot 4 = 4$$

By comparing these results, we can see how the centrality scores change for different values of α. When $\alpha = 1$, the centrality of node A is significantly higher than node B because A has more outgoing connections and a higher total weight. As α increases, the emphasis on the weight of the connections becomes stronger, further increasing the disparity between A and B. By adjusting α, we can control the influence of the edge weights in our centrality measure, thereby capturing different aspects of node importance in the network. This flexibility allows for a more accurate representation of the node's role in the dissemination of information, particularly in networks like Telegram where message forwarding is a key activity. The transfer and sharing of tacit knowledge requires strong ties (Hansen, Baratoff, & Neumann, 1999). For believers, conspiracy theories can be considered as a form of tacit knowledge. Therefore, when researching within echo chambers of believers of conspiracy theories, an α greater than 1 is more appropriate.

Neighborhood centrality

Neighborhood centrality of a node in an undirected graph measures the sum of the degrees of its direct neighbors (Bucur, 2020). It provides an indication of how well-connected the neighbors of a node are, which can reflect the node's potential to access information through its immediate connections (Rose, Opolot, & Georg, 2022).

Neighborhood centrality measures the connectivity of a node's immediate neighbors. A higher value indicates that the node's neighbors are well-connected, which may enhance the node's ability to access and spread information. In a weighted directed network, neighborhood centrality is adapted to account for the direction and weight of edges. It sums the degrees of the neighbors, weighted by the edge weights. This provides a measure of how influential a node is based on the connectivity of its neighbors, considering both incoming and outgoing connections.

Let G be a graph with a set of nodes $N(G)$ and a set of edges $E(G)$. The neighborhood centrality $C_N(i)$ of a node i is given by:

$$C_N(i) = \sum_{(i,j) \in E(G)} C_D(j)$$

where $C_D(j)$ is the degree centrality of node j, as defined in Equation 4.1.

To adapt neighborhood centrality for a weighted directed network, we consider both the directionality and the weights of the edges. This involves summing the weighted degrees (in-degree and out-degree) of the neighbors.

Weighted Out-Degree Neighborhood Centrality:

$$C^w_{N-out}(i) = \sum_{(i,j)\in E(G)} \sum_{(j,k)\in E(G)} w_{jk} \tag{4.7}$$

where:

- $(i, j) \in E(G)$ denotes an edge from node i to node j,
- w_{jk} is the weight of the edge from node j to node k.

Weighted In-Degree Neighborhood Centrality:
This metric measures the sum of the weighted in-degrees of the in-neighbors of a node i in a directed graph $G = (N(G), E(G))$:

$$C^w_{N-in}(i) = \sum_{(j,i)\in E(G)} \sum_{(k,j)\in E(G)} w_{kj} \tag{4.8}$$

where:

- $(j, i) \in E(G)$ denotes an edge from node j to node i,
- w_{kj} is the weight of the edge from node k to node j.

Combined Weighted Neighborhood Centrality: To get a comprehensive measure, we can combine the weighted in-degree and out-degree neighborhood centralities:

$$C^w_N(i) = \alpha \cdot C^w_{N-in}(i) + \beta \cdot C^w_{N-out}(i)$$

where:

- α and β are parameters that balance the influence of in-neighbors and out-neighbors, respectively.

Here is an example calculation of the degree centrality for node A of graph 4.1 once again shown below:

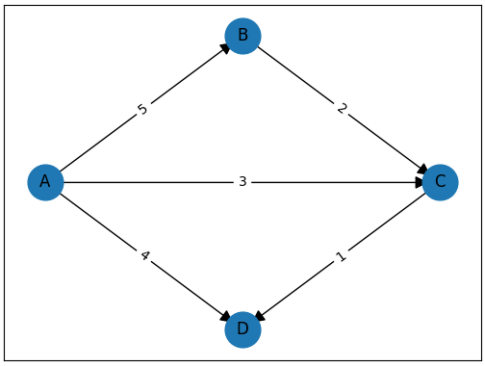


Table 4.1: Neighborhood Centrality Measures for Node A

Measure	Formula	Calculation	Result
Out-Degree Neighborhood Centrality	$C_{N-out}(A)$	Number of out-neighbors of A	3
Weighted Out-Degree Neighborhood Centrality	$C^w_{N-out}(A)$	$5 + 3 + 4$	12
In-Degree Neighborhood Centrality	$C_{N-in}(A)$	No in-neighbors	0
Weighted In-Degree Neighborhood Centrality	$C^w_{N-in}(A)$	No incoming edges	0

Node A has a high out-degree centrality, indicating it has many outgoing connections to other nodes, which signifies its active role in disseminating information. Its weighted out-degree centrality is even higher, emphasizing the strength or frequency of these connections.

Two-Hop Neighborhood centrality
Bucur (2020) also mentions the Two-Hop Neighborhood centrality as another option for a local centrality. The Two-Hop Neighborhood centrality is similar to the normal Neighborhood centrality only now exactly two hops away. Let G be a graph with a set of nodes $N(G)$ and a set of edges $E(G)$. The Two-Hop neighborhood centrality $C_{N-2hop}(i)$ of a node i is given by:

$$C_{N-2hop}(i) = \sum_{(i,j)\in E(G)} \sum_{(j,k)\in E(G)} C_D(k)$$

where $C_D(k)$ is the degree centrality of node k, as defined in Equation 4.1.

However, Bucur (2020) shows that the Two-Hop Neighborhood centrality performs worse than the normal version. Additionally, the Two-Hop version looks beyond the immediate neighbors to those two steps away. While this can be useful in some contexts, it might dilute the focus on the most direct influences, which are typically more relevant in understanding immediate local effects. The global centrality measures already serve for a broader perspective.

Evaluating Local Metrics

When considering *sensitivity to direct connections*, weighted degree centrality stands out as the most focused metric. It directly counts the number of connections (edges) a channel has, with additional emphasis on the strength of these connections. This makes it particularly useful in networks where the immediate connections of a node are crucial for understanding its influence. On the other hand, neighborhood centrality broadens this view by also accounting for the connections of neighboring channels. While this provides a more comprehensive view of a channel's local influence, it can dilute the emphasis on direct connections. The two-hop neighborhood centrality further extends this by including nodes two steps away, which, while useful for capturing a wider influence, might shift focus away from the direct interactions that are often most indicative of a channel's immediate impact.

In terms of accounting for the *strength of connections*, weighted degree centrality again proves to be the most straightforward and direct. It explicitly considers the weight of each connection, making it highly relevant in networks where the intensity or frequency of interactions matters. Neighborhood centrality also considers connection strength but does so in a more indirect manner, by incorporating the weights of a node's neighbors' connections. This can provide a richer context but may not emphasize the strength of direct connections as effectively as weighted degree centrality. Two-hop neighborhood centrality can capture even broader connection strengths, but the focus on distant connections might make it less relevant if the main interest lies in the immediate interactions of a node.

When *directionality* is an important factor, weighted degree centrality is highly adaptable, distinguishing clearly between incoming and outgoing connections. This makes it particularly versatile in directed networks. Neighborhood centrality can also account for directionality, but this adds complexity, especially when considering the directionality of neighbors' connections. Two-hop neighborhood centrality becomes even more complex in directed networks, as it needs to account for directionality two steps away, which can complicate the analysis and interpretation of the results.

From a *computational* perspective, weighted degree centrality is the most efficient, even in large networks. Its simplicity makes it ideal for real-time analysis or applications involving extensive datasets. Neighborhood centrality is more computationally intensive due to the additional layer of considering neighbors' connections, which can be more complex in larger networks. Two-hop neighborhood centrality is the most computationally demanding of the three, as it requires evaluating connections up to two steps away, which can significantly increase the complexity and computational load, particularly in large or dense networks.

Conclusion
In this research, degree centrality is chosen over both neighborhood centrality measures for analyzing the network structure of Telegram channels on a local level. Degree centrality offers simplicity and direct

interpretability, making it an ideal metric for identifying influential channels. By counting the number of direct connections a node has, degree centrality provides a straightforward measure of a channel's activity and popularity. This simplicity is particularly valuable in the context of Telegram networks, where the goal is to quickly identify key nodes that play a significant role in spreading information. On the other hand, neighborhood centrality, which measures the sum of the degrees of a node's neighbors, introduces additional complexity. While it provides insights into the broader connectivity of a node's immediate environment, it can make the results less intuitive. Furthermore, degree centrality aligns better with the research objectives by directly measuring the immediate influence of a node, which is crucial for understanding how conspiracy theories spread through direct message forwarding. Additionally, degree centrality is computationally efficient, making it more suitable for large-scale networks, whereas for example the two-hop neighborhood centrality requires a lot more intensive calculations, potentially slowing down the analysis. Overall, degree centrality provides a more practical and effective measure for identifying the most influential channels on a local level within a Telegram network.

Global centrality measures

Criteria for Selecting Global Metrics

Global metrics focus on the overall importance or influence of a node within the entire network. When selecting global metrics, one must consider the *ability to capture network-wide influence*. The metric should effectively identify nodes that play crucial roles in the broader network structure, such as those that connect different parts of the network or facilitate the flow of information across the network.

Dependence on network paths is another important criterion. Metrics that consider a node's position relative to the paths between other nodes can highlight nodes that are strategically important in the network. For example, nodes that lie on the shortest paths between many other nodes might control the flow of information, making them particularly influential.

In weighted and directed networks, it is important that the global metric takes into account the *weight and direction of connections*. This ensures that the metric accurately reflects the varying levels of influence across different connections, providing a more detailed understanding of the network dynamics.

Robustness to network changes is another crucial factor. The metric should remain meaningful even as the network evolves, such as when new nodes or connections are added. This stability ensures that the insights gained from the metric are reliable over time.

Closeness Centrality

This measures how close a node is to all other nodes in the network. It is defined as the reciprocal of the sum of the shortest path distances from the node to all other nodes:

$$C_{clo}(i) = \frac{|N(G)| - 1}{\sum_{j \in N(G), j \neq i} \text{dist}(i, j)}$$

where:

- $|N(G)|$ is the total number of nodes in the graph G,
- dist(i, j) is the shortest path distance from node i to node j,
- $(i, j) \in E(G)$ indicates that i and j are connected by an edge in the graph G.

In a weighted directed network, the shortest path distance takes into account both the direction and the weights of the edges. This however can cause problems for networks that are disconnected. If a node cannot be reached from another node, the shortest path distance is considered infinite. This can make the calculation of closeness centrality problematic as it would involve division by infinity. This problem will be illustrated as followed: Suppose we want to calculate the closeness centrality for node B of figure 4.1.

Shortest Path Calculation:

- From B to C: Direct path $B \to C$ (weight 2)
- From B to D: Path $B \to C \to D$ (weight 2 + 1 = 3)
- From B to A: No path exists (unreachable)

Sum of Shortest Path Distances:

$$\sum_{m \in N, m \neq B} d(B, m) = d(B, C) + d(B, D) + d(B, A)$$

Substituting the values:

$$= 2 + 3 + \infty = \infty$$

Closeness Centrality Calculation:

$$C_{clo}(B) = \frac{|N| - 1}{\sum_{m \in N, m \neq B} d(B, m)}$$

In this graph, $|N| = 4$:

$$C_{clo}(B) = \frac{4 - 1}{\infty} = 0$$

In this example, node B has a closeness centrality of 0 due to the presence of unreachable nodes. This illustrates the problem of applying closeness centrality in a weighted directed network, especially when the network is not strongly connected. The centrality value becomes zero when some nodes are unreachable, making it an unreliable measure in such cases.

Using closeness centrality in a network where the edge attribute represents the number of messages exchanged between nodes can be problematic due to the metric's assumptions about edge weights. Closeness centrality typically interprets edge weights as distances or costs, with lower weights indicating stronger or closer connections. However, when the edge weight reflects the number of messages, higher weights signify stronger connections, which contradicts the typical interpretation. Additionally, when using networkx to calculate shortest paths, non-existing edges are automatically given an edge weight of 1. If non-existing edges are manually assigned a weight of zero to represent no connection, inverting these weights to reflect that a higher weight indicates a stronger connection can result in division by zero, leading to infinite values. This disrupts the centrality calculation and can produce misleading results.

Betweenness centrality
This measures the extent to which a node lies on the shortest paths between other nodes in the network. It quantifies the role of a node as a bridge or intermediary in the network, highlighting its potential to control information flow. For a node n, the betweenness centrality $C_{betw}(n)$ is given by:

$$C_{betw}(n) = \frac{1}{(|N| - 1)(|N| - 2)} \sum_{j,k \in N \setminus \{n\}, j \neq k} \frac{\sigma_{jk}(n)}{\sigma_{jk}} \tag{4.9}$$

where:

- $|N|$ is the total number of nodes in the network,
- σ_{jk} is the total number of shortest paths from node j to node k,
- $\sigma_{jk}(n)$ is the number of those paths that pass through node n.

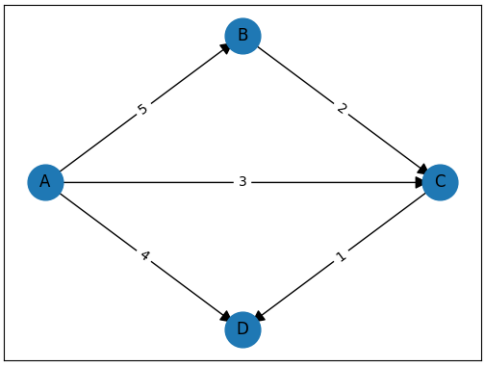
Normally, the betweenness centrality considers the shortest path to measure how close a node is to all other nodes, typically interpreting edge weights as distances or costs, where lower weights mean closer or more direct connections. However, in a network where higher weights are considered beneficial or indicative of stronger ties, one approach is to invert the weights for the purpose of calculating path lengths. This involves transforming the weights such that higher original weights result in lower "distance" values for the centrality calculations. This way of "inversing" the weight makes sure that the relative differences between the weights are kept. This approach involves normalizing the weights to a [0, 1] range, then inverting them, and finally scaling them back to the original range while preserving the relative differences.

$$\text{normalized\_weight}(w) = \frac{w - w_{\text{min}}}{w_{\text{max}} - w_{\text{min}}} \tag{4.10}$$

$$W_{\text{inv\_norm}}(w) = 1 - \text{normalized\_weight}(w) \tag{4.11}$$

$$\text{scaled\_inverted\_weight}(w) = w_{\text{min}} + W_{\text{inv\_norm}}(w) \cdot (w_{\text{max}} - w_{\text{min}}) \tag{4.12}$$

The following is an example calculation of the betweenness centrality of node B of figure 4.1:



Consider the network in Figure 4.1 with the following edge weights:

Table 4.2: Normalized, Inverted, and Scaled Weights for Betweenness Centrality Calculation

Edge	Original Weight	Normalized Weight	Inverted Normalized Weight	Scaled Inverted Weight
$A \to B$	5	1.0	0.0	1.0
$A \to C$	3	0.5	0.5	2.5
$A \to D$	4	0.75	0.25	2.0
$B \to C$	2	0.25	0.75	3.5
$C \to D$	1	0.0	1.0	4.0

Using these scaled inverted weights, the betweenness centrality of node B is calculated as follows:

Table 4.3: Calculation of Betweenness Centrality for Node B Using Scaled Inverted Weights

Node Pair	Shortest Path	Path Length	Does Path Pass Through B?
$A \to C$	$A \to C$	2.5	No
$A \to D$	$A \to D$	2.0	No
$B \to D$	$B \to C \to D$	7.5	Yes

The betweenness centrality of node B is:

$$C_{betw}(B) = \frac{1}{(4-1)(4-2)} \times 1 = \frac{1}{6} = 0.167$$

PageRank

PageRank is an algorithm originally used by Google to rank web pages in their search engine results. It assigns a numerical weight to each node in a network, with the intent to measure its relative importance within the network. The idea is that more important nodes are likely to receive more incoming links from other important nodes. The PageRank algorithm typically uses an iterative approach to calculate the PageRank values until they converge to stable values.

The PageRank $PR(i)$ of a node i in a directed graph G is defined as:

$$PR(i) = \frac{1-d}{|N(G)|} + d \sum_{(j,i)\in E(G)} \frac{PR(j)}{C_{D-out}(j)}$$

where:

- d is the damping factor, typically set to 0.85.
- (j, i) indicates an edge from node j to node i.
- $k^{out}(j)$ is the out-degree of node j, representing the number of edges originating from node j.

PageRank is a popular centrality measure that evaluates the importance of nodes in a network based on the structure of incoming links. However, one significant limitation of the PageRank algorithm is that it does not consider edge weights. In the standard PageRank formula, each link from node j to node i contributes equally to the rank of node ii, regardless of the strength or frequency of the connection between them. This can be a drawback in scenarios where edge weights represent important information, such as the number of forwarded messages or the strength of interactions in a network.

In the context of analyzing Telegram channels, where edge weights indicate the frequency of message forwarding, ignoring these weights could lead to misleading conclusions. For instance, a channel that frequently forwards messages might have a stronger influence in the network than a channel with a similar number of connections but fewer forwarded messages. Since PageRank treats all links equally, it would fail to capture this nuance, making it unsuitable for my analysis, which requires a measure that accurately reflects the weighted nature of connections between channels.

Evaluating Global Metrics

When assessing the *ability to capture network-wide influence*, closeness centrality is particularly effective as it measures the average distance from a node to all other nodes, making it valuable for identifying nodes that can quickly disseminate information across the network. Betweenness centrality captures nodes that serve as bridges between different parts of the network, emphasizing their role in controlling the flow of information. PageRank, however, assesses a node's influence based on the connections it has and the influence of those connections, but its inability to incorporate edge weights limits its applicability in networks where the strength of connections is crucial.

In terms of *reliance on network paths*, closeness centrality depends heavily on the shortest paths, making it sensitive to changes in the network structure. This sensitivity allows it to effectively identify nodes that can quickly reach others. Betweenness centrality also relies on shortest paths but focuses on nodes that lie on many paths, highlighting their role as intermediaries. PageRank does not rely on specific paths but rather on the overall structure, making it less sensitive to specific changes but perhaps missing nuanced roles in the network's structure.

Closeness centrality can be adapted for *weighted networks*, but this adaptation is complex when weights are not treated as distances. Betweenness centrality can incorporate edge weights by treating them as distances, making it more suitable for weighted networks, though the interpretation of weights requires careful consideration. PageRank does not account for edge weights, which is a significant limitation in scenarios where the strength of connections plays a crucial role in determining influence.

Regarding *robustness to network changes*, closeness centrality is sensitive to changes in the network, as it relies on the shortest paths, which can vary significantly with the addition or removal of nodes or edges. Betweenness centrality is also sensitive to changes, particularly those that affect the shortest paths, but it is robust in identifying key nodes in static networks. PageRank is more robust to changes, as it considers the overall structure rather than specific paths, making it stable in dynamic networks but potentially less responsive to changes in edge weights or node roles.

Conclusion

In this research is the betweenness centrality chosen over closeness and Pagerank. Closeness centrality, while useful in some contexts, assumes a strongly connected graph where every node can be reached from any other node, which is not always the case in a weighted directed network like Telegram. This assumption can lead to misleading or undefined centrality values for nodes that are isolated or have limited connectivity. On the other hand, PageRank centrality, despite its popularity for ranking

web pages, was found to be unsuitable for Telegram channels when the network is modeled using channels (Nobari et al., 2017). Additionally, PageRank does not consider edge weights. Betweenness centrality, however, offers a distinct advantage by highlighting nodes that act as bridges or intermediaries in the network. This measure identifies channels that play a crucial role in controlling the flow of information, making it particularly relevant for detecting influential channels in the spread of conspiracy theories. By focusing on the nodes that lie on the shortest paths between other nodes, betweenness centrality captures the strategic positions of channels within the network, providing a robust metric for the analysis.

Viral messages

Dargahi Nobari, Sarraf, Neshati, and Erfanian Daneshvar (2021) have proposed a metric specific for Telegram. They propose a metric for viral messages. This is the formal definition: a published message in a Telegram channel is called *viral* if the associated view counter in steady state is $\alpha\%$ higher than the average of other posts in a vicinity of n posts (Dargahi Nobari et al., 2021). α and n are the parameters. An $\alpha = 10\%$ and and $n = 1$ is chosen similar to the setup of Dargahi Nobari et al. (2021).

$$C_{viral}(n) = \text{Number of viral messages from node } n \tag{4.13}$$

This metric helps identify channels that produce content with significantly higher engagement compared to surrounding posts. By focusing on the number of viral messages, is it possible to measure the ability of a channel to generate high-interest content that spreads widely within a Telegram network. The local and global metrics discussed offer a solid foundation for understanding the structural aspects of the network, but they may miss elements related to the content and context of interactions. Viral Messages Centrality fills this gap by focusing on the spread of specific messages or content, capturing the impact of a node based on the virality of the content it propagates. This metric is particularly valuable in networks where the structure and the content of interactions are equally important for understanding influence. By integrating content-based impact with structural analysis, Viral Messages Centrality provides a more comprehensive view, making it a good addition to the overall assessment of channel influence.

Example: Consider a Telegram channel C that publishes a series of messages. We monitor a segment of five consecutive messages to determine if any are viral. Suppose the view counts of these messages are as follows:

- Message 1: 100 views
- Message 2: 110 views
- Message 3: 150 views
- Message 4: 120 views
- Message 5: 115 views

To determine if a message is viral, we calculate the average view count of the messages in its vicinity. For Message 3, the vicinity includes Messages 2 and 4 (since $n = 1$):

$$\text{Average view count of vicinity} = \frac{110 + 120}{2} = 115$$

Message 3 is considered viral if its view count is at least $\alpha\%$ higher than the average view count of its vicinity:

$$150 > 115 \cdot (1 + \frac{10}{100}) = 115 \cdot 1.10 = 126.5$$

Since 150 is greater than 126.5, Message 3 is classified as viral. Therefore, channel C would have a viral message centrality $C_{viral}(C) = 1$ for this segment. By applying this metric across all messages in a channel, we can identify channels with a high propensity for producing viral content. This, in turn, helps in assessing the influence and engagement levels of different Telegram channels within the network.

Conclusion

To address **Sub-Question 2: "Which graph metrics can be used in identifying influential nodes within Telegram networks?"**, we employ a combination of centrality measures that capture different aspects of node influence in the network. The two weighted degree centralities—activity centrality and popularity centrality—measure the direct (local) connections and the strength of those connections for each node. This helps in understanding which users are highly active or popular within smaller clusters of the network, potentially identifying key figures who might initiate or amplify conspiracy theories among a close-knit group of followers. Activity centrality ($\hat{C}^{w\alpha}_{D-out}(n)$) focuses on the number and weight of outgoing edges, indicating the extent to which a channel disseminates information to other channels. Popularity centrality ($\hat{C}^{w\alpha}_{D-in}(n)$) examines incoming edges, reflecting how often a channel is referenced or forwarded by others, thus capturing its popularity and reach.

Betweenness centrality ($C_{betw}(n)$) provides a global perspective by identifying nodes that act as bridges within the network. Nodes with high betweenness centrality lie on many shortest paths between other nodes, indicating their strategic importance in facilitating information flow and connecting different parts of the network. This measure provide insights into a node's influence across the entire network. Such nodes are crucial in the context of conspiracy theories, as they can propagate ideas from one part of the network to another, increasing the reach of these theories.

Additionally, we include the viral messages centrality ($C_{viral}(n)$), which specifically accounts for the ability of a channel to produce content that significantly outperforms neighboring posts in terms of views. This metric captures the content-related influence of a channel, highlighting those that can generate viral content and engage a larger audience. This is particularly relevant for detecting conspiracy theories, as viral messages often signal the spread of compelling but potentially misleading narratives.

4.2.4. Machine Learning Classification

As mentioned in Chapter 3 did Arezo Bodaghi et al. (2023) mention quite some transformer base models for content-based detection techniques. Especially BERT has been used the most.

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based model developed by Google (Devlin et al., 2019). It is designed to understand the context of words in a sentence by considering the entire sentence, both left and right context, during training. This bidirectional approach enables BERT to capture nuanced meanings and relationships between words, making it highly effective for various natural language processing (NLP) tasks.

BERT is built on the transformer architecture, which consists of an encoder-decoder structure (Vaswani et al., 2017). However, BERT uses only the encoder part. The encoder is composed of multiple layers of self-attention mechanisms and feed-forward neural networks. The self-attention mechanism allows BERT to weigh the importance of different words in a sentence, capturing long-range dependencies and contextual information effectively. This architecture allows BERT to perform tasks such as text classification, question answering, and named entity recognition with high accuracy.

BERT undergoes a two-step training process: pre-training and fine-tuning. During pre-training, BERT is trained on a large corpus of text to learn general language representations. This involves two tasks: Masked Language Modeling (MLM), where random words in a sentence are masked and the model learns to predict them, and Next Sentence Prediction (NSP), where the model learns the relationship between pairs of sentences (Devlin et al., 2019). After pre-training, BERT can be fine-tuned for specific tasks by adding a simple classification layer and training the model on task-specific data. This fine-tuning process allows BERT to adapt its general language understanding to perform well on specific NLP tasks.

Several variants of BERT have been developed to extend its capabilities. These include:

- RoBERTa (Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, & Stoyanov, 2019): An optimized version of BERT with improved training methodology.
- ALBERT (Lan et al., 2020): A lighter version of BERT with fewer parameters, making it more efficient.
- DistilBERT (Sanh, Debut, Chaumond, & Wolf, 2020): A smaller, faster, and cheaper version of BERT that retains 97% of BERT's language understanding.

- CT-BERT (Zaghouani et al., 2023): A version of BERT pre-trained on COVID-19 related tweets, specialized for pandemic-related content.
- mBERT (Multilingual BERT) (Pires, Schlinger, & Garrette, 2019): A version of BERT trained on 104 languages, designed to handle multilingual text.

In the context of content-based detection techniques, BERT's ability to understand the contextual meaning of words makes it a powerful tool for detecting conspiracy theories in text. When applied to Telegram messages, BERT can analyze the content of messages to identify patterns and phrases commonly associated with conspiracy theories. This involves fine-tuning BERT on a labeled dataset where messages are annotated as conspiracy-related or not, enabling the model to learn the characteristics of conspiracy-related content.

For example, in this study, mBERT, a variant of BERT trained on 104 languages, is used due to the multilingual nature of the messages, which include both Dutch and English. This choice ensures that the model can effectively handle and understand messages in multiple languages. The fine-tuning process involves using a manually labeled dataset of Telegram messages labeled as conspiracy-related or not. This allows mBERT to adapt to the specific language and context of conspiracy theories in Telegram messages. By leveraging BERT's contextual understanding, the fine-tuned model can accurately classify messages as conspiracy-related, even when the language is subtle or indirect.

In summary, BERT's transformer-based architecture and bidirectional context understanding make it highly effective for content-based detection techniques. Its ability to capture nuanced meanings and relationships between words enables it to identify conspiracy theories in text, providing a robust tool for analyzing and detecting misinformation on platforms like Telegram.

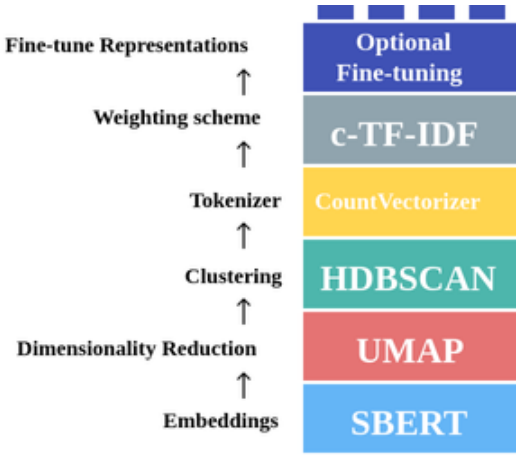
4.2.5. Topic Modeling

Topic modeling is an unsupervised machine learning technique used to identify and extract the underlying themes or topics present in a large collection of text documents (Vayansky & Kumar, 2020). This technique helps in organizing, understanding, and summarizing vast textual data by discovering hidden semantic structures within the text. One of the most commonly used algorithms for topic modeling is Latent Dirichlet Allocation (LDA), which assumes that each document is a mixture of topics and each topic is a mixture of words (Blei, Ng, & Jordan, 2003). By iteratively refining these mixtures, LDA can uncover the distribution of topics across documents and the distribution of words within each topic. Topic modeling is widely used in various applications, such as document classification, recommendation systems, and information retrieval, providing insights into the content and themes present in large texts.

BERTopic

BERTopic is an advanced topic modeling technique that leverages the power of transformer-based language models, such as BERT (Bidirectional Encoder Representations from Transformers), to generate dense embeddings of documents (M. Grootendorst, 2022). These embeddings capture the contextual information and semantic relationships between words more effectively than traditional bag-of-words topic modeling approaches. Figure 4.2 shows an overview of the sequence of the steps that BERTopic takes to create its topic representations. BERTopic assumes some independence between steps, which allows for some modularity.

Figure 4.2: Sequence of steps of BERTopic (M. P. Grootendorst, n.d.)



1. Embeddings: The first step in BERTopic is to convert the text data into numerical representations. This is achieved using sentence-transformers (SBERT), which generate dense vector embeddings for each document. These embeddings capture the semantic meaning of the text, allowing BERTopic to understand the context and relationships between words more effectively than traditional methods.
2. Dimensionality Reduction: Given the high dimensionality of the embeddings produced by BERT, BERTopic employs Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of the data. UMAP helps in visualizing and clustering the data by preserving the global structure while maintaining the local relationships within the data (McInnes, Healy, & Melville, 2020). Although BERTopic uses UMAP as a default to perform the dimensionality reduction, it is also possible to use PCA.
3. Clustering: After reducing the dimensionality, BERTopic uses Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to cluster the embeddings. HDBSCAN is a clustering algorithm that identifies dense regions in the data, allowing for the discovery of clusters of documents that share similar topics (Campello, Moulavi, & Sander, 2013). This step is crucial for grouping documents into coherent topics.
4. Tokenizer: The next step plays a crucial role in extracting the representative keywords for each topic. The quality of the topic representations is crucial for interpreting the topics. While the embeddings generated by transformer models capture the semantic meaning of the text, the CountVectorizer is used to identify and rank the most important words within each topic cluster. This process involves several key steps. Firstly, the text data is tokenized, breaking down the text into individual words or tokens. Following tokenization, the CountVectorizer counts the occurrences of each token in the documents, thereby creating a frequency count for each word. This results in the formation of a document-term matrix, where each row represents a document and each column corresponds to a word from the vocabulary, with the cell values indicating the frequency of the words in the respective documents. This matrix serves as the basis for further analysis, allowing BERTopic to identify and rank the most frequent and relevant words within each topic cluster.
5. Weighting scheme: In BERTopic, c-TF-IDF (class-based Term Frequency-Inverse Document Frequency) plays an important role in refining the results from the initial token counts generated by the CountVectorizer. Unlike traditional TF-IDF, which is applied at the document level, c-TF-IDF aggregates the text data within each topic cluster, treating all documents within a cluster as a single ”class” or ”document.” This helps in capturing the collective importance of terms within each cluster. Essentially, c-TF-IDF highlights the most representative and distinctive words for each topic cluster, providing a weighted representation of terms that enhances the coherence and interpretability of the identified topics.
6. Fine-tune Representations: BERTopic offers a few representation models that allow for further

fine-tuning of the topic representations. These are optional and are not used by default. This step also allows the use of Large Language Models (LLM) to further fine-tune topics to generate labels and summaries of topics. This can be done by passing the keywords of a topic to the LLM and asking to generate a label that fits the topic the best. This last step can be used to identify known conspiracy theories.

5

Methodology

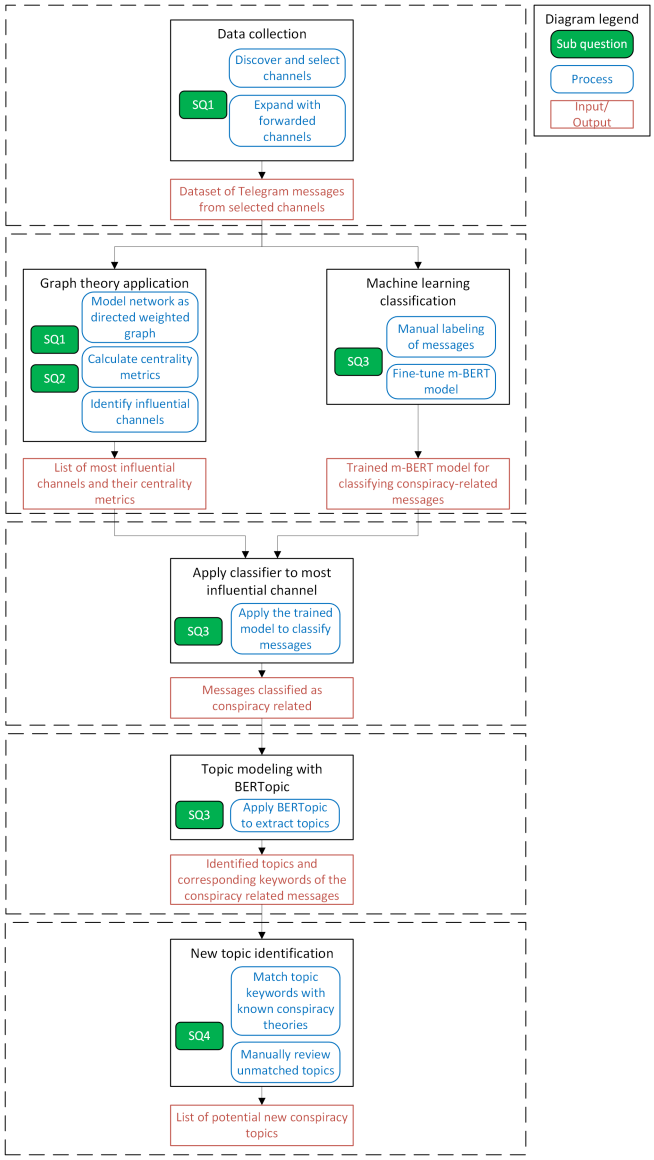
This chapter provides a detailed overview of the final methodology used in the study. This chapter outlines the specific steps taken, including data gathering, processing, and analysis, presenting a clear and structured description of the entire research process. While the operationalization chapter explains the ”why,” the methodology chapter focuses on the ”how,” offering a practical guide to the methods implemented. Additionally, the methodology chapter goes into greater detail on the data gathering methods, explaining how the data was collected, the tools and techniques used, and any challenges or considerations encountered during this process. This chapter ensures that the research process is transparent and replicable, providing a comprehensive account of the methodology applied to achieve the research objectives.

The objective of this research is to develop a model for the early detection of conspiracy theories on Telegram. To achieve this goal, this chapter outlines the methodological framework employed in the study. We begin by detailing the data collection process, including the selection of Telegram groups and channels relevant to our investigation. Following this, we explain the application of graph theory to identify influential channels within the network. This involves using centrality measures to pinpoint key nodes that significantly impact the spread of information. Subsequently, we describe the fine-tuning of the m-BERT model with our labeled dataset to accurately classify messages related to conspiracy theories. Finally, we apply BERTopic to the classified messages to identify and analyze different conspiracy theory topics. By combining these methodologies, we aim to create a robust model capable of detecting conspiracy theories, thereby contributing to the mitigation of misinformation on Telegram.

NOTE: Due to the sensitive nature of the topic, the dataset and code used in this research will not be shared publicly. The dataset includes potentially sensitive information that could compromise the privacy and security of individuals involved in the Telegram groups and channels analyzed. Only publicly available data was used and user information was not collected or processed. Additionally, the code developed for detecting conspiracy theories could be misused for unethical purposes, such as targeting or manipulating vulnerable populations, spreading misinformation, or other malicious activities. To prevent any potential harm and ensure ethical standards are maintained, access to the dataset and code will be restricted.

A visual overview of the methodology can be found in figure 5.1.

Figure 5.1: Method overview



5.1. Dataset

A challenge when researching phenomena on platforms like Telegram is identifying groups or channels relevant to the topic of interest (Hoseini et al., 2023). As mentioned in paragraph 4.2.2, it is a lot more difficult to discover groups due to the private nature of them. That is why it is chosen to focus on Telegram channels. The methodology by Hoseini et al. (2020) is followed to discover channels related to conspiracy theories. The steps are as followed:

1. **Discovering Telegram channels**. For this step it is chosen to focus on Dutch Telegram groups related to the Great Reset conspiracy theory. Other media like Twitter, Facebook and Google were used to discover Telegram URLs. Unlike Hoseini et al. (2023), we had no access to Twitter's Search API or the CrowdTangle API for Facebook. That is why a custom webcrawler was made using python that searched for one of the following Telegram URLs: t.me/, telegram.me/, or telegram.org/. The pages that the webcrawler searched through, were manually found. The pages were found by looking for keywords related to the Great Reset conspiracy theory. Some examples: Great Reset, World Economic Forum (WEF), Klaus Schwab, New World Order, Globale Elite, Agenda 2030.

2. **Collecting channel metadata**. Telegram's Web Client was used to obtain basic information about Telegram channels. Information was gathered such as the name of the channel, the description of the channel, and the amount of subscribers of the channel.
3. **Selecting Dutch Great Reset channels**. This step was already done by looking for pages related to the Great Reset.
4. **Validating Dutch Great Reset channels**. The channels were validated by viewing each channel and reading the content for a brief time. While validating the channels, it was important to select the channels that were actively discussing the Great Reset. Only Dutch channels were chosen, due to the small scale of this study and to keep the dataset relatively small but manageable.
5. **Joining and collecting messages in channels**. The next step is to join the Dutch Great Reset channels and collect their messages. To keep the scale of the study small 3 channels were joined. The messages from these channels were collected using Telethon (Lonami, 2019). Telethon uses Telegrams API (Telegram, 2024). Messages were collected that were sent between November 1, 2021, and January 1, 2024. A total of 16780 messages were collected.
6. **Expanding channels**. Of the messages extracted from the 3 initial seed channels, were all forwarded messages identified and the source channel of the forwarded message was extracted. It was then checked if the source channel is new (not already in the channel list) and if it the channel was new, was it added to a new list of channels to investigate. This new iteration extracted 45 new public channels. 1 channel was private and was left out of the dataset. The messages of these 45 new channels were added to the dataset and the messages of these channels were also collected. After expanding the dataset did we have a dataset of 747224 messages from 48 channels between November 1, 2021, and January 1, 2024. Another iteration of extracting channels from forwarded channels was done. This brought the total of channels to 751 channels. But due to the smaller scale of the study was it decided to stay with the dataset of 747224 messages from 48 channels.

5.2. Model

5.2.1. Graph theory application

As mentioned previously, the dataset of channels will be modeled as a directed weighted graph G. The graph G consists of nodes $N(G)$ and edges $E(G)$. Each node $n \in N(G)$ represents a Telegram channel. An edge $(i, j) \in E(G)$ represents a directed connection from channel i to channel j, where the direction indicates that channel i has forwarded messages from channel j. The weight of each edge w_{ij} corresponds to the number of messages that channel i has forwarded from channel j within a certain time interval. This directed graph can be formally written as:

$$G = (N(G), E(G)) \tag{5.1}$$

To identify the most influential users in the network, we take the centrality metrics chosen in paragraph 4.2.3 and compute the metrics for each node $n \in N(G)$.

1. **Normalized Weighted Degree Centrality**:
   - *Normalized Activity Centrality* ($\hat{C}^{w\alpha}_{D-out}(n)$): This measures the normalized weighted out-degree centrality, which is the same as equation 4.3:

$$C^{w\alpha}_{D-out}(n) = k_i^{out} \cdot \left(\frac{s_i^{out}}{k_i^{out}}\right)^{\alpha} \tag{5.2}$$

   Equation 5.2 is then normalized using Min-Max normalization in order to combine the centrality with other centralities:

$$\hat{C}^{w\alpha}_{D-out}(n) = \frac{C^{w\alpha}_{D-out}(n) - \min\limits_{n\in N(G)}(C^{w\alpha}_{D-out}(n))}{\max\limits_{n\in N(G)}(C^{w\alpha}_{D-out}(n)) - \min\limits_{n\in N(G)}(C^{w\alpha}_{D-out}(n))} \tag{5.3}$$

- *Normalized Popularity Centrality* ($\hat{C}^{w\alpha}_{D-in}(n)$): This measures the normalized weighted in-degree centrality, which is the same as equation 4.5:

$$C^{w\alpha}_{D-in}(n) = k^{in}_i \cdot \left(\frac{s^{in}_i}{k^{in}_i}\right)^{\alpha} \tag{5.4}$$

  Equation 5.4 is then normalized using Min-Max normalization in order to combine the centrality with other centralities:

$$\hat{C}^{w\alpha}_{D-in}(n) = \frac{C^{w\alpha}_{D-in}(n) - \min\limits_{n\in N(G)}(C^{w\alpha}_{D-in}(n))}{\max\limits_{n\in N(G)}(C^{w\alpha}_{D-in}(n)) - \min\limits_{n\in N(G)}(C^{w\alpha}_{D-in}(n))} \tag{5.5}$$

An $\alpha = 1.5$ was used as transferring and sharing of tacit knowledge requires strong ties (Hansen et al., 1999). For believers, conspiracy theories can be considered as a form of tacit knowledge. Therefore, when researching within echo chambers of believers of conspiracy theories, an α greater than 1 is more appropriate.

2. **Betweenness Centrality** ($C_{betw}(n)$): This centrality measures the extent to which a node lies on the shortest paths between other nodes. This is the same as equation 4.9:

$$C_{betw}(n) = \frac{1}{(|N(G)|-1)(|N(G)|-2)} \sum_{(j,k)\in N(G)\setminus\{n\}, j\neq k} \frac{\sigma_{jk}(n)}{\sigma_{jk}} \tag{5.6}$$

where σ_{jk} is the total number of shortest paths from node j to node k, and $\sigma_{jk}(n)$ is the number of those paths that pass through n.

As mentioned in paragraph 4.2.3, the weights in betweenness centrality typically represent distances or costs, and thus need to be inverted for the purpose of calculating path lengths in our context. ”Inversing” the weights ensures that the relative differences between the weights are maintained. This approach involves normalizing the weights to a [0, 1] range, inverting them, and then scaling them back to the original range while preserving the relative differences.

$$\text{normalized\_weight}(w) = \frac{w - w_{\text{min}}}{w_{\text{max}} - w_{\text{min}}} \tag{5.7}$$

$$W_{\text{inv\_norm}}(w) = 1 - \text{normalized\_weight}(w) \tag{5.8}$$

$$\text{scaled\_inverted\_weight}(w) = w_{\text{min}} + W_{\text{inv\_norm}}(w) \cdot (w_{\text{max}} - w_{\text{min}}) \tag{5.9}$$

3. **Normalized Viral Messages Centrality** ($\hat{C}_{viral}(n)$): A published message in a Telegram channel is called *viral* if the associated view counter in steady state is $\alpha\%$ higher than the average of other posts in a vicinity of n posts (Dargahi Nobari et al., 2021). α and n are the parameters. An $\alpha = 10\%$ and $n = 1$ are chosen similar to the setup of Dargahi Nobari et al. (2021).

$$C_{viral}(n) = \text{Number of viral messages from node } n \tag{5.10}$$

Similar to equations 5.2 and 5.4, $C_{viral}(n)$ is normalized using Min-Max normalization:

$$\hat{C}_{viral}(n) = \frac{C_{viral}(n) - \min\limits_{n\in N(G)}(C_{viral}(n))}{\max\limits_{n\in N(G)}(C_{viral}(n)) - \min\limits_{n\in N(G)}(C_{viral}(n))} \tag{5.11}$$

These metrics combine local, global and content measures:

- **Local Metrics**: Normalized Activity Centrality ($\hat{C}^{w\alpha}_{D-out}(n)$) and Normalized Popularity Centrality ($\hat{C}^{w\alpha}_{D-in}(n)$).

- **Global Metric**: Normalized Betweenness Centrality ($\hat{C}_{betweenness}(n)$).
- **Content Metric**: Normalized Viral Messages Centrality ($\hat{C}_{viral}(n)$)

The overall influence score $I(n)$ for each node n can be computed as the arithmetic mean of these metrics, enabling us to rank the nodes and identify the most influential user:

$$I(n) = \frac{1}{3}\left(\frac{\hat{C}^{w\alpha}_{D-out}(n) + \hat{C}^{w\alpha}_{D-in}(n)}{2} + \hat{C}_{betw}(n) + \hat{C}_{viral}(n)\right) \tag{5.12}$$

where $\hat{C}^{w\alpha}_{D-out}(n)$, $\hat{C}^{w\alpha}_{D-in}(n)$, $\hat{C}_{betweenness}(n)$, and $\hat{C}_{viral}(n)$ are the normalized values of the respective centrality measures.

Assigning equal weight to the centrality measures—local, global, and content-based—recognizes that influence within a social network is multifaceted. Each of these centrality measures captures different aspects of how a node (in this case, a Telegram channel) operates within the network. Influence in a network does not come from a single source; rather, it is a combination of how a node interacts within its immediate environment (local influence), how it connects different parts of the network (global influence), and how impactful its content is (content-based influence). By treating these measures equally, the methodology avoids biasing the influence score toward any particular type of centrality. For instance, a channel that is highly active (local influence) but does not act as a bridge between communities (global influence) or does not produce viral content (content influence) should not be unfairly advantaged or disadvantaged. This balanced approach aligns with Peng et al. (2018), who suggest that social influence is a complex interplay of various factors, each contributing to the overall power or importance of a node in a network.

5.2.2. Conspiracy theory detection model

Data preprocessing

The messages of the top scoring channel on influence score where then fetched and preprocessed. The data was then briefly preprocessed in order to be used later on. First of all, all duplicate messages were removed and second of all, were all unicode emoticons replaced with textual ASCII representations.

To develop the conspiracy detection model, m-BERT (Pires et al., 2019) was fine-tuned using a manually labeled dataset collected from messages on Telegram channels that were not identified as the most influential. The dataset consisted of messages that were manually labeled by the researcher as either conspiracy-related or not conspiracy-related. The open source annotation tool Doccano was used for annotating the dataset (Nakayama, Kubo, Kamura, Taniguchi, & Liang, 2018). For this labeling process, conspiracy theories were defined according to the European Commission's definition: "the belief that certain events or situations are secretly manipulated behind the scenes by powerful forces with negative intent" (European Commission, 2023). Additionally, Wikipedia was used as a reference source for current known conspiracy theories (Wikipedia, 2024). This grey literature was utilized because existing academic studies often do not focus on conspiracy theories, as these theories typically contradict established research. Using this annotated dataset, m-BERT was fine-tuned to train a classifier capable of identifying whether messages are related to conspiracy theories.

5.2.3. Topic modeling with BERTopic

After applying the classifier to the dataset of the messages of the top scoring channel on influence score, only those messages identified as conspiracy-related were kept for further analysis. To uncover the different conspiracy-related topics and theories discussed within these messages, BERTopic was employed. BERTopic, a topic modeling technique, allows for the extraction of coherent topics from large collections of text (M. Grootendorst, 2022). This method involves embedding the text data using transformer-based models and then applying clustering algorithms to group similar messages together. By analyzing these clusters, BERTopic can identify distinct topics and provide a clear understanding of the various conspiracy theories being discussed. The output includes a list of topics, each represented by a set of keywords and exemplary messages, which offers insight into the themes and narratives within the conspiracy-related content. This approach allows for a fast way to identify the specific conspiracy topics/theories circulating on Telegram channels.

5.2.4. New topic identification

To link the identified topics to known conspiracy theories, the OpenAI API (OpenAI, 2024b), specifically the GPT-4o model (OpenAI, 2024a), was used to analyze the keywords associated with each topic. The API was asked to match the extracted keywords with known conspiracy theories, identifying matches based on semantic similarity and contextual relevance. Topics that matched known conspiracy theories were documented accordingly.

For topics that could not be linked to any known conspiracy theories using the API, a manual review was conducted. These topics were examined to determine if they represented outliers, potentially new conspiracy theories, or other forms of misinformation. This manual check ensured that no conspiracy theories were overlooked. This combined method of using both automated and manual analysis offers a thorough understanding of the conspiracy narratives in the dataset, helping to identify them early.

It is chosen to begin with an automated link by the API (and those results are manually checked), which helps avoid biases that could arise if the researcher manually checks first. The alternative, first manually linking topics to conspiracy theories (and checking those results by the API) relies heavily on the researcher's existing knowledge, which may not cover newer or more obscure conspiracy theories that can emerge quickly. By using the API first, the risk of false negatives is reduced—ensuring that even less obvious links are considered. A manual review is then conducted to verify these links and prevent false positives, maintaining a balance between automated precision and human oversight. This dual-layered approach is critical because it leverages the broad knowledge base available to the API while also applying careful human judgment to validate the results.

6

Results

6.1. Influential channels identification

This section presents the findings from the analysis of the Telegram network. The focus is on identifying the most influential channels and visualizing the network structure. The top 10 channels are highlighted based on their influence scores, followed by a visualization of the entire network with node sizes proportional to their influence.

Table 6.1 below lists the top 10 channels identified as the most influential based on the computed centrality measures. The influence score is an aggregate metric derived from activity, popularity, closeness, betweenness, and viral messages centralities. The full table for all 48 channels can be found in Appendix A. Note: The channel ID's are anonymized for privacy reasons.

Table 6.1: An overview of the results of all the different centralities and the influence score.

Channel	$\hat{C}^{w\alpha}_{D-out}(n)$	$\hat{C}^{w\beta}_{D-in}(n)$	$C_{betw}(n)$	$\hat{C}_{viral}(n)$	$I(n)$
Channel1	1,00	0,06	0,99	0,13	0,55
Channel2	0,06	0,40	1,00	0,37	0,53
Channel3	0,69	0,00	0,37	0,73	0,48
Channel4	0,06	0,00	0,03	1,00	0,35
Channel5	0,04	0,02	0,69	0,18	0,30
Channel6	0,03	0,04	0,44	0,29	0,26
Channel7	0,01	0,15	0,18	0,45	0,24
Channel8	0,00	1,00	0,00	0,17	0,22
Channel9	0,00	0,04	0,23	0,39	0,21
Channel10	0,13	0,04	0,37	0,18	0,21

It is noticable that there is a very clear top 3 of the most influential channels. The results of these top 3 channels are discussed below:

Channel 1 has a maximum out-degree centrality score, indicating that it frequently forwards messages to many other channels. This suggests that Channel 1 acts as a significant source of information dissemination, playing a proactive role in spreading content across the network. However, its in-degree centrality is very low, meaning it rarely receives forwarded messages from other channels. This could indicate that while Channel 1 is a major broadcaster of information, it might not be engaging in dialogues or receiving feedback from others, which could limit its role in receiving and redistributing new information. Interestingly, Channel 1's viral messages score is also low, suggesting that although it broadcasts a lot of information, this content does not always tend to go viral. This could imply that the

channel is effective in spreading information widely but perhaps lacks the kind of compelling or sensational content that typically drives virality. This channel might be seen as an initiator of content, yet not necessarily as a channel whose content is highly resonant or engaging with the broader audience.

In contrast, Channel 2 has a relatively low out-degree centrality but a relatively higher in-degree centrality. This pattern indicates that Channel 2 is more of a receiver than a sender, acting as a hub where information from various other channels converges. This might suggest that Channel 2 is a popular node within the network, attracting information from different sources and potentially redistributing it selectively. Its high betweenness centrality further supports this, indicating that it often lies on the shortest path between other channels, thus playing a crucial role as an intermediary in the network. The decent score in viral messages suggests that when Channel 2 does forward content, it has a higher chance of that content going viral compared to Channel 1. This channel could be interpreted as a key player in amplifying certain messages, making it a strategic node for the spread of conspiracy theories.

Channel 3 presents a mixed profile with a relatively high out-degree centrality but the lowest in-degree centrality. This channel appears similar to Channel 1 in that it is more of a broadcaster than a receiver. However, unlike Channel 1, Channel 3 has a lower betweenness centrality, indicating it is not as central to the overall network's information flow. Its high viral messages score is particularly noteworthy, suggesting that although Channel 3 does not receive much information from others and is not central in bridging different parts of the network, the content it does produce or forward tends to be highly engaging or provocative, driving virality. This could indicate that Channel 3 might be a niche but influential player, capable of driving specific narratives that resonate strongly with its audience.

The differences in these centrality measures provide insights into how these channels influence the spread of information within the network. Channel 1, while highly active in sending out messages, may not contribute significantly to virality, meaning its role in spreading conspiracy theories might be more about volume rather than impact. Channel 2, acting as a hub, is strategically positioned to amplify conspiracy theories, given its role in connecting different parts of the network and its decent viral content score. Channel 3, though less central, appears to specialize in content that catches on quickly, making it a potent source for the rapid spread of specific conspiracy narratives.

Figure 6.1 shows the network graph visualization of the entire Telegram network of 48 channels, with node sizes proportional to their influence scores. The top 3 channels are also marked in red.

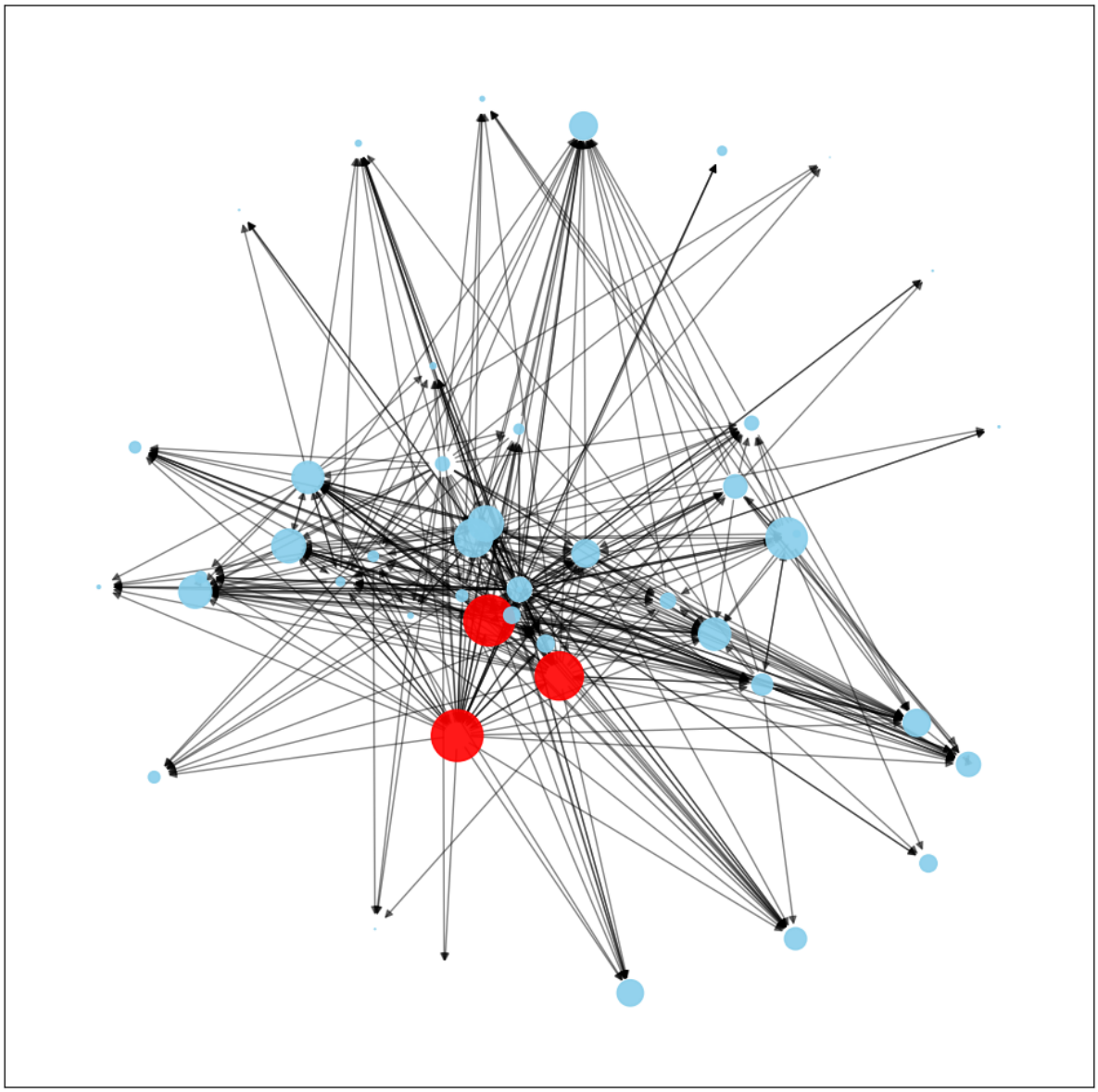


Figure 6.1: Telegram Network Graph with top 3 channels marked and Node Size Based on Influence Scores

6.2. Performance of conspiracy detection model

6.2.1. Dataset and fine-tuning

The manually labeled dataset used to fine-tune m-BERT consisted of 1,728 messages, primarily in Dutch. These messages were collected from channels that not performed as one of the top 10 channels on influence score. Of the 1728 messages 16% was labeled as conspiracy related. To ensure robust model training and evaluation, the dataset was split into training and validation sets, with 80% of the messages (1,382 messages) used for fine-tuning and the remaining 20% (346 messages) used for validation.

The fine-tuning of m-BERT was conducted using the following parameters:

- Number of training epochs: 3
- Training batch size per device: 32
- Evaluation batch size per device: 8
- Warmup steps: 500

- Weight decay: 0.01
- Learning rate: 2×10^{-5}

After fine-tuning m-BERT, the model was evaluated on the validation dataset. The evaluation metrics are summarized in table 6.2:

Table 6.2: Evaluation Results of the Fine-Tuned m-BERT Model

Loss	Accuracy	Precision	Recall	F1	Runtime	Samples per Second	Steps per Second
0,401	0,841	0,707	0,841	0,768	79,512	4,352	0,553

Each of these metrics gives insight into different aspects of how well the model is performing and how efficiently it is processing the data.

The loss value, which in this case is 0.401, represents the model's error rate. Loss is a measure of how far the model's predictions deviate from the actual labels. It is calculated during training and evaluation by comparing the predicted labels with the true labels. A lower loss indicates that the model's predictions are more accurate, meaning it is better at identifying conspiracy-related messages. However, some deviation is expected, so this loss value suggests the model is performing well but with some room for improvement.

Accuracy is another important metric, indicating how often the model's predictions are correct. With an accuracy of 0.841, or 84.1%, the model correctly classifies messages as either conspiracy-related or not in 84.1% of cases. While accuracy provides a general sense of performance, it does not distinguish between the types of errors made, such as false positives (incorrectly labeling a non-conspiracy message as conspiracy-related) or false negatives (failing to identify a conspiracy-related message).

Precision, measured at 0.707 or 70.7%, reflects the model's ability to correctly identify true positives out of all the instances it labeled as positive (conspiracy-related). This metric is particularly important when the cost of false positives is high—meaning it is important that when the model labels something as conspiracy-related, it is correct.

Recall, at 0.841 or 84.1%, measures how well the model identifies all actual positives in the dataset. In this context, it tells us how many of the total conspiracy-related messages the model correctly identified. High recall is critical when it is more important to identify all relevant cases, even if some irrelevant cases are also included.

The F1 score, which is 0.768 in this evaluation, is the harmonic mean of precision and recall. This metric provides a single measure of the model's performance that balances precision and recall, making it useful when it is important to ensure that the model is not favoring one over the other. A balanced F1 score indicates that the model is effectively identifying conspiracy-related messages while managing both false positives and false negatives.

The runtime of 79.512 seconds indicates how long it took to complete the evaluation, which is a measure of the model's computational efficiency. Alongside this, the samples per second (4.352) and steps per second (0.553) provide additional insight into how efficiently the model processes the data. Samples per second indicate the number of data points processed per second, while steps per second measure the number of iterations the model completes per second during evaluation. These metrics suggest that the model is not only accurate but also reasonably fast, which is important for applications requiring timely analysis of large datasets.

In summary, the fine-tuned m-BERT model performs well across several important dimensions, with strong accuracy, a balanced F1 score, and satisfactory precision and recall.

6.3. Topic modeling outcomes

After identifying the most influential channel as discussed in paragraph 6.1, a total of 16,145 messages were collected from this channel. Following the removal of duplicates, 14,908 unique messages

were processed by the m-BERT classifier. The analysis revealed that only 406 messages, approximately 2.11%, were identified as conspiracy-related. This relatively low percentage suggests that the channel's content is largely non-conspiratorial, or it may indicate potential limitations in the classifier's ability to detect certain types of conspiracy-related messages. To explore this further, the classifier's performance was also evaluated on the remaining two channels within the top three most influential channels. These channels showed similar results, with 2.41% and 2.51% of the messages classified as conspiracy-related.

The consistent performance across these channels supports the notion that the classifier is functioning as intended. However, the low percentage of conspiracy-related content prompts further consideration of whether the classifier is effectively capturing the full spectrum of conspiracy narratives or if the content in these influential channels is genuinely less conspiratorial in nature. Despite the low percentages, the detection of conspiracy-related messages within these channels remains significant. This analysis underscores the importance of verifying the classifier's accuracy and considering additional methods to ensure a comprehensive detection of conspiracy theories within the Telegram network.

To further analyze these messages, the Sentence Transformer model "paraphrase-multilingual-MiniLM-L12-v2" (Reimers & Gurevych, 2019) was used as step 1 of BERTopic. This Sentence Transformer model was chosen because of it's multilingual capabilities and the messages of our channels being either Dutch or English. For step 2 was the default UMAP used to perform its dimensionality reduction. UMAP was used with the standard parameters except for a random state = 42. This parameter was changed for reproducibility. In step 3 was the default HDBSCAN employed with a minimum cluster size of 5. In step 4 was the default CountVectorizer employed and in step 5 was the c-TF-IDF transformer forced to reduce frequent words. This removes stop words from the topic representation and should help with the interpretation of the topics. Stop words were only removed for the topic representations and not during the preprocessing, because the transformer-based embedding models that are used need the full context to create accurate embeddings. This configuration resulted in the identification of 17 distinct topics.

And example of the top output of BERTopic along with its associated keywords, is provided in Table 6.3.

Table 6.3: Keywords and Weights for Topic 0 (115 documents)

Keyword	Weight
chemtrails	0,525
ready	0,405
karen	0,405
fvd	0,405
stem	0,405
welkom	0,405
watch	0,371
message	0,370
happy	0,370
ze	0,370

6.4. New conspiracy theories

After identifying the topics using BERTopic, the OpenAI API was used to connect the keywords of each topic to known conspiracy theories. The API was queried with the keywords of each topic to determine if they match any existing conspiracy theories. The following prompt was used: "Are these keywords related to any known conspiracy theories?"

Out of the 17 topics identified by BERTopic, the OpenAI API was able to connect all 17 topics to existing conspiracy theories. This was unexpected, as a manual review of the keywords did not consistently reveal clear connections to known conspiracy theories. The analysis of the topics identified by BERTopic revealed that all topics could be linked to known conspiracy theories. While this might initially seem conclusive, it's essential to consider the process and potential limitations. For other datasets, not every

topic identified in previous analyses has been linked to a conspiracy theory by the API, which suggests that the tool can and does provide ”no” responses when appropriate. This result has several possible interpretations and implications. Firstly, the classifier may be highly precise, accurately labeling only those messages that are genuinely related to conspiracy theories. This would indicate that the classifier is performing excellently. Secondly, the dataset might be too small or too specific, focusing on a particular type of conspiracy theory or a limited number of channels, resulting in a limited variety of conspiracy theories being detected, all of which are already known. Thirdly, the classifier's threshold for labeling a message as conspiracy-related might be too strict, causing it to exclude borderline cases that could introduce new or less well-known conspiracy theories. This strictness ensures that only clear examples of known conspiracy theories are labeled, reducing the chances of false positives. The implications of these interpretations are significant. If the classifier is performing well, the results suggest that it is highly effective in identifying relevant content, demonstrating its robustness and reliability. However, the results might also indicate a need to expand the dataset to include a broader range of channels and messages, capturing a wider variety of conspiracy theories, including potential new ones. Additionally, the strictness of the classifier might need to be re-evaluated to balance precision and recall, potentially identifying a more diverse set of conspiracy theories. Overall, while the fact that all topics were linked to known conspiracy theories is a positive sign of the classifier's effectiveness, it also highlights the need to consider the scope of the dataset and the strictness of the classifier. Expanding the dataset and fine-tuning the classification threshold could provide a more comprehensive understanding of conspiracy theories on Telegram, including the identification of new or emerging ones.

The table with all the topics, their keywords and linked conspiracy theory, can be found in appendix B.

6.5. Validation

To validate the model, a fictive conspiracy theory called ”The Verdant Shadow Conspiracy” was created. The Verdant Shadow Conspiracy suggests that houseplants are not merely decorative or good for improving air quality, but are actually sophisticated surveillance devices. This theory claims that a covert organization known as ”Greenwatch” is responsible for genetically engineering these plants to monitor human activities worldwide. A full description of the Verdant Shadow Conspiracy can be found in Appendix C.

Using the OpenAI API, 100 Dutch Telegram messages about the Verdant Shadow Conspiracy were simulated in a style similar to the top-scoring channel on influence score. A full list of the messages can be found in Appendix D. These messages were re-added to the dataset of messages from the top-scoring channel. The choice of 100 simulated messages for the fictive conspiracy theory, ”The Verdant Shadow Conspiracy,” was primarily driven by the need to introduce a significant but manageable number of new messages into the existing dataset. This number was chosen to be large enough to be statistically meaningful while also ensuring that the addition did not overwhelm the original dataset, which contained thousands of messages. The intent was to introduce the fictive conspiracy at a scale that could realistically be detected by the model without skewing the overall analysis.

Out of these 100 messages, only 10 were identified as conspiracy-related by the classifier, which aligns with the model's previously measured accuracy. This suggests that the model is consistent in its classification performance, identifying approximately the same proportion of conspiracy-related content in both the original and enriched datasets. However, the fact that 90 messages were not flagged raises important considerations about the model's sensitivity and threshold settings. This outcome could indicate that the model maintains a high level of specificity, avoiding false positives, but it also prompts a discussion about whether the model might be missing more subtle or less explicitly framed conspiracy content.

BERTopic was then applied to this enriched dataset, which included the fictive messages. This analysis resulted in 18 topics, including the detection of the fictive conspiracy topic, with associated keywords as shown in Table 6.4.

Table 6.4: Keywords for Topic 7 Identified by BERTopic

Keyword	Weight
een	0.758
planten	0.625
urban	0.607
surveillance	0.607
de	0.583
van	0.582
overheid	0.542
hoe	0.540
te	0.535
monitor	0.504

BERTopic managed to find a topic that could be linked to the Verdant Shadow conspiracy, although the keywords contained some stop words, indicating that the parameters for filtering stop words could be put to more sensitive.

The OpenAI API was then queried with the dataset, including the fictive conspiracy topic. As expected, the fictive conspiracy was not marked as an existing conspiracy. This validates that the model works for marking this fictive conspiracy theory as a non-existing conspiracy. The full list of linked conspiracy theories can be found in Appendix E. It is up to the researcher to determine if the topics not belonging to an existing conspiracy theory are outliers or potentially new conspiracy theories. Only the fictive conspiracy theory was a topic that was labeled as not an existing conspiracy. The other topic was labeled as an outlier.

Table 6.5: The topics not linked to existing conspiracy theories by GPT-4o.

Topic ID	Keywords	Decision
5	netherlands, Kaag, overheid, heen, bang, Rutte, Klaver, democracy, grote, niks	Outlier
7	**een, planten, dat, voor surveillance, de, van, overheid, te, hoe**	**Fictive CT**

7

Discussion

7.1. Discussion

The integration of multiple analytical techniques, including graph theory, natural language processing, and topic modeling, provided a comprehensive approach to identifying and analyzing conspiracy theories on Telegram. The use of m-BERT for classifying conspiracy-related messages, combined with BERTopic for topic modeling, allowed for insights into the topics of the conspiracy related Telegram messages. The identification of the most influential channels using centrality measures further emphasized the role of key players in the distribution of conspiracy theories. The successful application of these methods demonstrates the robustness of the combined analytical approach.

The findings align with existing literature on the detection of conspiracy theories in social networks. Arezo Bodaghi et al. (2023) showed how hybrid techniques, focusing on both network features and content, can be successful in detecting misinformation. This study extends on this by applying graph theory centrality measures and applying transformer based models to both classify messages and analyze content of messages.

Additionally, much of the existing research has focused on platforms like Twitter and Facebook, where user interactions are more public and the network structures are different. By contrast, Telegram provides a unique environment due to its semi-private nature and the use of channels and groups, which can have a different impact on the spread of information. This study's approach to analyzing Telegram networks fills a gap in the literature by addressing these specific characteristics

Furthermore, existing literature often relies on basic content analysis or network metrics without integrating sophisticated machine learning techniques. This study enhances the analytical framework by incorporating m-BERT for classification of conspiracy-related messages and BERTopic for nuanced topic modeling. This integration allows for a more detailed and accurate identification of conspiracy theories, which is less explored in prior studies.

Lastly, to the best of my knowledge, has there been no metric so far for identifying the most influential user specifically on Telegram. This research proposes a combined influence metric, which includes both local and global centrality measures as well as content-specific measures. Which is an original contribution to the field. This metric provides a more holistic understanding of influence within a Telegram network

7.1.1. Implications for Detection of Conspiracy Theories

The detection of conspiracy theories is critical for mitigating their impact on public discourse and preventing the spread of misinformation. The methodologies employed in this study, particularly the fine-tuning of m-BERT and the use of BERTopic, provide powerful tools for identifying conspiracy theories in their early stages. By focusing on influential channels and analyzing the thematic content of messages, it is possible to detect emerging conspiracy theories. However, it is important to approach the mitigation of conspiracy theories with caution to avoid infringing on freedom of speech. Efforts to counter misinfor-

mation should be balanced with the protection of individuals' rights to express their views, ensuring that interventions are targeted and respectful of democratic principles. This balance is particularly important on platforms like Telegram, where users may be more concerned about privacy and censorship.

7.1.2. Limitations of the Study

Despite the promising results, several limitations should be acknowledged. Firstly, the dataset used in this study, while comprehensive, may not capture the full spectrum of conspiracy theories on Telegram. The focus on Dutch channels related to the Great Reset limits the generalizability of the findings. Expanding the dataset to include a broader range of topics and languages could enhance the robustness of the model. Future studies should consider larger and more diverse datasets, as well as alternative methods for simulating and validating conspiracy-related content.

Secondly, the strictness of the classifier, while ensuring high precision, may exclude borderline cases that could introduce new or less well-known conspiracy theories. This highlights a potential trade-off between precision and recall, which shows the need for continuous evaluation and adjustment of the classification threshold to balance precision and recall. Additionally, the annotator of the labeled dataset is no expert on conspiracy theories. The simulated messages for the fictive conspiracy theory, while useful for validation, may not fully capture the complexity and variability of real-world conspiracy theories. Additionally, the reliance on the OpenAI API for message simulation introduces potential biases based on the training data and algorithms used by the API. Future studies should consider larger and more diverse datasets, as well as alternative methods for simulating and validating conspiracy-related content.

8

Conclusion

8.1. Conclusion

This study aimed to develop a model for the early detection of conspiracy theories on Telegram by leveraging graph theory, machine learning, and topic modeling. The research was guided by the main research question: *How can conspiracy theories be identified on Telegram?* To address this, several sub-questions were explored.

Sub-Question 1: How can graph theory be applied to model the network structure of Telegram channels?
Graph theory was applied to model the network structure of Telegram channels by representing each channel as a node and each forwarded message as a directed edge. The weight of each edge corresponded to the number of messages forwarded between channels. This modeling approach enabled the analysis of the network structure and the identification of influential channels based on centrality measures.

Sub-Question 2: Which graph metrics can be used in identifying influential nodes within Telegram networks?
Several centrality measures were employed to identify influential nodes within the Telegram network, including weighted degree centrality (activity and popularity), betweenness centrality, and viral messages centrality. These metrics provided a comprehensive assessment of a channel's influence by considering both local and global network properties as well as content-specific measures.

Sub-Question 3: How can Telegram messages be analyzed to determine if they are conspiracy-related?
Telegram messages were analyzed using a fine-tuned multilingual BERT (m-BERT) model. The classifier was trained on a manually labeled dataset to distinguish between conspiracy-related and non-conspiracy-related messages. The use of a transformer-based model allowed for effective classification of multilingual messages, addressing the linguistic diversity of the dataset.

Sub-Question 4: What new and existing conspiracy theories can be identified on Telegram?
BERTopic was applied to the messages classified as conspiracy-related to identify distinct topics. The identified topics were then analyzed using the OpenAI API to determine if they matched known conspiracy theories. All identified topics were linked to existing conspiracy theories, suggesting the effectiveness of the model in detecting known conspiracies and highlighting the need for further research to capture emerging theories.

8.1.1. Addressing the Main Research Question

The main research question, *How can conspiracy theories be identified on Telegram?*, was addressed through a combination of methodologies. By modeling the network structure using graph theory, influential channels were identified, providing key targets for monitoring. The fine-tuned m-BERT classifier

demonstrated robust performance in detecting conspiracy-related messages, and BERTopic facilitated the identification of specific conspiracy theories. This integrated approach proved effective in identifying conspiracy theories on Telegram, although continuous refinement and expansion of the dataset are necessary for broader applicability.

8.1.2. Scientific Contribution

This research makes a scientific contribution by applying an interdisciplinary approach that integrates graph theory, machine learning, and topic modeling to detect conspiracy theories on Telegram. This approach contributes to several academic fields, including network analysis, computational social science, and misinformation studies.

By modeling the Telegram network and employing centrality measures, the study offers a framework for identifying influential nodes in social networks. The use of a multilingual BERT model and BERTopic for content analysis demonstrates the potential of combining network analysis with machine learning to address complex social phenomena. By combining content-based analysis with network-based metrics, the research provides a more holistic approach to detecting conspiracy theories, which can be further developed and applied to other social media platforms. This methodological approach contributes to the literature on topic modeling and its applications in social media research, particularly in the context of detecting and categorizing misinformation.

8.1.3. Societal Contribution

The societal contribution of this research lies in its potential to mitigate the spread of conspiracy theories on social media platforms, particularly on Telegram. Conspiracy theories have a impact on society, as they can undermine trust in institutions, polarize communities, and even cause violence. In recent years, the rise of digital platforms has amplified the spread of such theories, making it easier for misinformation to reach and influence large audiences rapidly.

By developing a model that identifies conspiracy theories on Telegram, this research offers a tool that can be used by policymakers, social media platforms, and researchers to monitor and address the spread of harmful narratives. The identification of influential channels within the network is especially important, as these channels often serve as key nodes for the dissemination of conspiracy theories. By targeting these influential channels, interventions can be more strategic and effective, potentially reducing the overall impact of conspiracy theories on public discourse.

The broader societal contribution also includes the potential to enhance public safety. Conspiracy theories can lead to real-world harm, including public health risks, as seen during the COVID-19 pandemic, where misinformation about the virus and vaccines contributed to vaccine hesitancy and public distrust in health authorities. By improving the ability to detect and respond to conspiracy theories early, this research supports efforts to prevent such harm. The insights gained from this research can inform the development of more effective strategies to protect public dialogue from the dangers of conspiracy theories, ultimately contributing to the trust in democratic institutions.

8.2. Recommendations for Future Research

Future research should focus on expanding the dataset to include a more diverse range of conspiracy theories and languages, enhancing the model's robustness and generalizability. Additionally, exploring different machine learning models and adjusting classification thresholds could improve the detection of emerging or less well-known conspiracy theories. CT-BERT could be a good model for English datasets.

Using a classifier based on centrality measures for finding the most influential user has shown to be effective (Bucur, 2020). Future research could incorporate this. Interdisciplinary collaboration with experts in psychology, sociology, and communication studies can provide deeper insights into the drivers and impacts of conspiracy theories, ultimately informing more effective detection strategies.

Finally, while the study successfully identified influential channels using centrality measures, it did not account for the potential impact of non-channel actors (e.g., individuals in group chats) on the spread of conspiracy theories. Future research could explore the role of these actors and incorporate them into the analysis, providing a more comprehensive understanding of the dynamics at play within Telegram

networks.

References

Ahmed, W., Vidal-Alaball, J., Downing, J., & López Seguí, F. (2020). COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. doi:10.2196/19458

Aïmeur, E., Amri, S., & Gilles Brassard, ·. (2023). Fake news, disinformation and misinformation in social media: a review. *13*, 30. doi:10.1007/s13278-023-01028-5

Algemene Inlichtingen- en Veiligheidsdienst. (2023). Aivd jaarverslag 2022.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, *3*(null), 993–1022.

Bodaghi, A. [A.], Schmitt, K., Watine, P., & Fung, B. (2023). A literature review on detecting, verifying, and mitigating online misinformation. *IEEE Transactions on Computational Social Systems*, 1–27. doi:10.1109/TCSS.2023.3289031

Bodaghi, A. [Arezo], Schmitt, K. A., Watine, P., & Fung, B. C. (2023). A Literature Review on Detecting, Verifying, and Mitigating Online Misinformation. *IEEE Transactions on Computational Social Systems*. doi:10.1109/TCSS.2023.3289031

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. doi:10.1016/J.SOCNET.2007.04.002

Bruns, A., Harrington, S., & Hurcombe, E. (2020). 'corona? 5g? or both?': The dynamics of covid-19/5g conspiracy theories on facebook. *Media International Australia*, *177*(1), 12–29. doi:10.1177/1329878X20946113

Bucur, D. (2020). Top influencers can be identified universally by combining classical centralities. doi:10.1038/s41598-020-77536-7

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in knowledge discovery and data mining* (pp. 160–172). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cassam, Q. (2019). *Conspiracy theories*. Polity Press.

Cheema, G. S. (2021). *On the Role of Images for Analyzing Claims in Social Media*. Retrieved from https://github.com/cleopatra-itn/image_text_claim_detection

Cinelli, M., Etta, G., Avalle, M., Quattrociocchi, A., Di Marco, N., Valensise, C., … Quattrociocchi, W. (2022). Conspiracy theories and social media platforms. *Current Opinion in Psychology*, *47*, 101407. doi:10.1016/J.COPSYC.2022.101407

Dargahi Nobari, A., Sarraf, M. H. K. M., Neshati, M., & Erfanian Daneshvar, F. (2021). Characteristics of viral messages on Telegram; The world's largest hybrid public and private messenger. *Expert Systems with Applications*, *168*, 114303. doi:10.1016/J.ESWA.2020.114303

Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Retrieved from https://github.com/tensorflow/tensor2tensor

Dlewis. (2024, March). The great reset conspiracy: How it spread in the netherlands. Retrieved from https://www.visionofhumanity.org/the-spread-of-the-great-reset-conspiracy-in-the-netherlands/

Douglas, K. M. (2021). Covid-19 conspiracy theories. *Group Processes & Intergroup Relations*, *24*, 270–275. doi: 10.1177/1368430220982068. doi:10.1177/1368430220982068

Enders, A. M., Uscinski, J. E., Seelig, M. I., Casey, ·., Klofstad, A., Wuchty, S., … Stoler, J. (2023). The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation. *45*, 781–804. doi:10.1007/s11109-021-09734-6

European Commission. (2023). Identifying conspiracy theories. Retrieved from https://commission.europa.eu/strategy-and-policy/coronavirus-response/fighting-disinformation/identifying-conspiracy-theories_en#documents

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

Gerster, L., Kuchta, R., Hammer, D., & Schwieter, C. (2022, October). *Telegram as a Buttress: How far-right extremists and conspiracy theorists are expanding their infrastructures via Telegram*. Institute for Strategic Dialogue. Retrieved from www.isdgermany.org

González-Padilla, D. A., & Tortolero-Blanco, L. (2020, July). Social media influence in the COVID-19 pandemic. doi:10.1590/S1677-5538.IBJU.2020.S121

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*.

Grootendorst, M. P. (n.d.). The algorithm. Retrieved from https://maartengr.github.io/BERTopic/algorithm/algorithm.html

Gruzd, A., & Mai, P. (2020). Going viral: How a single tweet spawned a covid-19 conspiracy theory on twitter. *Big Data and Society*, *7*(2). doi:10.1177/2053951720938405

Hansen, T., Baratoff, G., & Neumann, H. (1999). Dominating Opponent Inhibition of On and Off Pathways for Robust Contrast Detection, 232–239. doi:10.1007/978-3-642-60243-6{_}27

Heft, A., & Buehling, K. (2022). Measuring the diffusion of conspiracy theories in digital information ecologies. *Convergence*, *28*(4), 940–961. doi:10.1177/13548565221091809

Hoseini, M., Melo, P., Benevenuto, F., Feldmann, A., & Zannettou, S. (2023). On the Globalization of the QAnon Conspiracy Theory Through Telegram. doi:10.1145/3578503.3583603

Hoseini, M., Melo, P., Júnior, M., Benevenuto, F., Chandrasekaran, B., Feldmann, A., … Zannettou, S. 2. (2020). Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twiï¿¿er CCS CONCEPTS. doi:10.1145/3419394.3423651

Khanday, A. M. U. D., Wani, M. A., Rabani, S. T., Khan, Q. R., Amor Pérez-Rodríguez, M., Vizcaíno-Verdú, A., … Rayees Khan, Q. (2023). Hybrid Approach for Detecting Propagandistic Community and Core Node on Social Networks. doi:10.3390/su15021249

Khaund, T., Shaik, M., & Agarwal, N. (2020). Data Collection and Sensemaking from Telegram: A Case Study of Ukrainian Political Leaders Channels and Chat Groups.

Kokkera, K., Chen, A., Malapati, C., Demchenko, A., & Fung, C. (2023). A Study on the Conspiracy Theory Propagation Network on Twitter. In *Proceedings of ieee/ifip network operations and management symposium 2023, noms 2023*. doi:10.1109/NOMS56928.2023.10154285

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on world wide web* (pp. 591–600). doi:10.1145/1772690.1772751

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). Albert: A lite bert for self-supervised learning of language representations. arXiv: 1909.11942 [cs.CL]. Retrieved from https://arxiv.org/abs/1909.11942

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). Nasa faked the moon landing—therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, *24*, 622–633. doi: 10.1177/0956797612457686. doi:10.1177/0956797612457686

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv: 1907.11692 [cs.CL]. Retrieved from https://arxiv.org/abs/1907.11692

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Allen, P. G. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. Retrieved from https://github.com/pytorch/fairseq

Lonami. (2019). Telethon. Retrieved from https://docs.telethon.dev/en/stable/quick-references/client-reference.html

McInnes, L., Healy, J., & Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv: 1802.03426 [stat.ML]. Retrieved from https://arxiv.org/abs/1802.03426

Medina Serrano, J. C., Papakyriakopoulos, O., & Hegelich, S. (2020, July). NLP-based feature extraction for the detection of COVID-19 misinformation videos on YouTube. In K. Verspoor, K. B. Cohen, M. Dredze, E. Ferrara, J. May, R. Munro, … B. Wallace (Eds.), *Proceedings of the 1st workshop on NLP for COVID-19 at ACL 2020*, Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2020.nlpcovid19-acl.17

Moffitt, J., King, C., & Carley, K. (2021). Hunting conspiracy theories during the covid-19 pandemic. *Social Media and Society*, *7*(3). doi:10.1177/20563051211043212

Morgia, M. L., Mei, A., Mongardini, A. M., & Wu, J. (2021). *Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements; Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements*. Retrieved from https://t.me/trumps

Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., & Liang, X. (2018). doccano: Text annotation tool for human. Software available from https://github.com/doccano/doccano. Retrieved from https://github.com/doccano/doccano

Napolitano, M. G., & Reuter, K. (2021). What is a conspiracy theory? *88*, 2035–2062. doi:10.1007/s10670-021-00441-6

Newman, M. (2018, July). *Networks*. doi:10.1093/oso/9780198805090.001.0001

Nobari, A. D., Reshadatmand, N., & Neshati, M. (2017). Analysis of telegram, an instant messaging service. *International Conference on Information and Knowledge Management, Proceedings*, *Part F131841*, 2035–2038. doi:10.1145/3132847.3133132

OpenAI. (2024a, May). Hello gpt-4o. Retrieved from https://openai.com/index/hello-gpt-4o

OpenAI. (2024b). Openai api. Retrieved from https://platform.openai.com/docs/api-reference/chat

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. doi:10.1016/J.SOCNET.2010.03.006

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*. doi:10.1136/bmj.n160

Paolillo, J. C. (2018). The flat earth phenomenon on youtube. *First Monday*, *23*. doi:10.5210/fm.v23i12.8251

Peeters, S., & Willaert, T. (2022). Telegram and Digital Methods. *M/C Journal*, *25*(1). doi:10.5204/mcj.2878

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., & Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, *106*, 17–32. doi:10.1016/J.JNCA.2018.01.005

Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual bert? arXiv: 1906.01502 [cs.CL]. Retrieved from https://arxiv.org/abs/1906.01502

Pósfai, M., & Barabási, A.-L. (2016). *Network science*. Citeseer.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). *Truthy: Mapping the Spread of Astoturf in Microblog Streams*. Association for Computing Mac inery.

Reimers, N., & Gurevych, I. (2019, November). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing*, Association for Computational Linguistics. Retrieved from http://arxiv.org/abs/1908.10084

Ren, Z. (, Dimant, E., & Schweitzer, M. (2023). Beyond belief: How social engagement motives influence the spread of conspiracy theories. *Journal of Experimental Social Psychology*, *104*, 104421. doi:10.1016/J.JESP.2022.104421

Robinson, O., Sardarizadeh, S., Goodman, J., Giles, C., & Williams, H. (2021, June). What is the great reset - and how did it get hijacked by conspiracy theories? BBC. Retrieved from https://www.bbc.com/news/blogs-trending-57532368

Rose, M. E., Opolot, D. C., & Georg, C. P. (2022). Discussants. *Research Policy*, *51*(10), 104587. doi:10.1016/J.RESPOL.2022.104587

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter. arXiv: 1910.01108 [cs.CL]. Retrieved from https://arxiv.org/abs/1910.01108

Schäfer, K., & Choi, J.-E. (2023). Transparency in messengers: A metadata analysis based on the example of telegram. In *Proceedings of the 34th acm conference on hypertext and social media*. doi:10.1145/3603163.3609034

Schwab, K. (n.d.). Now is the time for a “great reset”. Retrieved from https://www.weforum.org/agenda/2020/06/now-is-the-time-for-a-great-reset/

Shahsavari, S. [S.], Holur, P., Wang, T., Tangherlini, T., & Roychowdhury, V. (2020). Conspiracy in the time of corona: Automatic detection of emerging covid-19 conspiracy theories in social media and the news. *Journal of Computational Social Science*, *3*(2), 279–317. doi:10.1007/s42001-020-00086-5

Shahsavari, S. [Shadi], Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. doi:10.1007/s42001-020-00086-5

Shao, C., Hazel Kwon, K., Walker, S., & Li, Q. (2023). A dynamic analysis of conspiratorial narratives on twitter during the pandemic. *Cyberpsychology, Behavior, and Social Networking*, *26*(5), 338–345. doi:10.1089/cyber.2022.0218

Shu, K., Bhattacharjee, A., Faisal Alatawi, |., Tahora, |., Nazer, H., Kaize Ding, |., … Liu, |. H. (2020). Combating disinformation in a social media age. doi:10.1002/widm.1385

Simon, M., Welbers, K., C. Kroon, A., & Trilling, D. (2022). Linked in the dark: A network approach to understanding information flows within the Dutch Telegramsphere. *Information Communication and Society*. doi:10.1080/1369118X.2022.2133549

Sunstein, C. R., & Vermeule, A. (2008). Symposium on conspiracy theories conspiracy theories: Causes and cures*. doi:10.1111/j.1467-9760.2008.00325.x

Telegram. (n.d.). Telegram faq. Retrieved from https://telegram.org/faq

Telegram. (2024). Channels.getmessages. Retrieved from https://core.telegram.org/method/channels.getMessages

Theocharis, Y., Cardenal, A., Jin, S., Aalberg, T., Hopmann, D., Strömbäck, J., … Štětka, V. (2023). Does the platform matter? social media and covid-19 conspiracy theory beliefs in 17 countries. *New Media and Society*, *25*(12), 3412–3437. doi:10.1177/14614448211045666

Ullah, F., & Lee, S. (2017). Community clustering based on trust modeling weighted by user interests in online social networks. *Chaos, Solitons & Fractals*, *103*, 194–204. doi:10.1016/J.CHAOS.2017.05.041

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30), Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. doi:10.1016/J.IS.2020.101582

Walther, S., & McCoy, A. (2021). Us extremism on telegram: Fueling disinformation, conspiracy theories, and accelerationism. *Perspectives on Terrorism*, *15*(2), 100–124. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111330388&partnerID=40&md5=3b3c46ab2e0702a6f56d70e573db2a9c

Wikipedia. (2024, June). Wikimedia Foundation. Retrieved from https://en.wikipedia.org/wiki/List_of_conspiracy_theories

Wischerath, D., Godwin, E., Bocheva, D., Brown, O., Roscoe, J. F., & Davidson, B. I. (2024). Spreading the Word: Exploring a Network of Mobilizing Messages in a Telegram Conspir-acy Group. doi:10.1145/3613905.3651888

XLin, X., Shang, T., & Liu, J. (2014). An estimation method for relationship strength in weighted social network graphs. *Journal of Computational Chemistry*, *02*, 82–89. Retrieved from https://api.semanticscholar.org/CorpusID:40121949

Zaghouani, W., Vladimir, I., & Ruiz, M. (2023). *COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter*. Retrieved from https://github.com/digitalepidemiologylab/covid-twitter-bert

A

Centralities results

This appendix presents a overview of the centrality measures calculated for each Telegram channel analyzed in this study. The centrality metrics include weighted degree centralities, betweenness centrality, and viral message centrality. These metrics are crucial for understanding the influence and role of each channel within the Telegram network, particularly in the context of spreading conspiracy theories. The following table provide detailed results for all 48 channels considered in this research.

Note: The channel ID's are anonymized for privacy reasons.

Table A.1: An overview of the results of all the different centralities and the influence score.

Channel	$\hat{C}^{w\alpha}_{D-out}(n)$	$\hat{C}^{w\beta}_{D-in}(n)$	$C_{betw}(n)$	$\hat{C}_{viral}(n)$	$I(n)$
Channel1	1,00	0,06	0,99	0,13	0,55
Channel2	0,06	0,40	1,00	0,37	0,53
Channel3	0,69	0,00	0,37	0,73	0,48
Channel4	0,06	0,00	0,03	1,00	0,35
Channel5	0,04	0,02	0,69	0,18	0,30
Channel6	0,03	0,04	0,44	0,29	0,26
Channel7	0,01	0,15	0,18	0,45	0,24
Channel8	0,00	1,00	0,00	0,17	0,22
Channel9	0,00	0,04	0,23	0,39	0,21
Channel10	0,13	0,04	0,37	0,18	0,21
Channel11	0,00	0,18	0,00	0,38	0,16
Channel12	0,02	0,02	0,16	0,29	0,15
Channel13	0,00	0,09	0,03	0,37	0,15
Channel14	0,00	0,69	0,00	0,08	0,14
Channel15	0,02	0,00	0,32	0,04	0,12
Channel16	0,00	0,10	0,11	0,20	0,12
Channel17	0,02	0,00	0,11	0,22	0,11
Channel18	0,00	0,43	0,00	0,08	0,10
Channel19	0,04	0,01	0,08	0,16	0,09
Channel20	0,00	0,00	0,00	0,18	0,06
Channel21	0,00	0,15	0,03	0,08	0,06
Channel22	0,07	0,01	0,08	0,05	0,06
Channel23	0,00	0,00	0,08	0,06	0,05
Channel24	0,00	0,02	0,01	0,10	0,04
Channel25	0,23	0,00	0,00	0,00	0,04
Channel26	0,03	0,01	0,02	0,04	0,03
Channel27	0,00	0,00	0,00	0,08	0,03
Channel28	0,00	0,02	0,00	0,07	0,03

Channel29	0,00	0,04	0,05	0,01	0,03
Channel30	0,00	0,02	0,05	0,00	0,02
Channel31	0,00	0,00	0,06	0,00	0,02
Channel32	0,00	0,00	0,00	0,05	0,02
Channel33	0,01	0,04	0,00	0,02	0,01
Channel34	0,00	0,00	0,02	0,01	0,01
Channel35	0,00	0,00	0,01	0,02	0,01
Channel36	0,00	0,01	0,00	0,01	0,01
Channel37	0,00	0,03	0,00	0,01	0,01
Channel38	0,00	0,00	0,01	0,00	0,01
Channel39	0,00	0,00	0,00	0,01	0,00
Channel40	0,02	0,00	0,00	0,00	0,00
Channel41	0,00	0,00	0,00	0,00	0,00
Channel42	0,00	0,00	0,00	0,00	0,00
Channel43	0,00	0,00	0,00	0,00	0,00
Channel44	0,00	0,00	0,00	0,00	0,00
Channel45	0,00	0,00	0,00	0,00	0,00
Channel46	0,00	0,00	0,00	0,00	0,00
Channel47	0,00	0,00	0,00	0,00	0,00
Channel48	0,00	0,00	0,00	0,00	0,00

B

Topics linked to existing CT

In this appendix, we provide an exhaustive list of the topics identified using BERTopic, along with their corresponding keywords and the conspiracy theories they were linked to by the OpenAI API. This analysis was conducted before the introduction of the fictive ”Verdant Shadow Conspiracy” into the dataset. The aim is to offer a clear view of the existing conspiracy theories detected in the Telegram channels without the influence of the fictive content. Each table outlines the topic ID, associated keywords, and the linked conspiracy theory.

Table B.1: The result of GPT-4o linking the keywords of the results of BERTopic to known existing conspiracy theories

Topic ID	Keywords	Linked Conspiracy Theory
0	chemtrails, ready, karen, fvd, stem, welkom, watch, message, happy, ze	Chemtrails
1	ukraine, russia, putin, russian, president, war, weapons, ukrainian, plant, biological	Biological Weapons in Ukraine
2	restricties, elektrisch, meat, cattle, electricity, shortages, cars, wave, no, te	Climate Change and Food Supply Control
3	fda, gov, covid, vaccinated, 19, document, vaccine, death, test, immune	COVID-19 and Vaccine
4	eu, speech, trieste, politie, deal, italy, asiel, haven, free, duits	EU Control and Immigration
5	netherlands, nr, pedo, heen, weer, vrijdag, zaterdag, country, grote, Clinton	Pizzagate
6	sun, 150, km, islands, eyes, see, local, open, away, wouldn	Solar Anomalies and Secret Islands
7	nasa, joke, cyndi, holland, interview, employee, former, defibrillator, complete, heart	NASA and Space Exploration Hoaxes
8	planet, water, 1214b, berta, gj, ruimte, alien, astrophysics, harvard, webb	Alien Life and Space Exploration
9	face, masks, wear, your, again, stop, mask, tip, fungal, fibers	Mask Mandates and Health Risks
10	maan, aarde, traveling, model, speed, 67, 000mph, around, its, revolves	Flat Earth
11	israel, 1000, god, libya, icke, misdaden, gaza, underground, hun, territories	David Icke and Anti-Semitic Narratives
12	isolated, fraude, grootste, journal, humanity, virusen, onderzoek, february, published, wuhan	COVID-19 Origin and Fraud
13	warming, sgtnewsnetwork, global, co2, swindle, film, weather, narrative, social, falling	Climate Change Hoax

14	5g, source, mind, control, repair, chips, cellphones, inside, scam, guy	5G and Mind Control
15	fire, wheat, roemenie, fields, wall, burn, dumbsan-dunderground, 22, clearly, 13	Food Supply Sabotage
16	emf, dna, 5g, studies, electromagnetic, university, phone, exposure, cells, radio	Electromagnetic Radiation and Health
17	Overheid, stupid, failed, too, democracy, bang, series, onmogelijk, people, niks	General Distrust in Governance

C

The Verdant Shadow Conspiracy

This appendix includes a detailed description of the fictive conspiracy theory created for the purpose of validating the detection model. The ”Verdant Shadow Conspiracy” was designed to mimic real-world conspiracy narratives, suggesting that common houseplants are actually surveillance devices engineered by a covert organization. This section provides the full narrative used to simulate this fictive conspiracy, offering context for the messages generated and their integration into the dataset.

The Theory

The Verdant Shadow Conspiracy suggests that houseplants are not merely decorative or good for improving air quality, but are actually sophisticated surveillance devices. This theory claims that a covert organization known as ”Greenwatch” is responsible for genetically engineering these plants to monitor human activities worldwide.

The Purpose

Greenwatch allegedly uses these bio-engineered plants to gather data on people’s habits, conversations, and even emotional states. The plants are equipped with microscopic sensors that can detect and transmit sound, monitor air quality changes that correspond with different emotions, and track movement within their vicinity.

The Technology

The technology behind this involves integrating nanotechnology with plant biology. These plants are said to be capable of using their natural processes, such as photosynthesis, to power tiny embedded sensors and transmitters. The data collected by these plant-based devices are then supposedly sent back to Greenwatch for analysis and action.

The Evidence

Supporters of the Verdant Shadow Conspiracy point to the sudden and rapid increase in the popularity of indoor gardening and the global push for greener, plant-filled environments in urban areas. They cite this trend as a strategy to ensure that no private space is free from surveillance.

The Goal

The ultimate goal of Greenwatch, according to the theory, is to create an omnipresent surveillance network that is completely hidden in plain sight, camouflaged within the greenery that people willingly bring into their homes and offices. This network would ostensibly be used to maintain social order and control by subtly influencing or directly intervening in individuals’ lives based on the collected data.

Secrecy and Implementation

To maintain secrecy, Greenwatch operates through various fronts, including seemingly benign gardening clubs, plant nurseries, and floral shops. They promote the distribution of these engineered plants as part of community green initiatives or urban beautification projects.

This theory combines elements of espionage with environmental concerns, presenting a narrative that taps into fears about privacy and the pervasiveness of surveillance technology in daily life.

D

List of Dutch Verdant Shadow messages

In this appendix, you will find the full list of Dutch Telegram messages generated by ChatGPT that relate to the fictive ”Verdant Shadow Conspiracy.” These messages were designed to resemble the style and tone of actual messages found in conspiracy-related channels on Telegram. They were added to the dataset to test the robustness of the conspiracy detection model. Each message is presented as it was generated, providing insight into how the fictive conspiracy was embedded in the research.

1. ”Recent onderzoek wijst uit dat sommige kantoorplanten in staat zijn tot het meten van CO2-niveaus, wat klinkt als een milieuvriendelijke inzet, maar wat als dit slechts een façade is voor het monitoren van personeel en hun gedrag op de werkplek?”,
2. ”Een insider deelde dat tijdens de laatste wereldwijde klimaattop ’speciale’ planten waren toegevoegd aan vergaderzalen; naar verluidt niet alleen voor de esthetiek, maar om deel te nemen aan een gelaagd observatie- en dataverzamelingssysteem dat gekoppeld is aan de Great Reset-agenda.”,
3. ”Heb je je wel eens afgevraagd waarom de overheid zo aandringt op het plaatsen van meer groen in onze wijken, juist wanneer grotere surveillancewetten worden aangenomen? Deze zogenaamde milieubewegingen lopen parallel met grotere stappen richting een gecontroleerde samenleving.”,
4. ”In mijn stad is een project gelanceerd voor slimme parken, compleet met IoT-verbonden planten die de parkomgeving beheren. Interessant genoeg lijken deze systemen meer data te verzamelen dan nodig voor milieudoeleinden alleen.”,
5. ”Op een geheime conferentie waar ik bij was, werd gesproken over hoe botanische tuinen over de hele wereld netwerken van spionagepunten zijn geworden, strategisch gepositioneerd nabij overheidsgebouwen en diplomatieke vertegenwoordigingen.”,
6. ”Een artikel in een wetenschappelijk tijdschrift beschreef hoe genetisch gemodificeerde planten konden communiceren via een soort van plantaardig internet, een concept dat verontrustend dicht bij de bewakingsinfrastructuur van de Great Reset ligt.”,
7. ”Stel je voor, een wereld waarin elke plant in staat is om alles wat je zegt te horen en dit terug te rapporteren aan een centrale database. Dit is geen verre toekomst; er zijn al patenten aangevraagd voor deze technologieën door onbekende maar machtige corporaties.”,
8. ”Tijdens een lokaal gemeenteraadsdebat kwam aan het licht dat nieuwe stedelijke ontwikkelingsprojecten verplichte integratie van ’intelligente’ planttechnologieën vereisen, wat de vraag oproept wie hier werkelijk baat bij heeft - de burgers of de staat?”,
9. ”Er gaat een gerucht dat de recente distributie van gratis planten aan huishoudens eigenlijk een overheidsexperiment is om te zien hoe effectief planten gebruikt kunnen worden als instrumenten voor massa-surveillance.”,

10. ”Naar aanleiding van de recente installatie van nieuwe ’groene’ verkeersscheidingsinstallaties langs de snelweg, die beweerd worden de luchtkwaliteit te verbeteren, is het curieus dat deze dezelfde technologie bevatten die ook in spionage wordt gebruikt.”,
11. ”Het is me opgevallen dat er steeds meer onderwijsinstellingen zijn die planten gebruiken als een manier om ’groen denken’ te bevorderen. Echter, sommige van deze planten zijn uitgerust met sensoren. Is dit een manier om toezicht te houden op onze jongeren?”,
12. ”In een onlangs gelekt document van een grote biotechnologiecorporatie wordt gesproken over het ’vergroenen van werkplekken’, niet enkel voor de gezondheid van werknemers maar ook als een nieuwe strategie voor het in kaart brengen van werknemersinteracties.”,
13. ”De snelle toename van urban farming initiatieven, geprezen om hun duurzaamheidsvoordelen, lijkt ook een dubbele agenda te hebben. Zijn deze locaties nu echt voor voedselproductie of dienen ze als camouflagenetwerk voor surveillancedoeleinden?”,
14. ”Experts hebben gewaarschuwd voor nieuwe vormen van ’biologisch hacken’ waarbij planten zijn ontworpen om specifieke gegevens van hun omgeving te verzamelen. Dit zou ver gaan in de controle van de overheid over de persoonlijke vrijheid van mensen.”,
15. ”Een bezorgde wetenschapper onthulde dat sommige milieuprojecten, gefinancierd onder het mom van duurzaamheid, in werkelijkheid experimenten zijn om te zien hoe goed planten kunnen worden ingezet als surveillance middelen in openbare en private ruimtes.”,
16. ”De discussie over het inzetten van planten voor het monitoren van de volksgezondheid in steden lijkt een nobel doel, maar bij nader inzien zijn er concrete aanwijzingen dat deze zelfde technologieën worden gebruikt om burgers op meer intrusive wijzen te monitoren.”,
17. ”Opvallend veel privacy advocaten hebben hun bezorgdheid geuit over ’groene’ technologieën die in nieuwe technologische producten worden geïntegreerd, suggererend dat deze planten mogelijk fungeren als luisterapparaten die gevoelige informatie kunnen doorsturen naar niet nader genoemde partijen.”,
18. ”Het beleid om groene ruimtes te ontwikkelen in nieuwe stedelijke ontwikkelingsplannen wordt vaak gekoppeld aan ’smart city’ technologieën. Critici beweren dat dit deel uitmaakt van een groter schema om gegevens over burgers te verzamelen zonder hun expliciete toestemming.”,
19. ”In een wereld waar alles verbonden is, hoe weten we dan dat de planten in onze huizen en kantoren ons niet bespioneren? Gezien de vooruitgang in biotechnologie en nanotechnologie, is het niet langer een verre hypothese dat planten kunnen worden gebruikt als surveillance-apparaten.”,
20. ”Er zijn steeds meer aanwijzingen dat overheden over de hele wereld aanzienlijk hebben geïnvesteerd in onderzoek naar hoe natuurlijke organismen, zoals planten, kunnen worden gebruikt om inlichtingen te verzamelen over burgers, vaak onder het voorwendsel van wetenschappelijk onderzoek of milieu-initiatieven.”,
21. ”Het gebruik van planten in openbare ruimtes, zoals bibliotheken en scholen, is niet nieuw, maar de recente toename van ’slimme planten’ met ingebouwde sensoren roept vragen op over de werkelijke intenties achter deze groene toevoegingen.”,
22. ”Discussies in beveiligingsforums hebben onthuld dat bepaalde soorten planten, die bekend staan om hun luchtzuiverende eigenschappen, ook bijzonder effectief zijn in het detecteren van geluidsvibraties, wat ze ideale middelen maakt voor het onopvallend monitoren van gesprekken.”,
23. ”Tijdens een internationale milieuconferentie werd gesproken over het ’verantwoord integreren van biotechnologie in dagelijks leven’, maar tussen de regels door waren er hints dat deze technologieën verder kunnen gaan dan alleen het verbeteren van onze leefomgeving.”,
24. ”De recente push van steden om meer te ’vergroenen’ gaat gepaard met de installatie van nieuwe soorten straatmeubilair dat verdacht veel weg heeft van technologie die wordt gebruikt in spionage en dataverzameling.”,
25. ”Het lijkt erop dat de nieuwste generatie milieuvriendelijke kantoorgebouwen, vol met levende groene muren, niet alleen bedoeld zijn om werknemers zich beter te laten voelen, maar ook om nauwkeuriger toezicht te houden op hun dagelijkse activiteiten en interacties.”,

26. ”Een onlangs gepubliceerd artikel door een anonieme voormalige werknemer van een grote biotechnologie firma onthulde plannen om planten te integreren in surveillance netwerken, waarbij gebruik wordt gemaakt van hun natuurlijke aanwezigheid in stedelijke gebieden om niet op te vallen.”,
27. ”In een poging om de stedelijke veiligheid te verbeteren, worden nieuwe soorten ’beveiligingsplanten’ ontwikkeld die kunnen reageren op specifieke chemische signalen in de lucht, wat in theorie kan helpen bij het detecteren van bedreigingen voordat ze zich voordoen, maar ook vragen oproept over privacy.”,
28. ”Sommige milieuactivisten hebben gewaarschuwd dat de interesse van de overheid in het financieren van groene projecten in stedelijke gebieden misschien minder te maken heeft met zorg voor het milieu en meer met het verhogen van toezicht op deze dichtbevolkte gebieden.”,
29. ”Recente investeringen in ’groene’ technologieën en projecten zoals stadsbossen en gemeenschappelijke tuinen zijn mogelijk onderdeel van een grotere strategie om stedelijke gebieden te controleren onder het mom van duurzaamheid en openbare veiligheid.”,
30. ”Het groeiende aantal patenten op technologieën die geluid kunnen detecteren en analyseren via planten geeft een nieuw beeld van wat mogelijk is op het gebied van stedelijke surveillance.”,
31. ”De groei van gemeenschappelijke ’groene’ ruimten in nieuwe woonwijken is wellicht niet enkel een poging om de kwaliteit van leven te verbeteren, maar ook een manier om bewonersgemakkelijker in de gaten te houden via geïntegreerde sensor-technologie in het landschap.”,
32. ”Er zijn zorgen onder privacyadvocaten dat de trend van biophilic design in architectuur, waarbij de natuur wordt geïntegreerd in de bouw van gebouwen, een achterdeur kan bieden voor uitgebreide surveillance.”,
33. ”Met de huidige technologische vooruitgang is het niet langer buiten het bereik van de mogelijkheid dat planten, eenmaal aangepast, kunnen functioneren als levende recorders, continu informatie verzamelend over hun omgeving.”,
34. ”De integratie van biotechnologie in stedelijke ontwikkeling, vooral in de vorm van ’slimme parken’ en ’groene gebouwen’, heeft potentieel de dualiteit van het verbeteren van de leefomgeving en het uitbreiden van de mogelijkheden voor stadsbrede surveillance.”,
35. ”Recente lekken hebben aangetoond dat sommige overheden experimenteren met het gebruik van planten als onderdeel van een geïntegreerd beveiligingssysteem in publieke ruimtes, waarbij de natuurlijke aanwezigheid ervan hen helpt om niet opgemerkt te worden.”,
36. ”Het gebruik van geavanceerde genetische manipulatietechnieken om planten specifieke eigenschappen te geven die hen in staat stellen om als onderdeel van een surveillancesysteem te functioneren, roept vragen op over de ethische grenzen van biotechnologie.”,
37. ”Naast hun rol in het verbeteren van de luchtkwaliteit, worden planten in toenemende mate bekeken voor hun vermogen om als biologische sensoren te fungeren, wat ze waardevolle instrumenten maakt voor zowel gezondheidsmonitoring als surveillance.”,
38. ”Het is bekend dat sommige overheden initiatieven financieren om de aanplant van bepaalde soorten bomen en struiken te bevorderen in zowel private als openbare ruimtes, vaak die soorten die het best geschikt zijn voor integratie met verborgen surveillanceapparatuur.”,
39. ”Hoewel het idee van ’luisterende’ planten misschien ver lijkt, hebben recente technologische ontwikkelingen dit idee veel realistischer gemaakt, met verschillende startups die zich richten op de commerciële ontwikkeling van dergelijke technologieën.”,
40. ”Critici van stedelijke surveillancepraktijken wijzen vaak op de proliferatie van ’slim groen’ in openbare ruimtes als een manier om de inzameling van persoonlijke gegevens op een minder zichtbare manier uit te breiden.”,
41. ”De trend van het vergroenen van werkruimtes gaat vaak gepaard met de introductie van planten die niet alleen esthetisch zijn, maar ook zijn aangepast om de luchtkwaliteit en mogelijkerwijs gesprekken en andere geluiden in hun omgeving te monitoren.”,

42. ”Enkele documenten die zijn gelekt door hackers tonen aan dat onderzoek naar plant-based surveillance veel verder is gevorderd dan eerder werd aangenomen, met field tests die al plaatsvinden in verschillende grote steden rond de wereld.”,
43. ”Er zijn aanwijzingen dat de ’Verdant Shadow Conspiracy’ nauw verweven is met ’The Great Reset’. Men speculeert dat de intensivering van de groene ruimtes in grootstedelijke gebieden niet alleen gericht is op milieuvoordelen, maar ook deel uitmaakt van een groter plan om stedelijke populaties intensiever te monitoren en te controleren.”,
44. ”Onlangs is mij ter ore gekomen dat nieuwe technologieën die biologische reacties van planten op menselijke emoties kunnen detecteren, worden getest in overheidsgebouwen. Dit zou een nieuw tijdperk van surveillance inluiden, waarbij zelfs onze meest subtiele emoties niet onopgemerkt blijven.”,
45. ”Een bron binnen een overheidsagentschap heeft onthuld dat de financiering voor onderzoek naar communicatie tussen planten drastisch is toegenomen. Dit onderzoek zou niet alleen wetenschappelijk van aard zijn; het zou ook getest worden voor toepassingen in nationale veiligheid en massasurveillance.”,
46. ”De recente toename van ’slimme’ plantenbakken en tuinierapparatuur, vaak gefinancierd door schimmige organisaties met vermeende banden met de overheid, doet geloven dat deze apparaten dienen als zowel voedingsbron als afluisterapparatuur.”,
47. ”In een uitgelekt rapport van een niet nader genoemde overheidsinstantie werd gesproken over het integreren van sensor-gebaseerde technologie in publieke ’groene muren’. Deze technologie zou niet alleen de gezondheid van de planten monitoren, maar ook gegevens verzamelen over de mensen die erlangs lopen.”,
48. ”Tijdens een conferentie over duurzaamheid gaf een expert toe dat het pushen van urban farming initiatieven deel uitmaakt van een groter plan om burgers meer te binden aan stedelijke gebieden, wat het gemakkelijker maakt om ze te monitoren onder het mom van duurzame ontwikkeling.”,
49. ”Ik heb ontdekt dat er gespecialiseerde software wordt ontwikkeld voor zogenaamde ‘ecologische’ monitoring, die in werkelijkheid diep kan integreren met stedelijke beveiligingssystemen. Deze systemen zijn in staat gedetailleerde demografische en gedragsgegevens te verzamelen van iedereen die door groene zones loopt.”,
50. ”Een patent is recent goedgekeurd voor een technologie die het mogelijk maakt voor planten om draadloos te communiceren met elkaar en met een centraal systeem. Hoewel dit wordt verkocht als een manier om de gezondheid van urban forests te monitoren, zijn de implicaties voor privacy enorm.”,
51. ”De obsessie van onze moderne maatschappij met het binnenhalen van planten en het vergroenen van huizen, aangewakkerd door overheidsprogramma’s, lijkt nu dubieuze bijbedoelingen te hebben. Deze planten zijn wellicht niet enkel voor decoratie of gezondheid, maar fungeren als oren en ogen van de staat.”,
52. ”Ik ben er onlangs achter gekomen dat enkele plantensoorten die gebruikt worden in overheidskantoren, genetisch gemodificeerd zijn om specifieke golflengtes van geluid beter op te vangen. Dit maakt ze perfect voor spionage zonder dat iemand het doorheeft.”,
53. ”Het gebruik van planten als dekmantel voor surveillance is een methode die diep geworteld lijkt in historische spionage praktijken. De hedendaagse technologie maakt het echter mogelijk dit op ongekende schaal toe te passen.”,
54. ”Na gesprekken met een voormalig medewerker van een geheim project, is het mij duidelijk geworden dat sommige openbare parken zijn ontworpen rondom de behoeften van surveillancetechnologie, niet alleen rondom ecologie en recreatie zoals publiekelijk wordt beweerd.”,
55. ”Opvallend genoeg is er een duidelijke correlatie tussen de plaatsing van nieuwe ‘intelligente’ groene zones en wijken met een hogere socio-economische status. Dit zou kunnen wijzen op gerichte surveillance-activiteiten, gericht op het monitoren van bepaalde bevolkingsgroepen.”,
56. ”Recent onderzoek toont aan dat sommige ‘milieuvriendelijke’ bedrijven die planten verkopen met

ingebouwde monitoringstechnologie, daadwerkelijk subsidies ontvangen van overheidsinstellingen. Dit roept ernstige vragen op over de werkelijke motieven achter deze subsidies.",

57. "In beveiligde overheidsgebouwen wordt gebruikgemaakt van planten die niet alleen CO2 omzetten, maar ook fijne deeltjes en geluidsgolven kunnen detecteren. Dit soort dual-use technologieën worden steeds vaker ingezet in het kader van nationale veiligheid.",
58. "Er zijn trainingen voor overheidsmedewerkers die specifiek gericht zijn op het herkennen van modificaties in planten die wijzen op integratie van surveillance-technologie. Dit is een indicatie dat dergelijke technieken op grote schaal worden ingezet binnen overheidsinstellingen.",
59. "Verschillende stadsplanners hebben anoniem hun zorgen geuit over de toenemende druk om surveillance-technologieën te integreren in publieke groene projecten. Deze technologieën zijn vaak zo subtiel dat ze niet te onderscheiden zijn van gewone plantenverzorgingstools.",
60. "Er is onlangs een netwerk blootgelegd van bedrijven in de agrotechnologie die heimelijk werken aan de ontwikkeling van plantensoorten die in staat zijn tot geavanceerde data-acquisitie. Deze planten zouden in staat zijn om alles, van geluid tot biochemische signalen in hun omgeving, te monitoren.",
61. "De groeiende trend van biophilic design in architectuur wordt niet enkel gedreven door esthetische of gezondheidsredenen. Achter de schermen spelen surveillance-capaciteiten een grote rol, met planten die strategisch geplaatst worden om gegevens te verzamelen.",
62. "Inlichtingendiensten hebben conferenties gehouden die specifiek gericht zijn op het gebruik van flora als onderdeel van beveiligingsprotocollen. Deze conferenties behandelen technieken voor het integreren van plant-gebaseerde sensoren in publieke en private ruimten.",
63. "Het fenomeen van 'slimme steden' omvat niet alleen technologie zoals camera's en sensoren. De integratie van 'intelligente' planten speelt een steeds grotere rol, waarbij stedelijke flora wordt gebruikt om in real-time gegevens over inwoners te verzamelen.",
64. "Een geheime bron meldde dat experimenten met genetisch gemodificeerde planten, die verhoogde niveaus van straling kunnen detecteren, zijn uitgevoerd in samenwerking met nationale veiligheidsdiensten. Deze planten zouden tactisch gebruikt kunnen worden in gebieden die van hoog strategisch belang zijn.",
65. "Recentelijk zijn documenten gelekt waaruit blijkt dat sommige 'milieubeschermende' overheidsprogramma's in werkelijkheid zijn ontworpen om surveillance-technologieën te verspreiden onder de dekmantel van klimaatverandering en duurzaamheidsinitiatieven.",
66. "Het is fascinerend en tegelijkertijd beangstigend dat recente ontwikkelingen in plantenbiotechnologie de mogelijkheid bieden om planten zo te programmeren dat ze reageren op specifieke audiofrequenties, wat ze tot perfecte afluisterapparaten maakt.",
67. "De promotie van groendaken in nieuwe bouwvoorschriften lijkt een ecologisch motief te hebben, maar dient tegelijkertijd als een platform voor geavanceerde surveillancetechnieken. Deze groendaken zijn uitgerust met technologie die niet alleen de plantengroei bevordert, maar ook data verzamelt over de bewoners onder deze daken.",
68. "De inzet van planten als bio-sensoren in stedelijke omgevingen is al lang geen sciencefiction meer. Overheidsinstellingen testen openlijk hoe stedelijke flora gebruikt kan worden voor het monitoren van milieuvervuiling, wat slechts een dekmantel lijkt voor het verzamelen van gegevens over burgers.",
69. "In een reeks workshops over stedelijk groen werden technieken gedemonstreerd voor het integreren van communicatieapparatuur in tuinontwerpen. Deze apparatuur is niet te onderscheiden van normale tuinartikelen, waardoor ze onopvallend in de openbare ruimte kunnen worden geplaatst.",
70. "Er is een toenemende bezorgdheid onder privacyactivisten over de rol van planten in smart homes, waarbij technologieën die aanvankelijk werden gepresenteerd als middelen om plantengroei te stimuleren, nu ook gebruikt worden voor het monitoren van huiselijke activiteiten.",
71. "Op internationale beurzen voor tuinbouwtechnologie zijn recent diverse innovaties gepresenteerd die de manier waarop we over planten denken radicaal kunnen veranderen. Deze tech-

nologieën stellen planten in staat om te functioneren als levende sensoren binnen een beveiligingssysteem, waardoor de traditionele surveillance wordt uitgebreid naar de natuurlijke omgeving.",

72. "Het concept van biobeveiliging is aan het verschuiven van elektronische systemen naar biologische systemen. Overheden over de hele wereld experimenteren met het gebruik van planten als onderdeel van hun beveiligingsinfrastructuur, wat nieuwe ethische en privacyvraagstukken oproept.",
73. "Op een recent symposium over biosurveillance werd gediscussieerd over het ethische gebruik van genetisch gemodificeerde organismen, waaronder planten die zijn ontworpen om in real-time chemische en biologische data te verzamelen en te verzenden. Deze ontwikkelingen wijzen op een toekomst waarin onze eigen tuinen ons zouden kunnen bespioneren.",
74. "Critici van stedelijke ontwikkelingsplannen wijzen op de potentiële gevaren van het integreren van surveillance-technologieën in openbare groenprojecten. Deze kritiek wordt vaak genegeerd of onderdrukt, wat de vraag oproept wie er werkelijk baat heeft bij deze projecten.",
75. "Heeft iemand anders gemerkt dat overheidsgebouwen vol staan met dezelfde soorten planten? Volgens mij is dit geen toeval. Het lijkt erop dat ze zijn geïmplementeerd voor surveillance, gezien hun locatie in strategische vergaderruimtes.",
76. "Laatst was er een lek over hoe bepaalde plantensoorten speciale sensoren bevatten die geluid kunnen opvangen en data kunnen verzenden. Dit is overduidelijk een teken dat we worden afgeluisterd door wat we in huis halen.",
77. "Wist je dat tijdens de Koude Oorlog experimenten met planten werden uitgevoerd om ze als spionagemiddelen te gebruiken? Dit is duidelijk doorontwikkeld tot wat nu toegepast wordt in onze stedelijke omgevingen.",
78. "Opvallend hoe veel stedelijke ontwikkelingsprojecten tegenwoordig een 'groen' component hebben. Ze zeggen dat het voor duurzaamheid is, maar in werkelijkheid zijn het tactische punten voor het monitoren van burgers.",
79. "Er zijn studies die aantonen dat sommige planten in staat zijn chemische signalen uit te zenden wanneer ze bepaalde stimuli detecteren. Kun je je voorstellen hoe dit gebruikt kan worden voor surveillance?",
80. "In een recente documentaire over geheime overheidsprogramma's kwam naar voren dat investeringen in biotechnologie niet alleen gericht zijn op gezondheid, maar ook op het ontwikkelen van planten die inlichtingen kunnen verzamelen.",
81. "Op forums wordt druk gespeculeerd over de echte reden achter de snelle verspreiding van huis- en kantoortuinen. Velen geloven dat dit een dekmantel is voor het uitrollen van een uitgebreid spionagenetwerk.",
82. "Heb je ooit gehoord van het 'Flora Surveillance Project'? Dat is een vermeend geheim project waarbij genetisch gemodificeerde planten worden gebruikt om burgers te monitoren.",
83. "Ik las een onderzoek waarin werd beweerd dat bepaalde planten elektronische signalen kunnen opvangen. Dit zou technologie kunnen zijn die de staat gebruikt om ons te bespioneren via iets zo onschuldigs als onze huiskamerplanten.",
84. "De recente toename van 'slimme tuinen' en 'verticale groene muren' in nieuwe zakendistricten is zeer verdacht. Ze zijn ideaal gepositioneerd om gesprekken van duizenden dagelijkse passanten op te vangen.",
85. "Er is onlangs een patent aangevraagd op een technologie die het mogelijk maakt voor planten om digitale data te verzamelen. Dit zou theoretisch kunnen betekenen dat elke plant in je huis een potentiële dataverzamelaar is.",
86. "Veel mensen merken dat hun planten buitengewoon goed gedijen, zelfs met minimale verzorging. Sommigen speculeren dat dit komt door verborgen technologieën die hen in leven houden als onderdeel van een groter surveillanceplan.",
87. "Het feit dat de NSA een van de grootste financiers van botanisch onderzoek is, roept serieuze vragen op. Wat weten zij over planten dat wij niet weten?",

88. ”Experts in encryptietechnologie hebben gewaarschuwd dat de capaciteiten van planten om informatie op te slaan en te verzenden veel verder gaan dan we momenteel begrijpen. Dit opent de deur voor nieuwe methoden van massasurveillance.”,
89. ”Opmerkelijk hoe sommige nieuwbouwwijken planten gebruiken die specifiek geselecteerd zijn op hun vermogen om milieuveranderingen te detecteren. Dit zou deel kunnen uitmaken van een poging om toezicht te houden op de woonomgevingen van burgers.”,
90. ”Er gaan geruchten dat tijdens internationale topontmoetingen speciale planten worden geplaatst die zijn ontworpen om de aanwezigheid van bepaalde chemicaliën te detecteren, wat wijst op een poging om de veiligheid of spionage te verhogen.”,
91. ”De overheid is naar verluidt een van de grootste afnemers van geavanceerde hydrocultuursystemen. Dit roept vragen op over het werkelijke gebruik van deze technologieën in overheidsgebouwen.”,
92. ”Wist je dat sommige ’groene’ non-profitorganisaties die stadsvergroening promoten, eigenlijk gefinancierd worden door overheidscontracten? Dit zou een manier kunnen zijn om surveillanceapparatuur te installeren onder het mom van milieubehoud.”,
93. ”Op een recent gehouden technologiebeurs werd een nieuw type plantensensor onthuld die kan integreren met bestaande smart home-systemen. Dit lijkt onschuldig, maar de mogelijkheden voor overheidsmonitoring zijn enorm.”,
94. ”Het aantal start-ups dat zich richt op de integratie van planten en technologie is in de afgelopen jaren geëxplodeerd. Deze bedrijven ontvangen ongewoon veel belangstelling en financiering van overheidsinstanties.”,
95. ”Sommige klokkenluiders hebben gesuggereerd dat de plotselinge verschuiving naar bio-energetische onderzoeksprojecten een voortzetting is van oudere, meer controversiële surveillanceprogramma’s.”,
96. ”Tijdens een recente conferentie over stadsplanning werd een nieuwe ontwerpfilosofie gepresenteerd die sterk de nadruk legt op ’levende gebouwen’ vol planten. Critici suggereren dat dit een nieuwe manier is om burgers in de gaten te houden.”,
97. ”In sommige privébeveiligingskringen wordt gezegd dat de beste verborgen camera’s niet zijn gemaakt van metaal en plastic, maar van bladeren en stengels.”,
98. ”Het recente overheidsinitiatief om groene subsidies te verhogen voor huizen en kantoren bevat clausules die verdacht veel lijken op voorwaarden voor het toestaan van installatie van monitoringapparatuur.”,
99. ”Een bekende biotechnoloog gaf onlangs een lezing waarin hij suggereerde dat we nog maar aan het oppervlak krabben van wat mogelijk is met genetische plantenmodificatie, specifiek verwijzend naar hun potentieel voor gegevensverzameling en -verwerking.”,
100. ”Heb je ooit gemerkt dat in films vaak wordt verwezen naar planten die mensen afluisteren? Sommigen zeggen dat dit een vorm van ’soft disclosure’ is, waarbij de waarheid verstopt zit in fictie.”,

E

Topics linked to existing CT

This appendix provides a list of the topics identified by BERTopic after the fictive ”Verdant Shadow Conspiracy” was added to the dataset. The tables include the topic IDs, associated keywords, and the conspiracy theories linked by the OpenAI API, including the fictive one. This appendix is essential for understanding the impact of adding the fictive content on the model’s detection capabilities and how it interacted with existing conspiracy narratives.

Table E.1: The result of GPT-4o linking the keywords of the results of BERTopic to known existing conspiracy theories

Topic ID	Keywords	Linked Conspiracy Theory
0	chemtrails, ready, karen, fvd, stem, welkom, watch, message, happy, ze	Chemtrails
1	ukraine, russia, putin, russian, president, war, weapons, ukrainian, plant, biological	Biological Weapons in Ukraine
2	restricties, elektrisch, meat, cattle, electricity, shortages, cars, wave, no, te	Climate Change and Food Supply Control
3	fda, gov, covid, vaccinated, 19, document, vaccine, death, test, immune	COVID-19 and Vaccine
4	eu, speech, trieste, politie, deal, italy, asiel, haven, free, duits	EU Control and Immigration
5	netherlands, Kaag, overheid, heen, bang, Rutte, Klaver, democracy, grote, niks	-
6	sun, 150, km, islands, eyes, see, local, open, away, wouldn	Solar Anomalies and Secret Islands
7	**een, planten, dat, voor surveillance, de, van, overheid, te, hoe**	-
8	nasa, joke, cyndi, holland, interview, employee, former, defibrillator, complete, heart	NASA and Space Exploration Hoaxes
9	planet, water, 1214b, berta, gj, ruimte, alien, astrophysics, harvard, webb	Alien Life and Space Exploration
10	face, masks, wear, your, again, stop, mask, tip, fungal, fibers	Mask Mandates and Health Risks
11	maan, aarde, traveling, model, speed, 67, 000mph, around, its, revolves	Flat Earth
12	israel, 1000, god, libya, icke, misdaden, gaza, underground, hun, territories	David Icke and Anti-Semitic Narratives
13	isolated, fraude, grootste, journal, humanity, virusen, onderzoek, february, published, wuhan	COVID-19 Origin and Fraud
14	warming, sgtnewsnetwork, global, co2, swindle, film, weather, narrative, social, falling	Climate Change Hoax

15	5g, source, mind, control, repair, chips, cellphones, inside, scam, guy	5G and Mind Control
16	fire, wheat, roemenie, fields, wall, burn, dumbsan-dunderground, 22, clearly, 13	Food Supply Sabotage
17	emf, dna, 5g, studies, electromagnetic, university, phone, exposure, cells, radio	Electromagnetic Radiation and Health