PhD-FSTM-2023-051
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 27/06/2023 in Esch-sur-Alzette
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

Ninghan CHEN

Born on 8 April 1994 in Guiyang, (China)

DECODING THE REAL WORLD: TACKLING VIRTUAL ETHNOGRAPHIC CHALLENGES THROUGH DATA-DRIVEN METHODS

Dissertation defence committee

Dr Jun PANG, dissertation supervisor
Professor, Université du Luxembourg

Dr Frank HOPFGARTNER
Professor, University of Koblenz

Dr Martin THEOBALD, Chairman
Professor, Université du Luxembourg

Dr Nikolaos ALETRAS
Professor, University of Sheffield

Dr Xihui CHEN, Vice Chairman
Research Associates, Université du Luxembourg

"Torture the data, and it will confess to anything."

Ronald Coase

Abstract

Decoding the Real World:
Tackling Virtual Ethnographic Challenges through Data-Driven Methods

by Ninghan CHEN

As the Internet has become an integral part of our daily lives, virtual world, particularly online social networks (OSNs), have evolved into crucial platforms for learning and idea exchange. These online interactions generate a considerable amount of data, reflecting real-world human behaviours and experiences. This leads to the question: how can this wealth of data from the virtual world be utilised to investigate real-world phenomena and challenges? One solution lies in virtual ethnography, an ethnographic approach enhanced by computational tools. In the past three years, the COVID-19 pandemic has brought to light a unique problem emanating from this abundance of data: the infodemic. This term, referring to an overabundance of information, accurate or not, has compounded the challenges presented by the pandemic. Misinformation and fake news have inundated OSNs, fostering confusion, fear, and harmful behaviours among individuals in the real world.

To combat the infodemic, governments and healthcare bodies have deployed interventions on OSNs. These interventions aim to amplify trustworthy information, control the spread of misinformation and fake news, and understand public sentiment and policy reactions. This thesis provides an exhaustive examination of the infodemic, employing Social Network Analysis (SNA) as the computational tool. It underscores the significance of three SNA applications—social characteristics, information diffusion, and sentiment analysis—in addressing the aforementioned OSN-based interventions. Concerning social characteristics, our objective is to identify users who genuinely contribute to information diffusion. We introduce two novel measures to evaluate the actual performance of individual users and user subgroups in diffusing COVID-19 information. We also shed light on the heightened mental distress experienced by influential users during the COVID-19 pandemic.

In terms of information diffusion, we propose two prediction models. These models consider the content of messages, users' susceptibility, and influence to estimate a message's eventual reach and identify users likely to disseminate the message. We validate our models through experiments, and the results indicate that our models surpass existing methods.

For sentiment analysis, we devise a Graph Neural Network-based text classification framework to extract vaccine attitudes from text posts on social media. We use the vaccine attitudes of users' friends as contextual information to minimise the interference of linguistic nuances like sarcasm. Lastly, to confirm the consistency between virtual world data and real-world phenomena, we conduct a comprehensive cross-validation. This involves comparing virtual and real-world data concerning COVID-19 vaccine hesitancy across different regions and time periods.

Acknowledgements

This thesis has been prepared within the framework of DTU DRIVEN. The Doctoral Training Unit "Data-driven computational modelling and applications" (DTU DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

This work marks not only the works of four years of my Ph.D. journey, but also a milestone in my research career and life. In addition to the papers and thousands of lines of code, this journey has enriched me with the ability to think, investigate, write, and present.

I would like to express my sincere gratitude to my supervisor Jun, for providing me the opportunity to pursue a Ph.D. under his guidance. His continuous encouragement to explore uncharted territories, and unwavering belief in my potential is invaluable. I am deeply grateful to Xihui for his dedication, invaluable guidance, knowledge, inspiring discussions, and collaborative research efforts that have contributed immensely to my journey. Thanks also to Sjouke for his encouragement and advice, which have been instrumental in my progress.

Thanks to Martin Theobald and Conchita D'Ambrosio for their continuous guidance and insightful discussions. I consider myself fortunate to have them as part of my CET committee, encouraging me to perform in-depth research.

Thanks to my defence jury, and particularly to Prof. Frank Hopfgartner and Prof. Nikolaos Aletras. I greatly appreciate Prof. Frank Hopfgartner's active involvement in my defence and his insightful feedback. My sincerest thanks also extend to Prof. Nikolaos Aletras. His mentorship, which began five years ago, has profoundly shaped my path and altered my destiny.

Thanks to everyone in the SaToSS group. As a vibrant and supportive group, you have been my valued companions throughout this journey. Thanks in particular to Reynaldo, Aoran, Semen, Sergiu, Sevdenur and Oscar. The memories of shared laughter and every dark joke over lunch will be cherished forever.

Thanks to my friends, both on-site and remote, especially Cui, Juntong, Siwen and Zhiqiang, your willingness to listen to my ramblings and indulge my culinary exploits have played a key role in maintaining my sanity.

Finally, I extend my profound gratitude to the two extraordinary people who brought me into this world: my dad and mum. Your love, encouragement, and unwavering support have nourished my body, mind, and soul. By the way, thanks to my cat, Sanwan. Despite your occasional misadventures in code deletion, your presence has been a constant source of comfort and companionship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Virtual and Real World

"In the near future, corporate networks reach out to the stars. Electrons and light flow throughout the universe. "

Ghost in the Shell

Since the World Wide Web's inception and its subsequent public accessibility, the virtual world has emerged as a revolutionary conduit for human communication. It has empowered individuals to share their opinions, insights, experiences, and perspectives through a plethora of Internet-based applications [WH08]. This development incited a persistent debate regarding the virtual world's relationship with the real one: are they distinct social spaces, or do they overlap? This question was humorously captured in a famed 1993 New Yorker cartoon, which posited: "On the Internet, no one knows you're a dog." This quip underscores the virtual world's inherent anonymity, highlighting its stark contrast with the tangible world.

However, this clear-cut distinction between the virtual and physical worlds is currently under reassessment, given the intensifying incorporation of the virtual domain into everyone's daily life [MS00]. The virtual world, initially viewed as a separate social space from other facets of human action and experience, is increasingly recognised as an integral part of the real world [GSBC09]. The traditional binary between the virtual and real worlds is becoming progressively obsolete as they interact and reciprocally influence each other [MS00]. This dynamic exchange is evident when real-world events spur extensive discourse in the virtual world, which in turn reverberates back into the real world. A salient example of this interplay is the role of social media platforms, such as Twitter and Facebook, in facilitating global social movements like #BlackLivesMatter and #MeToo. The discourse generated by these online discussions has had profound implications in the real world, leading to policy changes and heightened public awareness around various social issues [PBR19]. Consequently, the virtual world is increasingly perceived not as an isolated entity but as a component of the real world, further blurring the lines between these once distinct domains. This perspective is epitomised by the US Supreme Court's recognition of virtual social platforms like Twitter and Facebook as the new "public square" — primary spaces for communication, learning, and idea exchange [BM19]. Research has also shown that the virtual world has become a vital platform for mobilising people, creating social change, and garnering global support, from the Arab Spring to the war in Ukraine [SA13, PMFT22].

Since the virtual world can be used as an expression of human action and experience, *How to study phenomena and problems occurring in the real world through the virtual world?* To answer this question, virtual ethnography, an ethnographic methodology has been proposed. There are two dominant meanings of virtual ethnography [BH17]. The first refers to the traditional ethnographic study of digital cultures, in which researchers use participant observation and qualitative methods to study social phenomena in the virtual world [PHP$^+$15]. The second definition of virtual ethnography refers to the use of computational methods to enhance ethnographic research. This approach involves the collection and analysis of large amounts of data generated by social actions in the virtual world, using computational tools [BH17]. This mixed method of research can enrich and deepen the researcher's understanding of virtual world data. In this thesis, we focus on the second meaning of virtual ethnography.

1.2 COVID-19 Infodemic and Virtual Ethnography

In the past three years, the world has experienced an unprecedented COVID-19 pandemic. As a result of physical isolation and social detachment, people increasingly rely on online social networks(OSNs) for information, advice, and mental health support. However, OSNs have also become a breeding ground for misinformation and fake news, leading to risky behaviours that endanger public health. While the pandemic is receding, it is still critical to take stock of lessons for future large-scale infectious diseases of similar type. These experiences include both approaches to fighting the virus itself and to combating the *infodemic* caused by the virus. The term "infodemic" outlines the perils of misinformation during disease outbreaks mainly on social media [CQG$^+$20, HHW$^+$21]. Apart from accelerating virus transmission by distracting social reactions, the infodemic increases cases of psychological diseases such as anxiety, phobia and depression during the pandemic [DBG$^+$20].

In response to the infodemic, governments and healthcare institutions have enacted social media-based interventions involving *amplifying of reliable information*, *controlling the spread of misinformation and fake news*, and *understanding public sentiment and reactions to policies*. Virtual ethnography, a widely employed research methodology, empowers researchers to probe how users form relationships on social media, examine their information diffusion behaviour, and analyse user experiences and impacts during the pandemic.

Nevertheless, with the burgeoning volume of data generated by OSNs, the deployment of virtual ethnography presents two significant challenges. The primary challenge involves the collection of large, comprehensive data sets for research [LPA$^+$09]. Given that analyses hinged on the virtual world are contingent on extensive data sets, conventional data collection methods such as manual statistics fall short. This necessitates the exploration of novel data collection methods for large-volume data.

The second challenge pertains to the selection of appropriate tools to analyse the vast quantities of data stemming from social platforms [BC12, CR10]. Unlike traditional research methods such as surveys and interviews, researchers often lack direct access to users' attitudes or sentiments when handling social platform data. Instead, they must extract relevant information from these platforms, which may appear in various formats such as text, social networks, and videos. Consequently, researchers need to opt for more efficient and accurate tools to extract and analyse the necessary information.

1.3 Social Network Analysis

To address the second challenge stated above, Social Network Analysis (SNA) emerges as a compelling solution. SNA encapsulates a set of advanced computational methods crafted to study various social structures and discern associations among participants within a community [OR02]. This analytical approach affords crucial insights for the selection of participant samples and the comprehension of participant interactions in virtual ethnographic studies. Fundamentally, SNA relies on graph theory, providing a mathematical framework that aptly describes network structures and interaction patterns between individual nodes. SNA finds diverse applications in OSNs, spanning areas such as social characteristics, which emphasise the characterisation and quantification of individuals within a network, their interrelationships, and the broader network structure; Information diffusion, focusing on interaction patterns and content exchange among users; and sentiment analysis, centring on individual user behaviour. In this thesis, we embark on a thorough exploration of the pivotal roles these three applications of SNA play in combating the infodemic. First, the strategy for amplifying reliable information can be shaped by a thorough analysis of social characteristics. This will involve the identification and quantification of influencers who can facilitate the effective spread of trustworthy information. Second, we tackle the issue of controlling the diffusion of misinformation and fake news by gaining a comprehensive understanding of information diffusion patterns. This will involve the study and prediction of information diffusion patterns among users within the social network. Third, we achieve a nuanced understanding of public reactions and attitudes towards policy measures through exhaustive sentiment analysis. This process requires extraction and examination of individual user sentiment and the overall sentiment trends within the public domain. This investigation will provide invaluable insights into the interplay between social networks and public health crises, as well as guide the development of strategies to manage future public crises.

1.3.1 Social Characteristics

The examination of social characteristics is pivotal in understanding individual roles and behaviours in OSNs [Cen10]. As we grapple with the infodemic, the significance of understanding these social characteristics intensifies. The strategic amplification of trustworthy information can be notably improved through a comprehensive understanding of these characteristics. The identification and quantification of influential nodes can facilitate the efficient diffusion of information.

Social characteristics hone in on individual nodes or subgroups within the network, offering insights into key users and their respective influence within the network. A basic characteristic is the node's degree [Fre78], representing the number of connections a node has with other nodes. This metric is instrumental in identifying well-connected nodes or key influencers. More advanced metrics like centrality [Fre78, ZL17] and PageRank [PBMW99] estimate the significance of nodes within the network. Centrality incorporates both the number of connections and the significance of these connections. In contrast, PageRank is an iterative algorithm that allocates weights to connections based on the importance of users. These metrics assist in revealing the hidden structure of relationships within the network and identifying the most influential users within the community. Examining social characteristics in virtual ethnographic research is critical for understanding the complex interactions between nodes, their relationships, and the overarching structure of OSNs.

1.3.2 Information Diffusion

Information diffusion is the process by which information, influences, and other forms of content spread from one user to another in social networks [BRMA12, GGLT04]. By analysing information diffusion in a social network, we can gain insights into how information spreads, identify key sources, and understand how this spread affects individual behaviour. In the context of an infodemic, understanding the patterns and mechanisms of information diffusion is essential to control the spread of misinformation and fake news. By studying these diffusion patterns, we can identify the primary sources and channels of misinformation, predict its potential spread, and develop effective strategies to counter it.

The diffusion model typically consists of three elements: sender, receiver, and transmission medium [Cho15]. In the case of OSNs, senders and receivers are users within the network, while the OSNs serve as the diffusion medium. Several classical information diffusion models exist, including threshold models [Gra78], independent cascade models [GGLNT04, KKT03], and epidemic models [KE05]. Generally, at the beginning of the diffusion process, all nodes are inactive. When an original node sends a message, it becomes active and influences connected neighbouring nodes, possibly activating them based on the chosen diffusion model's activation mechanism. This iterative process continues until the diffusion stops due to no more nodes being activated or a predefined stopping condition is met.

Threshold models assert that a user becomes active or adopts information when the proportion of active neighbours reaches a certain threshold [Gra78]. The activation threshold varies between nodes, representing the resistance to adopting information or social pressure required to change a node's state. Threshold models have been widely used to study the spread of opinions, innovations, and information in social networks. Independent cascade models represent information diffusion as a process analogous to the spread of diseases or infections [KKT03]. In these models, the probability of a user becoming active depends on the number of active neighbours. Once active, a user has a single chance to activate each neighbour. Independent cascade models are useful for studying viral content or opinion spread on social media platforms. Epidemic models, inspired by infectious disease spread, assume that diffusion occurs through direct contact between active and inactive nodes [Mol95]. The diffusion process follows a SIR (Susceptible, Infected, Recovered) model [Bai75], with nodes transitioning from susceptible (inactive) to infected (active) and eventually recovering (inactive again) over time.

Several factors can affect information diffusion in a social network, such as network structure, node attributes, content features, and diffusion dynamics. Network structure, including the number of nodes, connection density, presence of communities, and distribution of centralities, significantly impacts information spread [BLL+14]. For example, a highly connected network may facilitate faster and wider diffusion than a sparsely connected network. Individual node characteristics, such as influence, credibility, and expertise, also affect diffusion. Influential nodes with numerous connections can amplify information spread, while less influential nodes may have limited impact. Content features, including novelty, relevance, and emotional appeal, influence the rate and extent of information diffusion. Emotionally charged content may spread quickly and widely within a network [BWJ+17]. Finally, diffusion dynamics, such as timing, sequence, and duration of information spread, can affect overall diffusion patterns [NLL14, CSG+20]. Rapid information spread may create a burst of activity, while slower spread may result in a more gradual diffusion process.

1.3.3 Sentiment Analysis

Sentiment analysis is a broad field that involves studying and analysing people's perspectives, attitudes, and feelings towards entities and their attributes as expressed in written texts [AXV+11]. Sentiment analysis includes not only the analysis of whether the emotions are positive or negative, anxious or happy but also the attitude or opinion expressed about something, such as support or opposition. In the face of an infodemic, sentiment analysis becomes an indispensable tool for understanding public reactions and attitudes towards policy measures, health guidelines, and other pandemic-related topics. By analysing the sentiment of social media posts, comments, and reactions, we can gauge the public's overall sentiment towards these issues and identify prevailing trends in public opinion.

The development of sentiment analysis can be divided into four main levels, document level [Tur02], sentence level [HL04], aspect level [SF16], and social context level [SRI19]. At the document level, the entire document is treated as a single entity, and the analysis is applied to the document as a whole to determine an overall sentiment [PL04].

In sentence-level sentiment analysis, each sentence within the document is treated as an individual entity, and the sentiment analysis methods are applied to each sentence separately to determine the sentiment expressed in that sentence. Once the sentiment of each sentence has been analysed, the results are then summarised to provide an overall sentiment for the document [KB06, HZ11].

Aspect-level sentiment analysis, also known as feature-based sentiment analysis, is a technique used to identify and analyse the sentiment expressed towards specific aspects or features of an entity. At this level, the goal is to identify all entities and their aspects in a text and assign a polarity score to each aspect of the entity [TQL16]. The main challenge in aspect-level sentiment analysis is to accurately distinguish between the sentiment expressed towards different aspects when there are multiple aspects in a text. For example, in the sentence "This hotel offers decent amenities, but the service is terrible", the sentiment towards the amenities and the service are different aspects, and the sentiment towards the amenities is positive, while the sentiment towards the service is negative. Aspect level can be divided into three steps: identification, classification and aggregation. The identification step is to extract aspect terms [JG10]; the classification step is to classify the sentiment of each aspect [TQL16]; and the aggregation step is to combine the sentiment of each aspect [LYL19].

At social context level, the analysis takes into account the social context in which the text was created by a user, including the personal characteristics and preferences of the user, his/her past behaviour and attitudes, and the social network. Social context level sentiment analysis can be used to analyse the sentiment expressed by individual users towards a particular topic or event, taking into account their past posts and the posts from his/her neighbourhoods [SFG+15, GHW17]. This approach can provide more accurate and personalised sentiment analysis results, which can be useful for targeted analysis of attitudes towards specific topics, such as vaccines or political issues.

1.4 Research Questions

As discussed in the previous section, to combat the infodemic, governments and healthcare organisations start to deploy OSNs responses. However, these responses fundamentally hinge upon a comprehensive understanding and analysis of OSNs, which in turn, are predicated on the availability of substantial data. This reality foregrounds the first research question we aim to address:

Research Question 1

How can we gather a large amount of data?

OSNs have become a crucial communication channel for healthcare organisations to disseminate official guidelines and professional advice, as well as information on effective COVID-19 prevention measures such as masking, vaccination, and social distancing. To combat the infodemic, individual users, healthcare professionals, and social activists with large followers were encouraged to share this information on social media. However, the effectiveness of these users in promoting the speed and extent of information diffusion can vary greatly, making it challenging to measure their performance in this regard. Here, we use the term 'bridging performance' as an analogy to evaluate how quickly and widely information can diffuse on social media because of the sharing of a user. Traditional social characteristics methods may not accurately capture the bridging performance of individual users in spreading crisis-related information. For example, medical professionals on Twitter may not have thousands of followers like super tweeters, but their professional endorsement can greatly contribute to the popularity of the tweets they share. This leads to our second question:

> Research Question 2
>
> How to design a measurement to capture the actual bridging performance of social media users in terms of spreading COVID-19-related information?

In addition to spreading accurate and useful information, combating the spread of false information and disinformation is also critical for fighting against the infodemic. To achieve this, cascade prediction has emerged as a widely accepted approach to understanding the prevalence of information based on its early adopters. By predicting the spread of information, healthcare authorities can quickly respond to misinformation before it causes significant harm. Studies have indicated that when a user decides to retweet a message, three factors come into play: the message content, the influence of active friends, and the user's susceptibility. A user's influence reflects their ability to persuade others to share their message, while susceptibility refers to the likelihood of the user being influenced by others. However, existing cascade prediction models only consider the influence of active friends and the user's own susceptibility, ignoring the content of the message itself. Furthermore, these models fail to account for the fact that a user's influence and susceptibility are not only specific to the user but also to the topic or content of the message. For example, a sports journalist who prefers pop music may be more influential when tweeting about sports than music-related topics. As a recipient of information, the journalist may be more cautious in sharing sports news than music-related tweets. To address these limitations, we propose our third research question:

> Research Question 3
>
> How to accurately predict the popularity of information?

In the fight against the infodemic, understanding public attitudes and reactions to policies is crucial. One such policy that has garnered a lot of attention is vaccination, which is an effective measure to prevent pandemics caused by infectious diseases like COVID-19. Research has shown that information on social media plays an important role in shaping people's attitudes toward vaccination, and thus, it is important to analyse public attitudes towards vaccines on social media platforms. To achieve this, researchers have conducted numerous studies using virtual ethnography to analyse pro- and anti-vaccine users on social media, to understand public attitudes toward vaccination and to track their changes. Sentiment analysis in SNA is found to be a useful tool for researchers to identify the sentiments expressed in texts about vaccination. However, existing sentiment analysis methods for vaccine attitudes face three main challenges. First, the global nature of pandemics requires an approach that can handle multilingual texts. Most feature-based approaches, which classify posts by predefined text features (e.g., keywords), typically focus on a single language, leading to the failure of such approaches when it comes to analysing multilingual content data. Second, linguistic features, such as sarcasm and irony, are prevalent in vaccine-related posts during the pandemic, which significantly compromises the performance of existing sentiment analysis models. For example, a user might express support for vaccination by ridiculing the "microchip for vaccination" conspiracy theory, saying "I won't do it for their vaccines, I'm waiting for 6G". Such sarcasm can be difficult to detect with traditional sentiment analysis methods. Third, research has shown that users' attitudes toward vaccination are closely related to those of their neighbours in social networks, including friends and friends of friends. Those with negative attitudes toward vaccination tend to be associated with users with negative attitudes, while those with positive attitudes tend to connect with users with positive attitudes. However, trying to detect users' vaccine attitudes through communities fails to capture real-time vaccine dynamics, as people's opinions can change rapidly based on new information and events. To address these three challenges, this leads to our fourth research question:

> Research Question 4
>
> How to accurately extract users' attitudes from their posts?

When it comes to obtaining data on vaccine attitudes and hesitancy, researchers have the option of using surveys or social media to gather information. However, both methods are susceptible to potential bias and errors. Social media data, in particular, has been found to have three possible sources of error: *measurements*, *coding* and *missingness* [HM17, Bak17]. Measurement errors can occur when social media users do not express their true attitudes in their posts. Coding errors may arise from flaws in the methods used to capture public opinion. missingness can occur due to non-representative social media users, where not all people express their opinions online. For example, Twitter tends to have a younger user base, while Facebook attracts more older users. Given the potential for error in social media data, it is important to develop appropriate methods to correct the bias and accurately measure users' vaccine attitudes from social media. We, therefore, pose our fifth research question:

> Research Question 5
>
> Is it possible to accurately measure individuals' vaccine hesitancy from social media using appropriately designed methods?

Part I		Part II	Part III		Part IV	
RQ 1		RQ 2	RQ 3		RQ 4	RQ 5
Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8	Chapter 9

Table 1.1: Thesis structure.

1.5 Thesis Overview

This thesis is organised into four parts. In Part I, we provide an overview of the data collection method and datasets used in this study, followed by an exploratory and quantitative analysis of the datasets. Chapters 3 and 4 focus on addressing research question 1. The second part of the thesis focuses on social characteristics in OSN. Chapter 5 explores the quantification of the bridging performance of users, which is relevant to research question 2. Part III proposes two prediction models for information diffusion in Chapters 6 and 7, respectively. In Chapter 6, we focus on cascade prediction, examining the spillover effect of information exposure on users' decisions to participate in the diffusion of specific information. In Chapter 7, we introduce a novel deep learning cascade prediction model, CasSIM, which is capable of simultaneously achieving two highly sought-after objectives: popularity prediction and final adopter prediction. These models are developed based on social network features and aim to address research question 3. The final part of the thesis, Part IV, focuses on sentiment analysis. In Chapter 8, we address research question 4 by successfully implementing the extraction and continuous tracking of users' vaccination stances using a deep learning framework that utilises text posts on Twitter. In Chapter 9, we explore research question 5, wherein we validate the potential of social media data to complement social surveys in monitoring public hesitancy regarding the COVID-19 vaccine. Table 1.1 outlines the structure of this thesis and the contributions of each chapter are detailed below.

- In Chapter 2, we introduce the preliminary knowledge of graph and graph neural networks (GNNs) used for information prediction and sentiment analysis. We also introduce the necessary knowledge of cascade and cascade prediction, including the definition of the problem and basic methods.

- In Chapter 3, we introduce our proposed data collection method for two Twitter datasets, which is specifically designed to gather detailed information on the COVID-19 period, such as user activity in a specific geographic location, as well as their social network. Furthermore, a segment of the dataset is annotated to provide additional insights into users' attitudes towards COVID-19 vaccines. Our primary objective is to produce high-quality datasets that can support various virtual ethnography studies pertaining to COVID-19 on OSNs.

- In Chapter 4, in order to assess the suitability of our dataset for virtual ethnography studies we explore the temporal and spatial distribution of the collected datasets and conduct a linguistic distribution analysis using a variety of figures. Alongside these data visualisation, we also undertake several analyses to demonstrate the potential applications of these datasets in facilitating virtual ethnography studies. Our results indicate a correlation between the volume of tweets and COVID-19 cases in the GR and the country of interest, although this association is only evident during specific periods of the pandemic. Furthermore, we charted topic changes in each country and region at the onset of the pandemic and identified notable differences between the GR and the country of interest.

- In Chapter 5, we concentrate on identifying influential users on Twitter who have a significant impact on the spread and popularity of trustworthy information, and who have played a crucial role in countering the adverse effects of misinformation. Furthermore, we investigate the influence of the COVID-19 pandemic on the subjective well-being (SWB) of these users.

- In Chapter 6, we centre around casacde prediction, where we investigate the spillover effect of information exposure on users' decisions to participate in the diffusion of certain information. We focus on the diffusion of information related to COVID-19 preventive measures due to its special role in consolidating public efforts to slow down the spread of the virus. Through the Twitter dataset we collected, we validate the existence of the spillover effects. Building on this finding, we propose extensions to three cascade prediction methods based on GNNs. Experiments conducted on our dataset demonstrated that the use of the identified spillover effects significantly improves the state-of-the-art GNNs methods in predicting the popularity of not only preventive measure messages but also other COVID-19 messages.

- In Chapter 7, we propose a new deep learning cascade prediction model CasSIM that can simultaneously achieve two most demanded objectives: popularity prediction and final adopter prediction. Compared to existing methods based on cascade representation, CasSIM simulates information diffusion processes by exploring users' dual roles in information propagation with three basic factors: users' susceptibilities, influences and message contents. With effective user profiling, we are the first to capture the topic-specific property of susceptibilities and influences. In addition, the use of GNNs allows CasSIM to capture the dynamics of susceptibilities and influences during information diffusion. We evaluate the effectiveness of CasSIM on three real-life datasets and the results show that CasSIM outperforms the state-of-the-art methods in popularity and final adopt prediction.

- In Chapter 8, we leverage the textual posts on Twitter to extract and track users' vaccination stances in near real time by proposing a deep learning framework. To address the impact of linguistic features such as sarcasm and irony commonly used in vaccine-related discourses, we integrate into the framework the recent posts of a user's social network neighbours to help detect the user's genuine attitude. Based on our annotated dataset from Twitter, the models instantiated from our framework can increase the performance of attitude extraction by up to 23% compared to state-of-the-art text-only models. Using this framework, we successfully validate the feasibility of using OSNs to track the evolution of vaccination attitudes in real life.

- In Chapter 9, we validate whether social media data can be used to complement social surveys to monitor the public's COVID-19 vaccine hesitancy. Taking advantage of recent artificial intelligence advances, we propose a framework to estimate individuals' vaccine hesitancy from their social media posts. With the vaccine-related tweets from our Twitter dataset, we compare vaccine hesitancy levels measured with our framework against that collected from multiple consecutive waves of surveys. We successfully validate that Twitter can be used as a data source to calculate consistent public acceptance of COVID-19 vaccines with surveys at both country and region levels. In addition, this consistency persists over time although it varies among socio-demographic sub-populations. Our findings establish the power of OSNs in complementing social surveys to capture the continuously changing vaccine hesitancy in a global health crisis similar to the COVID-19 pandemic.

- Chapter 3 is based on paper entitled "A Multilingual Dataset of COVID-19 Vaccination Attitudes on Twitter" [CCP22a], published in the journal of Data in Brief.
- Chapter 4 is based on paper entitled "An Exploratory Study of COVID-19 Information on Twitter in the Greater Region" [CZP21], published in the journal of Big Data and Cognitive Computing.
- Chapter 5 is written using the content of two papers [CCZP22a]: one was published in the proceedings of 2022 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), and the other one was submitted to ACM Transactions on the Web.
- Chapter 6 is written using the content of two papers [CCZP21, CCZP22c]: one was published in the proceedings of 2021 International Conference on Advances in Social Networks Analysis and Mining (ASONAM) and the other one was published by the journal of Entropy.
- Chapter 7 is based on paper entitled "A Tale of Two Roles: Exploring Topic-specific Susceptibility and Influence in Cascade Prediction", is accepted by the journal track of 2023 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD).
- Chapter 8 is based on paper entitled " 'Double vaccinated, 5G boosted!': Learning Attitudes towards COVID-19 Vaccination from Social Media.", is under major revision of ACM Transactions on the Web.
- Chapter 9 is based on paper entitled "Measuring COVID-19 Vaccine Hesitancy: Consistency of Social Media with Surveys." [CCP$^+$22b], published in the proceedings of 13th International Conference on Social Informatics (SocInfo).

Chapter 2

Preliminaries

In this chapter, we provide an overview of essential background information on key concepts related to graphs and Graph Neural Networks, as well as cascades and cascade prediction. This material is presented in a separate introductory chapter because it forms the foundation for much of the thesis's subsequent development. The background knowledge required for the rest of this thesis, including topics such as subjective well-being mining, users' bridging performance in information diffusion, and vaccine hesitancy extraction, is more focused within specific chapters. This chapter aims to concentrate on the crucial subset of concepts required to understand our contributions and discussions throughout the thesis, rather than offering a comprehensive summary of the topics covered in this work.

2.1 Graph Representation

In this section, we present the fundamental concepts of graphs. A graph serves as a versatile structure for modelling relationships among entities across numerous domains, including business networks, biological systems, and social networks, which are the primary focus of this thesis.

Definition 1 (Graph)**.** *A graph can be formally defined as* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$*, where* $\mathcal{V}$ *represents the set of nodes,* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ *denotes the set of edges, and* $\mathcal{X} = \{x_i \in \mathbb{R}^m \mid i \in \{1, 2, \ldots, n\}\}$ *signifies the node features, with* $n = |\mathcal{V}|$ *and* m *indicating the dimensionality of the node features. A graph can be summarised by an adjacency matrix* $A \in \mathbb{R}^{n \times n}$*.* $\forall i, j \in \{1, 2, \ldots, n\}$*,* $A_{ij} = 1$ *if there is an edge between nodes* i *and* j*, otherwise* $A_{ij} = 0$*.*

For simplicity, many graph learning models assume that there are no self-connections, i.e., edges from and to the same node. For instance, in a social network graph $\mathcal{G}$, a node $v \in \mathcal{V}$ represents a social media user, while an edge (v, v') signifies a relationship between users v and v'. Node features $\mathcal{X}$ may encompass user-posted text messages or demographic data about the users, and users can not follow themselves.

Moreover, graphs can be either directed or undirected, depending on whether the edges have an orientation. In an undirected graph, edges have no orientation and are represented as unordered pairs of nodes. In contrast, a directed graph (also known as a digraph) contains edges with an orientation, denoted as ordered pairs of nodes. The directed nature of edges in a digraph implies an asymmetry between nodes, whereas undirected graphs exhibit symmetric relationships.

Definition 2 (k-hop Neighbourhood)**.** *Let $\mathcal{N}_i^k$ be the set of neighbours of node v_i within k hops, i.e., $\{v \mid d_{\mathcal{G}}(v, v_i) \leq k\}$ where $d_{\mathcal{G}}(v, v_i)$ is the shortest distance between v and v_i in the graph $\mathcal{G}$. Note that node v_i is also in $\mathcal{N}_i^k$ as $d_{\mathcal{G}}(v_i, v_i) = 0$.*

Comprehending the definition of neighbours is vital for the subsequent section, as it elucidates the mechanism of the graph learning model and formalises the path of information propagation.

2.2 Graph Neural Networks

Graph neural networks (GNNs) are neural network models designed for processing data that is structured in the form of a graph [SGT+09]. The goal of GNNs is to learn a representation vector for a node h_v or the entire graph h_G, using the graph structure and node features $\mathcal{X}$. The central concept of GNNs is the message-passing framework, where nodes aggregate information from their neighbours $h_u^{(k)}$, with $u \in \mathcal{N}(v)$, to update their own representations $h_v^{(k)}$.

Initially, node representations h_v^0 are assigned the input node features $h_v^{(0)} = X_v$. After k iterations of message passing or aggregation, a node's representation captures the information within its k-hop neighbourhood.

$$a_v^{(k)} = \mathsf{AGGREGATE}^{(\mathsf{k})}(\{h_u^{(k-1)} : u \in \mathcal{N}(v)\}) \tag{2.1}$$

$$h_v^{(k)} = \mathsf{COMBINE}^{(\mathsf{k})}(h_v^{(k-1)}, a_v^{(k)}) \tag{2.2}$$

where $h_v^{(k)}$ denotes the feature vector of node v during the k-th iteration, and $\mathcal{N}(v)$ represents the neighbourhood of node v. $\mathsf{AGGREGATE}^{(\mathsf{k})(\cdot)}$ is the aggregation function (e.g., summation, average, or maximum) utilised by each node to aggregate messages from its neighbours. $\mathsf{COMBINE}^{(\mathsf{k})}(\cdot)$ is the node update function, where each node updates its representation based on the aggregated messages and its current representation, typically using a nonlinear transformation like a neural network layer.

With the representation vector of every node at the k-th layer, the representation of the graph $\mathcal{G}$ can thus be calculated by a function as follows:

$$h_{\mathcal{G}} = \mathsf{READOUT}(\{h_v^k \mid v \in \mathcal{V}\}) \tag{2.3}$$

The READOUT function can be simply implemented as the mean of nodes' vectors or other complex pooling functions depending on the specific requirements of scenarios in practice.

In the main thesis, we will investigate the construction of AGGREGATE and READOUT to improve model representation and prediction capabilities for diverse downstream tasks. Here, we present two well-known GNNs variants: Graph Convolutional Networks (GCN) [KW17] and Graph Attention Network (GAT) [VCC+18].

GCN is a semi-supervised learning algorithm for graph representation and GAT is a variant of GCN which introduces the attention mechanism to distinguish the significance of neighbours. In GCN, element-wise mean pooling is employed for aggregating messages. For each node v_i, aggregate messages from its neighbors $u \in \mathcal{N}(v_i)$ using the element-wise mean pooling operation:

$$a_v^{(k)} = \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} h_u^{(k-1)} \tag{2.4}$$

$$h_v^{(k)} = \sigma(\boldsymbol{W}^{(k)} a_v^{(k)}) \tag{2.5}$$

where $\boldsymbol{w}^{(k)} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{d}}$ is a trainable weight matrix of layer k, σ is a non-linear activation (ReLU) and $\|$ is the concatenation operator. The process is repeated for a predefined number of iterations or until convergence, allowing the nodes to capture information from increasingly distant neighbours.

In contrast, GAT refines the message aggregation step by incorporating the attention mechanism. This enables the model to assign different weights to neighbours based on their importance. GAT adaptively learns a normalisation coefficient, α_{uv}, for edge $\mathcal{E}_{uv}$, to aggregate feature vectors from node v's neighboring nodes, $N(v)$, and combine them with its own feature vector. The update steps for GAT are similar to those for GCN, with the primary difference lying in the message aggregation step:

$$a_{uv}^{(k)} = \text{Softmax}(\text{LeakReLU}(\gamma[\boldsymbol{W} h_u^{(k-1)} \| \boldsymbol{W} h_v^{(k-1)}])) \tag{2.6}$$

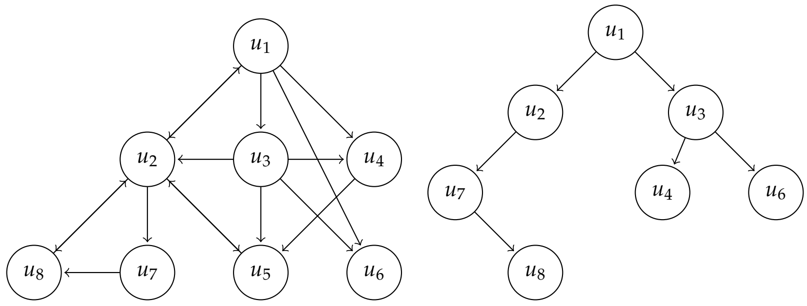
$$h_v^{(k)} = \sigma(\sum_{(u,v) \in \mathcal{E}} a_{uv}^{(k)} \boldsymbol{W}^{(k)} h_u^{(k-1)}) \tag{2.7}$$

where $\boldsymbol{w}^{(k)} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{d}}$ is a trainable weight matrix of layer k, σ is a non-linear activation (ReLU) and $\|$ is the concatenation operator.

2.3 Cascade and Cascade tree

A cascade records the process of the diffusion of a message. It stores all activated users and the time when they are activated. In this thesis, a user is activated in diffusing a message when he/she retweets the message. When a message m is firstly posted by a user, it will be perceived by the user's followers who will adopt the message and relay the message. This cascading process will continue on the social graph until no further sharing occurs.

Definition 3 (Cascade)**.** *A diffusion cascade of m at the time window T is a node set of users who adopted m. $C_m^T = \{u_1, u_2, \ldots, u_{n_T^m}\}$. Note that n_T^m is the number of adopters of m in time window T.*



(a) A social graph (b) A cascade

Figure 2.1: Example of a cascade.

A cascade tree is a directed, acyclic graph that represents the structure of a cascade, capturing the relationships between nodes involved in the diffusion process [WSL+17]. In a cascade tree, nodes correspond to the users participating in the cascade, and directed edges represent the diffusion of information between nodes. The first user who posted the message is regarded as the root of the cascade tree. Users who retweeted the message, but received no further retweeting comprise the leaf nodes. Note that a tweet with the quotation to another tweet is also considered as a retweet of the quoted message. An edge from u to u' is added to the cascade if u' follows u and u' re-tweeted the message after u, indicating u activated u'. If many of the users who u' follows ever retweeted the message, meaning u' may be activated by any of them, we select the one who lastly retweeted as the parent node of u'. Figure 2.1(b) shows a cascade of the social network in Figure 2.1(a). In this example, user u_4 can be activated by the messages retweeted by either u_1 or u_3. Since u_3 retweeted after u_1, we add the edge from u_3 to u_4 indicating that the retweeting of u_3 activated u_4.

We denote the root node of a cascade C by $r(C)$. We call a path that connects the root and a leaf node a *cascade path*, which is actually a sequence of nodes ordered by their activation time. For instance, (u_1, u_3, u_4) is a cascade path in our example indicating that the diffusion of a message started from u_1 and reached u_4 in the end through u_3. In this thesis, we represent a cascade tree as a set of cascade paths. For instance, the cascade in Figure 2.1(b) is represented by the following set $\{(u_1, u_2, u_7, u_8), (u_1, u_3, u_4), (u_1, u_3, u_6)\}$.

2.4 Cascade Prediction

Cascade prediction is a method used in social network analysis to predict the spread of information or influence within a network. Let $\mathcal{M}$ be a set of messages. We use the term "*message*" to refer to a piece of information that can be disseminated over social media. It can be a tweet on Twitter or an image on Instagram. In this thesis, we focus on textual messages and our approach can be straightforwardly extended to other message types if their representations can be effectively calculated and adapted. For any message $m \in \mathcal{M}$, we have the set of active users that had adopted this message up to the time t_0 after the message was first posted, denoted by $C_m^{t_0}$.

The observation time t_0 depends on the requirements of downstream applications as well as the popularity of social media platforms. It can be of hours on Twitter and Weibo, and years for citation networks. In general, the cascade prediction methods can be divided into two classes: popularity prediction and final adopter prediction.

Popularity prediction. The problem of *popularity prediction* is to predict the final number of active users, i.e., $n_m^\infty = |C_m^\infty|$. In practical applications, the final time can be determined as a given fixed time period t or by approximating the cessation of growth or slow growth of the cascade. This is also commonly called the *macroscopic cascade prediction* in the literature [YCW$^+$15, YTS$^+$19].

Final adopter prediction. The goal is to predict the set of users who will forward the target message. This is different from the *microscopic cascade prediction* in the literature [YCW$^+$15, YTS$^+$19] which aims to predict the next active user according to the observed ones.

For both types of problems, methodologically, existing approaches to cascade prediction problems can be categorised into three main groups: diffusion model-based approaches, generative approaches, and cascade representation-based models.

Diffusion model-based methods. This line of methods iteratively run their diffusion models to simulate the information propagation process as viral contamination [PMV20]. Two typical diffusion models are used: Independent Cascade (IC) [SHL17, WSLC15] or Linear Threshold (LT) [KKT03]. Earlier stochastic methods require manual assignment of influence probabilities for each user pair, which is not tractable in practice. To address the deficiency, embedding learning-based methods are proposed such as TIS [WSLC15] and EMBED-IC [BLG16] and CELFIE [PMV20]. User-specific susceptibility and influences are represented as latent parameters which are estimated according to observed cascades. The activation of a user can thus be determined by users' susceptibility and influence vectors. One advantage of such methods is that they can well characterise the diffusion process and output the activation state of each user. However, they suffer from high computation overhead and the strong assumptions on diffusion models makes them suboptimal for cascade prediction [YTS$^+$19, SRZ$^+$22, CZZ$^+$19].

Generative methods. With The time stamps of users' sharing behaviours, a cascade of early adopters is abstracted as an event sequence and thus temporal point processes can be applied to simulate the arrivals of events. According to the employed point processes, we have two types of generative methods: the ones based on Reinforced Poisson process [SWSB14] and those based on the self-exciting Hawkes process [CSC$^+$17]. Due to the assumption of temporal point processes, this line of methods over-simplify information diffusion and are thus limited in prediction performances.

Cascade representation-based models. The idea of this class of methods extract features of observed cascades as representation vectors and employ machine learning models to infer cascades dynamics. Earlier works rely manually crafted features from user profiles [CJY$^+$13] and message contents [HDD11]. Deep learning overtakes feature-engineering methods recently due to its overwhelming performance. DeepCas [LMGM17] is the first end-to-end deep learning method for popularity prediction. It samples diffusion paths from cascade graphs and make use of recurrent neural networks (RNN) to embed these sequential paths. Following DeepCas, a number of methods are proposed by extending RNNs to calculate cascade representations by interpreting cascades as sequences [WZLC17, YTS$^+$19, WCL18]. With social relations between adopters, some studies model cascades as cascade graphs and use various methods to calculate their embedding vectors with more effective sampling methods [TLH$^+$21] or graph embedding methods [CZZ$^+$19, SRZ$^+$22, XZZ$^+$23]. In spite of their promising performance, deep learning cascade prediction faces some inherent challenges as stated in [CZZ$^+$19, TLH$^+$21, ZYXT21]. New methods are continuously developed to address them. For instance, Zhou et al. [ZYXT21] studied the impact of the long-tailed distribution of cascade sizes on cascade prediction. In addition, except FOREST [YTS$^+$19], deep learning based methods focus on either popularity prediction or microscopic prediction, i.e., forecasting the next single adopter, and thus cannot predict popularity and final adopters simultaneously. Cao et al. [CSG$^+$20] proposed a different approach CoupledGNN by modelling the cascading effects with GNNs [KW17], i.e., users' sharing behaviours are influenced by their neighbours in social networks. However, this method oversimplified the diffusion process by ignoring users' double roles in information diffusion and thus produced suboptimal prediction performances.

Part I

Dataset Collection and Multi-faceted Exploration

Chapter 3

Data Collection

In the previous chapter, we highlight how governments and healthcare organisations can leverage these insights to implement effective responses, including diffusing trustworthy information, controlling the spread of misinformation, and understanding public opinions and reactions to policies. However, one of the challenges in virtual ethnographic research is obtaining data that meets the requirements of the study, such as containing information about the user's social networks along with all the messages posted by the user over time.

In this chapter, we address this challenge by describing a dataset collection method for Twitter and presenting two comprehensive datasets spanning for two years. These datasets serve as a rich resource for researchers studying various aspects of virtual ethnography, such as social characteristics, information diffusion, and sentiment analysis. These aspects can be explored to gain a deeper understanding of user behaviour and interactions on social media platforms, particularly during critical events like pandemics.

In summary, our datasets offer a valuable resource for researchers seeking to understand the use of social media during the COVID-19 pandemic. It can inform more effective responses to future public health crises by providing insights into the role of OSNs in shaping public opinion and the potential for social media-based interventions.

3.1 Introduction

In this chapter, we outline the method employed to collect the Twitter dataset and present two datasets compiled using this method for various downstream research purposes, spanning a two-year period from October 22, 2019, to December 31, 2021. Our datasets can offer a comprehensive understanding of how individuals utilised Twitter for communication during the COVID-19 pandemic. Alongside the posted messages, we include users' social networks on Twitter, enabling researchers to examine the spread of information across these networks and assess the influence of users' surroundings on their perspectives. Additionally, we choose vaccine attitudes as a case study and manually annotate 17,934 tweets to encompass perceptions of vaccination (i.e., positive, negative, and neutral). This annotated data delivers a

Table 3.1: Social media datasets related to COVID-19 vaccines.

Dataset	Period	Platform	Size	Annotation	Region	Topic
Pierri et al. [FSMS21]	2020/12/20-2021/03/13	Twitter/ Facebook	3M/10M	No	Italy	COVID-19
CoVaxxy	2021/01/12-up to date	Twitter	Real-time updates	No	USA	Misinformation
MMCoVaR	2020/02/1-2021/05/08	Twitter/News articles	24,184	Yes	Worldwide	Misinformation

more detailed and nuanced understanding of how individuals perceive and discuss vaccines on social media platforms.

Distinct from existing datasets, we concentrate on Western European countries, such as Belgium, Germany, France, and Luxembourg. These countries have experienced significant impacts from the pandemic and exhibit similarities in their pandemic control strategies, making them ideal representatives. By focusing on these countries, we can gain insight into the diverse uses of social media and its effects on individuals' attitudes and behaviors towards the pandemic. As all tweets are publicly accessible via Twitter APIs and only tweet IDs are published, our collection is exempt from IRB review and adheres to Twitter's terms of service.

Researchers can employ our dataset to investigate various topics, including message diffusion, the role of OSNs in shaping attitudes, and the impact of social media-based interventions. Ultimately, our dataset serves as a valuable resource for researchers aiming to understand social media usage during the COVID-19 pandemic and inform more effective responses to future public health crises.

3.2 Related Work

Learning attitudes towards COVID-19 vaccination. A considerable amount of literature has studied vaccination attitudes utilising social media as data source. It has been well established that users' narratives on social media disclose their individual perceptions of vaccination [JVR+20, GCH19a]. Compared to self-reporting questionnaires, social media can provide large-scale data for fine-grained analysis [SA21, LL21a], and capture uninterrupted perspectives due to the passive collection of social media data [HMT+17].

We identify two types of methods in the literature extracting vaccination attitudes from social media data. The methods of the first category determine users' attitudes according to communities they belong to. Pro- and anti-vaccine users are found to spontaneously form communities [CMK+20]. A community can be a chat group users join themselves, or formed by users' interactions, e.g., mutual friendship. Shahsavari et al. [SHW+20] used community detection to identify anti-vaccine communities. Johnson et al. [JVR+20] retrieved users' vaccination attitudes from the stances of Facebook fan pages and investigated the characteristics of the vaccination-related communities formed. The second type of methods extract attitudes from user-generated content, mainly with machine learning and deep learning models. Melton et al. [MOASN21] and Xue et al. [XCH+20] obtained the topics of a set of tweets and used the sentiment of each topic as the vaccination attitudes of the tweets in that topic. Hatmal et al. [HAHO+21] calculated users' attitudes towards the vaccine side effects with traditional machine learning methods including random forest and XGBoost. Hussain et al. [HTH+21] proposed a model combining traditional dictionary based methods with the deep learning model BERT to predict public attitudes. By comparing with existing analysis based on surveys, the authors showed their model achieved similar results. Such consistency is also confirmed by Yousefinaghani et al. [YDM+21].

The second type of methods allow us to conduct temporal analysis and monitor the changes of attitudes while the community-based methods will fail due to the relative stable community structures. To enable these methods relying on machine learning or deep learning, we conduct the first attempt to release a large number of tweets annotated with vaccination attitudes.

Datasets related to COVID-19 vaccines. Although social media posts have been used to study vaccination attitudes since the outbreak of the pandemic [LL21b], only a few datasets are publicly available. Pierri et al. [FSMS21] published a dataset collected from Twitter and Facebook recording Italian users' discussions about vaccination. DeVerna et al. [DPT+21] released the CoVaxxy dataset composed of English-language tweets about the COVID-19 vaccination generated from the US. Chen et al. [MXKP21] published the MMCoVaR dataset which contains only 24,184 tweets related to COVID-19 vaccines, spanning less than one month. Table 3.1 summarises the basic information of these three datasets.

Our dataset differs from the above datasets from three aspects. First, the released tweets span over a period of about 14 months, which covers a sufficient amount of time before and after the administration of the first COVID-19 vaccine. This enables temporal studies on vaccination attitude changes along with time. Second, the collected tweets are from four Western European countries which can well portray the first group of COVID-19 vaccine receivers and other European countries hit badly by the pandemic. Last not least, we have annotated a large number of tweets with the vaccination attitudes. These labelled data make it possible to utilise the established NLP methods based on deep learning to learn vaccination attitudes from tweets, and facilitate the development of new methods.

3.3 Dataset Collection Method

To collect high-quality data for virtual ethnographic research, we design a data collection method that ensures three key attributes: locality, topicality, and connectivity. First, locality pertains to the geographic location of users from whom we collect data. To study social media behaviour in a specific region, it is crucial to obtain data from users in that region. Thus, our method focuses on identifying and capturing data from users in the selected study area, ensuring geographical relevance. Second, topicality relates to the specific topics of interest in the research. To capture data from users active on a particular topic, our method emphasises identifying and collecting data from users discussing the topic of interest, ensuring a dataset rich in content relevant to the research question. Third, connectivity refers to the social connections between users on the platform. Understanding the structure and dynamics of these connections is critical for studying information spread and the influence of users' surroundings on their perspectives. Therefore, our method captures data on users' connections, including followers, followees, and interactions such as retweets, quotes, and mentions. This comprehensive method ensures a thorough dataset for researchers to analyse and draw conclusions from.

Our data collection method consists of two steps. In the first step, we compile a collection of active users from the target area. A user is considered "active" if they regularly participate in discussions and interact with others on a certain topic, such as COVID-19 in this thesis. We focus on these users because active users are more likely to express their true opinions through tweets. In the second step, we collect the tweets of the identified active users. This ensures that the dataset is enriched with relevant content and reflects the opinions and discussions of users engaged with the subject matter.

Step 1. Active user identification. In accordance with Twitter policy, the only means of obtaining Twitter users is through the meta-information associated with tweets. Instead of searching tweets directly, we refer to a publicly available tweet dataset [CLF20], which is continuously updated and contains tweet IDs related to COVID-19 from users worldwide. We download tweet IDs from the dataset that fall within the specified period and fetch the corresponding tweets using the Twitter API. In total, we download 51,966,639 tweets, from which we identify 15,551,266 Twitter users. To retain active users, ensuring locality, topicality, and connectivity, we perform two consecutive filtering operations: *location-based filtering* and *activity-based filtering*.

Location-based filtering. We use geographic information in the tweet metadata to determine users' areas of origin. If a user has multiple tweets from different areas, a rare occurrence among our collected users, we use the areas reported in the earliest tweet. The tweet metadata has two fields for storing the originator's location: *Geo* and *Place*. The *Geo* field records the user's device-generated location, while the *Place* field stores user-provided geographic information. *Geo* information is accurate and follows a unified format that can be automatically parsed. However, only about 2% of tweets include *Geo* values. Consequently, we use the *Place* field, which is typically ambiguous when the *Geo* field is unavailable.

To standardise the location format and remove ambiguity, we leverage ArcGis Geocoding, widely used in previous research for the same purpose [HHSC11]. For instance, the user-input *Place* field value *Moselle* is converted into a machine-readable location, including city, state, and country: *Moselle, Lorraine, France*. Over 70% of the selected users have at least one geographic field filled. We remove user IDs not located in the target area, ultimately retaining the seed users from the areas we want.

Activity-based filtering. We propose two methods for filtering active users based on two different graphs: *user interaction graph* and *social network graph*.

The first method is based on the *user interaction graph*, which filters out inactive users from the seed users. The user interaction graph is undirected and weighted. When user u retweets or mentions a tweet generated by user u' or vice versa, an edge is created between the two users (e.g., u and u') [EJR+10]. The weight of an edge represents the number of retweets or mentions between two users. Using our dataset in this thesis as an example, we first eliminated edges with weights below a certain value to exclude occasional interactions. Then, we remove vertices with a degree lower than a certain value to ensure active users frequently interact with multiple users.

The other method is based on the user's *social network graph*. We employ an iterative approach to collect social networks and users. For each seed user, we obtain their followers and retain only those users who have a mutual following relationship with the seed user, as these users are more likely to reside in the target area. Then, we download the new users' locations from their profiles and add only those users from the target area to the social graph. If a user in the graph has a following relationship with the newly added user, we also add edges accordingly. After the first round, we continue to check the newly added users by adding their friends who follow each other and do not already exist in the current social network graph. This method continues until no new users can be added. In the dataset used for this thesis, the filtering method reaches termination after five iterations. We take the largest weakly connected part of the social graph. After this step, we obtain a large social network graph, and since most users in the graph are relatively inactive, we construct a subgraph by removing all users who post or retweet fewer than a certain number of tweets. Note that we retain some of these inactive users when the remaining graph is no longer connected after removing them. Finally, we obtain a social graph with active users.

Both of these filtering methods have their own advantages and disadvantages. The user interaction graph method emphasises users who actively participate in discussions and interactions related to the topic of interest. This method is simple, has a low time cost of collection, ensures that selected users contribute significantly to ongoing conversations, and produces a larger social graph containing more user and social information. However, it is important to note that in Twitter data, even if user u retweets content from another user u', it does not necessarily mean that u follows u'. As a result, this approach might not fully capture users' social networks since it only considers direct interactions between users.

On the other hand, the social network graph approach captures the entirety of users' social networks, allowing researchers to analyse the influence of users' environment on their perspectives. This method provides a more comprehensive view of the connections between users and represents their true social relationships, offering a more realistic depiction of user interactions. However, social network-based approaches are complex, time-consuming, and typically rely on a small number of seed users, yielding a smaller network that may impact the analysis of user behaviour and attitudes.

In conclusion, the choice of the most suitable filtering method depends on the specific goals and objectives of the study. Researchers should carefully consider their research goals and the desired level of network authenticity to determine the most appropriate method for addressing their particular research questions.

Table 3.2: Tweet examples.

Label	Example (Translated to English)
Positive	We have a new weapon against the virus: the vaccine. Hold together, again.
Negative	My daughter, a nurse at the AP-HP, on the vaccine "Ah ah ah! They don't even dream about it, they start with the old ones so that we can attribute the side effects to age".
Positive but dissatisfaction	It's bad enough for individuals to refuse #COVID19 #vaccines for themselves. But forcing a mass vax site to shutdown, knowing it means vaccines may go to waste, is criminal. Call it pandemicide.
Neutral	Have any diabetics been vaccinated? I need some information
Off-topic	a 10% discount on pet vaccinations next week.

Step 2. Timeline tweet streaming. In this step, we gather tweets that are originally posted or retweeted by the users identified in the previous step during the study period. These tweets will be used to analyse user behaviour and interactions throughout the study period. We utilise the Twitter Academic Research API to search for tweets based on the IDs of active users. The API permits up to 500 tweets per download request. To ensure comprehensive coverage, we construct a request for each user on a monthly basis. This approach allows us to achieve an acceptable coverage rate since it is less probable that a user will post more than 500 tweets within a single month. It is worth mentioning that if researchers wish to focus on the content of tweets related to a specific topic, they can incorporate keywords in the search that correspond to the term of interest. By adding these keywords, the search results will be refined, enabling the collection of tweets that are more closely related to the selected topic.

3.4 Collected Datasets

In this thesis, based on the two-step data collection method outlined earlier, we collecte two datasets related to COVID-19, named *EU-Vax* and *GR-ego*, respectively. Each of these datasets employs one of the filtering methods discussed in the activity-based filtering section of Step 1. Specifically, the *EU-Vax* dataset uses a filtering method based on the user interaction graph, while the *GR-ego* dataset uses a filtering method based on the social network graph. As mentioned earlier, these two datasets have different focuses. The *EU-Vax* dataset, collected using the interaction graph-based filtering method, is more effective in gathering users who are active on the topic and can include a larger number of users. Conversely, the social network-based filtering approach employed in the *GR-ego* dataset may not accumulate as extensive a network, but it can more accurately represent the real-life social networks of users. In this thesis, we utilise different research networks depending on the specific research objectives.

In this section, we delve into the details of these two datasets collected based on the collection method in the previous section.

3.4.1 The EU-Vax Dataset

The EU-Vax dataset is a COVID-19 vaccine-related Twitter dataset that focuses on users from four European countries: Luxembourg, Belgium, Germany, and France.

These countries are selected as representatives because of their similar pandemic control strategies and close economic and political interactions. The EU-Vax dataset utilises a user interaction graph-based filtering method, which offers a more comprehensive view of discussions and interactions across the four countries. Our data collection spans two years, from October 2019 to the end of 2021, encompassing the three months before the COVID-19 pandemic outbreak. The dataset covers the period from the early stages of the pandemic to the commencement of vaccination campaigns, ranging from 22 January 2020 to 15 March 2021.

To compile the EU-Vax dataset, an initial set of seed users from the four target countries is identified using location-based filtering. User IDs not located in these countries are removed, retaining a total of 767,583 users. Subsequently, the user interaction graph-based filtering method from Step 1 is applied. Edges with weights less than 2 are eliminated to exclude occasional interactions, and vertices with a degree less than 2 are removed to ensure that active users engage in frequent interactions with multiple users. This method results in the retention of 54,381 active users.

With this set of active users in the target countries, Step 2 is followed to collect tweets related to COVID-19 vaccines or vaccinations initiated or retweeted by these users. As in previous studies, COVID-19 vaccine-related keywords are used to filter tweets. Same as previous studies [YDM+21, DPT+21], we use keywords related to COVID-19 vaccines to filter tweets. Two methods have been adopted to select vaccine-related keywords in the literature [DPT+21]. One is called *snowball sampling* which iteratively enriches the initial set of keywords according to the newly downloaded messages. The other method directly constructs the set of keywords based on expert knowledge and contexts. As many keyword lists are publicly available and produce rather good coverage [HTH+21, SZS+18], we decided to refer to them and only selected the ones with the best coverage to keep the list short. As the tweets originated from our targeted countries are written in multiple languages, which are different from those studied in the previous works, we translated the selected keywords when necessary. After multiple rounds of manual validation, we used all the keywords containing the following strings: *'vax'*, *'vaccin'*, *'covidvic'*, *'impfstoff'*, *'vacin'*, *'vacuna'*, *'impfung'*. *'sputnikv'*, *'astrazeneca'*, *'sinovac'*, *'pfizer'*, *'moderna'*, *'janssen'*, *'johnson'* and *'biontech'*. In this step, we download a total of 2,198,090 tweets, and the IDs are made publicly available.

3.4.2 The GR-ego Dataset

The GR-ego dataset focuses on the Greater Region of Luxembourg (GR), a cross-border region centred around Luxembourg and encompassing adjacent regions of Belgium, Germany, and France. This region is selected primarily due to its intense connections of international residents from diverse cultures, making it unique as a global financial centre. Additionally, these countries well represent the first batch of countries administering COVID-19 vaccines. This dataset is collected using the social network graph-based filtering method, as mentioned earlier. The primary aim of the GR-ego dataset is to provide a more accurate representation of users' real social networks within this specific geographic area, allowing for a deeper understanding of the interactions and connections between users in this region.

To compile the GR-ego dataset, an initial set of seed users from the GR is identified. These seed users are determined based on location-based filtering, retaining tweets originating from the Greater Region. In total, 128,310 tweets from 8,872 users in the region are obtained.

Next, we apply the social network graph-based approach, as mentioned in the activity-based filtering section. In our dataset, the filtering method reaches termination after five iterations. We then select the largest weakly connected part of the social graph. After this step, a total of 12,256,152 users and 21,203,130 subsequent relationships are obtained. Since a majority of the users in the graph are relatively inactive, we construct a subgraph by removing all users who posted or retweeted fewer than 3 tweets. It is important to note that we retain some of these inactive users because removing them would have disconnected the remaining graph. Ultimately, we obtain a social graph with 14,756 users and 148,647 edges.

Finally, we collect 37,281,824 tweets spanning from October 22nd, 2019 to December 31st, 2021, according to the method outlined in Step 2. We have released the IDs of our collected tweets via GitHub[1].

3.4.3 Data Annotation

In this thesis, one of our goals is to comprehend public opinions and reactions to COVID-19 vaccination policies. In order to facilitate the application of sentiment analysis methods, such as machine learning and deep learning-based natural language processing (NLP), we have created a manually annotated dataset of tweets. In this section, we detail how we selected tweets to annotate, determined the annotation labels, and conducted the manual annotation.

Tweet selection. Since the number of tweets in our EU-Vax Dataset exceeds our capacity to annotate, we select a number of tweets that can well represent the linguistic features of COVID-10 vaccination related tweets. Specifically, we first sort the downloaded tweets in the descending order by their numbers of times being retweeted. We then remove the most frequently retweeted tweet from the ordered list and added it to the list of tweets to annotate iteratively until every active user has at least one posted or retweeted message to annotate. In total, we select 17,934 tweets.

Annotation labels. We annotate each tweet with one of the following labels:

- **Positive (PO)**: The originator expresses his/her support for the vaccines and vaccination in the sense that vaccine or vaccination can effectively protect the public, and will be or has been vaccinated.

[1] `https://github.com/NinghanC/SWB4Twitter`

- **Negative (NG)**: The originator expresses doubts or disbelief about the effectiveness of the vaccines or vaccination in combating the pandemic, or hesitates or refuses to be vaccinated.
- **Neutral (NE)**: No explicit attitude or and intention are expressed.
- **Positive but dissatisfaction (PD)**: The originator expresses dissatisfaction or complaints about the current policies or measures against COVID-19, but still holds a positive attitude towards vaccination.
- **Off-topic (OT)**: The content is related to neither COVID-19 vaccines.

We give some examples for each label in Table 3.2. Note that in addition to the labels for affective attitudes, i.e., positive, negative and neutral, we add a new label PD. We notice that there exists a large portion of tweets expressing the originators' negative feelings or disagreement about the way the governments handle the pandemic such as complaints about lock-downs. However, the originators still show their belief in vaccines as an effective and ultimate measure to beat the virus. Such tweets use terms which are negative inherently and if not explicitly separated, they will confuse NLP methods with the ones that should be labelled as NG. The example tweet of PD in Table 3.2 contains negative terms such as "bad" and "criminal" but still clearly conveys the support of the user for COVID-19 vaccination.

Annotator training and consolidation. We hire 10 bachelor students to manually annotate the sampled tweets. All annotators are proficient in at least two of the four official languages of the targeted countries, and, in the meantime, are active on Twitter. The author of this thesis acts as the coordinator in charge of annotator training and annotation consolidation. Each annotator receives a tutorial from the coordinator explaining the semantics of all labels with examples. We also distribute a guideline illustrating the workflow on the Doccano platform[2] we build to collect annotators' input. To ensure that all annotators hold the correct understanding, we conduct a pilot annotation process in which all annotators are first asked to annotate 100 tweets. The coordinator verifies their annotations and provides additional explanations if necessary. We repeat the process with another 100 tweets. After two rounds of training, annotators succeed in understanding the labels and also become familiar with the Doccano platform.

We first select one annotator to annotate all the tweets and this full annotation takes approximately 60 hours. We then randomly assign to each of the remaining 9 annotators around 2,000 tweets and ask them to validate the labels. Meanwhile, the coordinator goes through all annotated tweets. When disagreeing with the labels given by the first annotator, they add new labels. This validation takes the coordinator about 60 hours and each of the other annotators 4 hours. In this way, our annotation strategy ensures each tweet is labelled three times. To solve the conflicts, the coordinator consolidates all annotations. The label agreed by at least two annotators is set as the final annotation. For those with three different labels, the coordinator communicates with the other two annotators and picks the most appropriate one.

[2] github.com/doccano/doccano

Table 3.3: Inter-annotator agreement (PO: Positive, NG: Negative, NE: Neutral, PD: Positive but dissatisfaction, OT: Off-topic).

Label	AOA	Feliss' kappa	Krippen-dorf's Alpha
PO	0.72	0.73	0.73
NG	0.82	0.88	0.88
NE	0.74	0.78	0.77
PD	0.61	0.63	0.62
OT	0.83	0.87	0.86

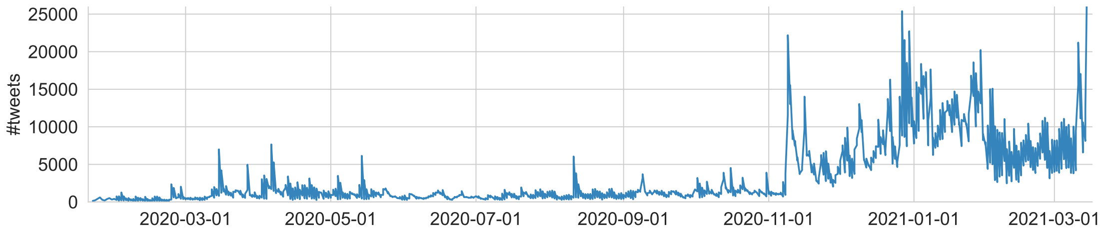


Figure 3.1: The temporal distribution of tweets on daily basis.

Annotator agreement. To ensure the quality of our annotations, we leverage three widely accepted metrics to quantitatively evaluate the inter-annotator reliability for each label: Average Observed Agreement (AOA) [FLP13], Fleiss' kappa [FLP13], and Krippendorff's Alpha [Kri70]. AOA is the average observed agreement between any pair of annotators. The term "observed agreement" in AOA refers to the proportion of labels two annotators agree with. Both Fleiss' kappa and Krippendorff's Alpha are applicable to measure the agreement between a fixed number of annotators, with the difference that Krippendorff's Alpha can handle missing labels. The values of all the three measurements range from 0 to 1, where 0 indicates complete disagreement and 1 indicates absolute agreement. For Fleiss' Kappa, 0.41-0.60, 0.61-0.80, and 0.81-1.0 are considered as moderate agreement, substantial agreement, and excellent agreement, respectively [FLP13]. Krippendorff's Alpha is more rigorous than normal standards [Kri70]. Values between 0.667 and 0.800 are deemed acceptable, while values greater than or equal to 0.8 are considered highly reliable [Kri70].

Table 3.3 summarises the inter-annotator agreement for each annotation label. We can see that for all labels, AOA scores range from 0.61 to 0.83. This implies that most of the annotations have at least two annotators in agreement. The values of the other two measurements are close. The annotators achieve the highest rank of agreement according to Fleiss' kappa and Krippendorff's Alpha for both the labels NE and OT and the second-highest rank on labels PO and NE. The annotators' agreement on PD falls drastically compared to other labels, but still remains moderate according to the ranking criteria of the Fleiss' Kappa measurement. This can be explained by our difficulties during annotation in dealing with the special linguistic features of PD tweets, i.e., frequently used negative terms or sarcastic expression. A closer look will lead to another observation that the extent of agreement on the label PO is slightly lower. A careful manual investigation reveals that a large proportion of disagreed annotations also attribute to the sarcasm and irony made to express their opinions about anti-vaccination. This identified challenge to handle sarcasm is consistent with the previous finding that people frequently are confused by sarcasm, which makes comprehension difficult [RRV13]. We take the following tweet as an example: *I am very disappointed! 16 days after my first injection of the vaccine against #covid19 I still don't get the 5g.* This tweet uses ironic expression joking about anti-vaccination comments but in fact delivers a definite supporting attitude for vaccination. Such tweets produced misunderstandings among annotators, which are solved in our consolidation phase.

Statistics. We depict the distribution of annotated tweets over the vaccination attitude labels in Figure 3.2. Adding up those labelled both PO and PD, we can see that more than 60% of the tweets express a positive attitude toward vaccination while about 20% are associated with negative attitudes.

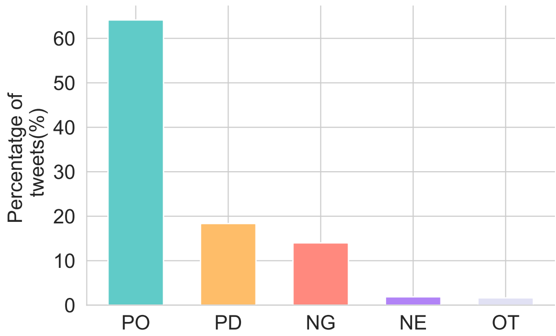


Figure 3.2: Attitude distribution of annotation tweets.

3.5 FAIR

The publication and colloction of our dataset strictly adheres to the principles of FAIR. We achieved the principle *Findable* and *accessible* by uploading our dataset through Zenodo[3], which is publicly accessible. Our data is stored in the CSV format which can be exported in any other format with little efforts. We strictly adhered to the Twitter Developer Agreement and Policies[4] in the collection and distribution of data. Our release is also compliant with the EU General Data Protection Regulation (GDPR).

3.6 Conclusion and Potential Applications

We collect two new multimodal datasets focusing on COVID-19 to facilitate research on monitoring vaccine hesitancy with the recent advances of deep learning in NLP. With four countries in Western Europe as the targeted regions, our dataset will help understand the vaccination attitudes changes during the pandemic in Europe among the first receivers of the vaccines. Moreover, researchers of NLP models can benefit from the multilingualism nature of our dataset to study new effective methods to address multilingualism in opinion or sentiment extraction.

[3]https://doi.org/10.5281/zenodo.5851407
[4]https://developer.twitter.com/en/developer-terms/agreement-and-policy

Our dataset has some limitations. First, we cannot take Twitter users' attitudes equally to public vaccination attitudes as Twitter users sometimes cannot well represent the general population. Social media can now only be used as a complementary source of information. One of our goals to release our dataset is not to replace social studies, but to demonstrate one potential application to social media in public health surveillance. Second, we only considered affective stances towards vaccination. Although our annotation can help develop new deep learning NLP models, the analysis results obtained from the trained model are limited. Third, the sampling methods used in the data collection and annotation can bring extra bias. For instance, our selection of tweets to annotate may give more significance to the users who are more influential on Twitter as their tweets are more likely to be retweeted. This can be solved by random sampling but it will significantly increase the number of tweets to annotate.

Chapter 4

Multi-faceted Exploration

In the previous chapter, we thoroughly explain the collection method of the two datasets, outlining their characteristics and presenting basic descriptive data. In this chapter, we explore the temporal and spatial distribution of the collected datasets and conduct a linguistic distribution analysis using a variety of figures. Alongside these data visualisation, we also undertake several analyses to demonstrate the potential applications of these datasets in facilitating virtual ethnography studies.

4.1 Introduction

In this chapter, we explore the two datasets - GR-ego and EU-Vax - separately, aiming to gain insights into public reactions and attitudes related to the COVID-19 pandemic and vaccination efforts. We conduct data-driven exploratory studies using machine learning and representation learning methods, enabling us to observe differences in user behaviour across the GR and related countries.

For the GR-ego dataset, we investigate how Twitter users in the GR and related countries react differently over time to the evolving COVID-19 situation. We find that there is a correlation between tweet volume and the number of COVID-19 cases in these areas, although this correlation only exists during specific periods of the pandemic. Additionally, we examine the changes in topics discussed in each country and region from January 22, 2020, to June 5, 2020, highlighting the main differences between the GR and its surrounding countries.

In the case of the EU-Vax dataset, we focus on the issue of vaccine hesitancy as a major factor contributing to the stagnant uptake ratio of COVID-19 vaccines in Europe and the US. We recognise the importance of quickly and accurately understanding public attitudes towards vaccination, and acknowledge the effectiveness of social media platforms in capturing public opinion. This dataset comprises 2,198,090 tweet IDs collected from Western Europe, with 17,934 of them annotated with the originators' vaccination stances. Our annotation facilitates the use and development of data-driven models to extract vaccination attitudes from social media posts, further confirming the power of social media in public health surveillance.

To lay the groundwork for future virtual ethnography-related research, we demonstrate the potential application of our data in practice, we track temporal changes in public attitudes towards both COVID-19 itself and COVID-19 vaccination. This

Figure 4.1: Word cloud of the tweets posted between 11/9/2020 and 11/11/2020.

exploration of the datasets not only serves as a foundation for further research but also highlights the value of these datasets in understanding and analysing public sentiment and behaviour in the context of pandemic and vaccination efforts.

4.2 Exploration of the EU-Vax Dataset

In this section, we focus on the EU-Vax dataset, which is collected to study public opinion and reactions to the COVID-19 vaccination policy in Western Europe. We perform a comprehensive statistical analysis and visualisation of this dataset to better understand its characteristics and trends. Moreover, we evaluate and compare the performance of established NLP benchmarks to assess the dataset's applicability for sentiment analysis tasks. To demonstrate the potential use of these data in practice, we apply the dataset to track temporal changes in vaccination attitudes and to analyse sudden surges that warrant further attention.

4.2.1 Dataset Statistics

We illustrate through diagrams the temporal and spacial distributions of the EU-Vax Dataset as well as the language distribution. Along with the diagrams, we conduct some analysis to showcase the potential use of our tweet dataset.

Temporal distribution. In Figure 3.1, we display the number of collected tweets on daily basis. Sudden surges of daily posts may indicate the switch of public attention and should be investigated to identify possible events that cause the changes. We can see that COVID-19 vaccines or vaccination were rarely discussed before November 9, 2020 when a sudden surge occurred. Since then, the discussion around them remains popular. To understand the events that promoted the increase, we extracted the tweets that were generated in the 3 days after November 9, 2020 and created a word cloud to identify the frequently used words (see Figure 4.1). With the highlighted words, we infer that the event may be related to the efficacy of the Prizer/Biotech vaccine. With a check on the news, we confirmed our inference. Pfizer announced the 90% efficacy of its vaccine co-developed with Biotech.[1] With the similar approach, we checked the peak on December 27, 2021 and the sudden increase on March 15, 2021. We discovered that the tweet increase on 12/27/2021 attributed

[1]https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against

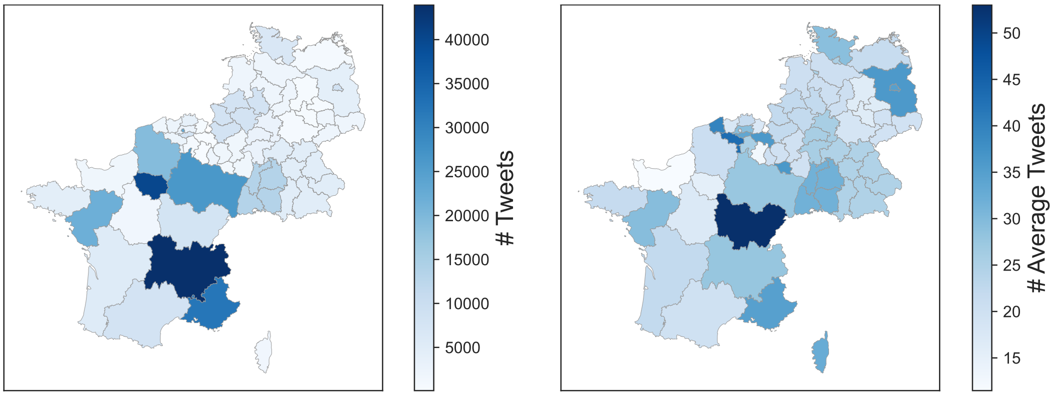


Figure 4.2: The geographic distribution of the total numbers of tweets (left) and average numbers of tweets per user (right).

to the start of the EU Mass Vaccination Campaign while the other increase may be caused by the news that France and Germany joined other European countries to temporarily halt the use of the Oxford-AstraZeneca vaccine.

Spacial distribution. In Figure 4.2, we depict the total number of tweets posted by users in all the states of the targeted countries (on the left) and the corresponding average number of tweets posted by a user (on the right). We use the latitudes and longitudes provided by the geolocation API to match users to the corresponding states. If the returned geographical information can only map to the level of countries, we exclude them when drawing Figure 4.2. The populated regions generate more tweets than those with less population. This is especially obvious in France. The states where big cities such as Paris and Lyon reside have the most tweets. However, the users in such populated regions are not always be more active on average in participating in the vaccine-related discourses. Moreover, the states adjacent to other countries tend to be more active. For instance, in France, the users in the states adjacent to Switzerland on average posted more than 40 tweets while users in the state with Paris just posted about 20.

Language distribution. As we mentioned previously, a character of our tweet dataset is the multilingualism that is inherent in Europe. In Figure 4.3, we show the percentages of tweets in the top 5 most used languages. As the official language of France, Belgium and Luxembourg, French is the dominant language which is used in more than 60% of the collected tweets. The multilingualism is considered as a challenge in NLP to extract subjective opinions from texts. Researchers will benefit from our tweet dataset and the annotated tweets in developing and validating new NLP methods to address this challenge.

4.2.2 Experimental Evaluation and Validation

We aim to achieve two goals through experiments in this section. First, by training and testing adapted NLP methods for sentiment or opinion classification, we examine the validity and quality of our annotation of vaccination stances. Second, to illustrate the usability of our datasets, we conduct some preliminary temporal studies tracking vaccine hesitancy and identifying changes requiring further analysis.

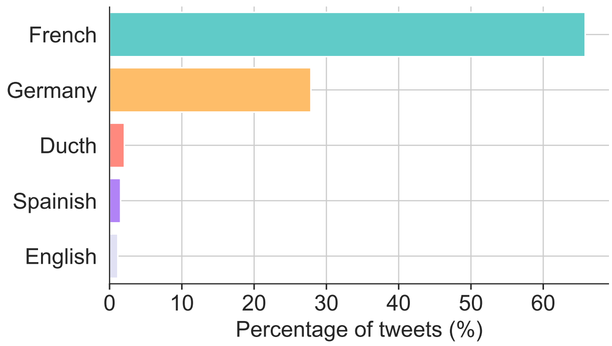


Figure 4.3: Distribution of tweet over languages.

Vaccination attitude calculation with NLP. The application of deep learning and machine learning has revolutionised NLP, especially in extracting opinions or sentiments from textual contents. Compared to machine learning models which rely on manually constructed features, deep learning models can learn effective features automatically with little manual intervention. Extensive empirical evidence has proved the overwhelming performance of deep learning models in NLP studies [CJS+20].

We trained and ran several well established NLP models based on machine learning and deep learning with our annotated tweets. A good performance of the trained models in classifying tweets will attest the utility and trustworthiness of our annotation.

Experiment setup. We select Random Forest (RF) and Support Vector Machines (SVM) as the representative machine learning models due to their wide use. Regarding deep learning models, we use BERT [DCLT19], RoBERTa [LOG+19], and DistilBERT [SDCW19]. Such deep learning models are pretrained and produce a low-dimensional representation for any given piece of text, which can be used as input for downstream classification methods.

We preprocess the tweets by removing mentions of other users with '@', quoted hyperlinks and 'RT' which stands for "retweet". We remove tweets with the label 'off-topic' due to their small proportion. To test the models' capability in dealing with multilingualism, we construct three datasets: the original annotated tweets, the French-language annotated tweets and the German-language annotated tweets. We divide each of these three datasets into training, testing and validation set with the ratio 80%, 10% and 10%, respectively. To train the selected machine learning models, we use TfidfVectorizer [PVG+11] to convert the preprocessed tweets into the bag of n-gram vectors. We use grid search as the optimiser for SVM and RF. For deep learning methods, we adopt their publicly available implementation for text embedding and keep their default settings. The text embeddings are then sent to a fully-connected ReLU layer with dropout. A linear layer is added on the top of the final outputs for regression with softmax as the activation function. We use CrossEntropyLoss as the loss function and Adam as the optimiser [KB15]. All models are trained for 30 epochs for optimisation with the learning rate of 0.00001, and batch size of 32. We set the maximum sequence length as 128, which defines the maximum numbers of tokens in a tweet that can be processed.

Result analysis. Table 4.1 shows the classification performance evaluated with conventional measurements. First, we can see that the deep learning methods outperform the machine learning methods in classifying both multilingual tweets and those in single languages, and their performances are close. This confirms the findings in [HRS+20]. Second, the results show that multilingualism affects the classification performance of deep learning models, although we use a pre-trained model specifically for classifying tweets in multiple languages.

By comparing with the models' performance on other classifying tasks in the literature [CJS+20, HRS+20], we observe that the models can achieve the same-level performances. This implies that our annotation is trustworthy and useful for future research on vaccination attitude learning. The results we showed in Table 4.1 can thus be referred to as benchmarks for comparison.

Table 4.1: Classification results for different benchmarks.

	Mulitilingual				French				German			
Model	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
RF	0.4317	0.3219	0.4471	0.4749	0.5676	0.4829	0.4754	0.5510	0.51481	0.4978	0.4611	0.4503
SVM	0.4001	0.3816	0.4380	0.4263	0.5004	0.4256	0.4141	0.4998	0.4954	0.4037	0.4093	0.4719
mBERT	0.6622	0.5769	0.6132	0.6466	0.7016	0.6933	0.7004	0.7184	0.6999	0.6875	0.6919	0.7038
XLM-RoBERTa	0.6801	0.5848	0.6271	**0.6618**	0.7023	0.7018	0.7145	**0.7086**	0.7102	0.6971	0.7081	**0.7079**
Distil-mBERT	0.6768	0.5834	0.6287	0.6601	0.6978	0.6916	0.7084	0.7065	0.7094	0.7004	0.7068	0.7071
CamemBERT	-	-	-	-	0.7147	0.7136	0.7222	**0.7120**	-	-	-	-
GottBERT	-	-	-	-	-	-	-	-	0.7165	0.7046	0.7199	**0.7136**

Temporal analysis of vaccination attitudes. We present a description of an application of our tweet dataset and annotation to illustrate one potential use of our release. Specifically, we visualise the temporal evolution of the various vaccination attitudes and analyse the possible causes to some changes that require more attention.

We apply XLM-Roberta model trained in the previous experiment to calculate the vaccination attitudes of the tweets in the tweet dataset. Figure 4.4 shows the distribution of the predicted vaccination attitudes. We can see the distribution is similar to that shown in Figure 3.2 with slightly more tweets with negative attitudes.

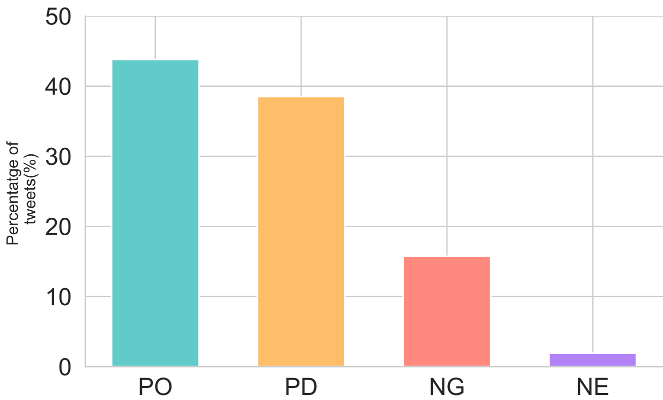


Figure 4.4: Attitude distribution of all tweets.

We draw the temporal evolution of the numbers of tweets that are classified as NE, PO and PD in Figure 4.5. Note we ignore the tweets with label OT due to their small amount. Recall that significant discussion on COVID-19 vaccines started from November 8, 2020 as shown in Figure 3.1. Therefore, we focus on the attitude changes from that date. We can see that the number of tweets containing different attitudes toward vaccination changes over time. Compared to the temporal distribution of daily tweets shown in Figure 4.5, we observe that the growth in the total number of tweets is not accompanied by a proportional change of vaccination attitudes. Specifically, the number of neutral tweets varies less from day to day and remains stable at the same level compared to tweets with other attitude labels. Based on previous research reporting that the content of tweets is highly correlated with real-world situations [PCHG20], we make a hypothesis that real-world events may contribute to the fluctuating numbers of tweets with different attitudes.

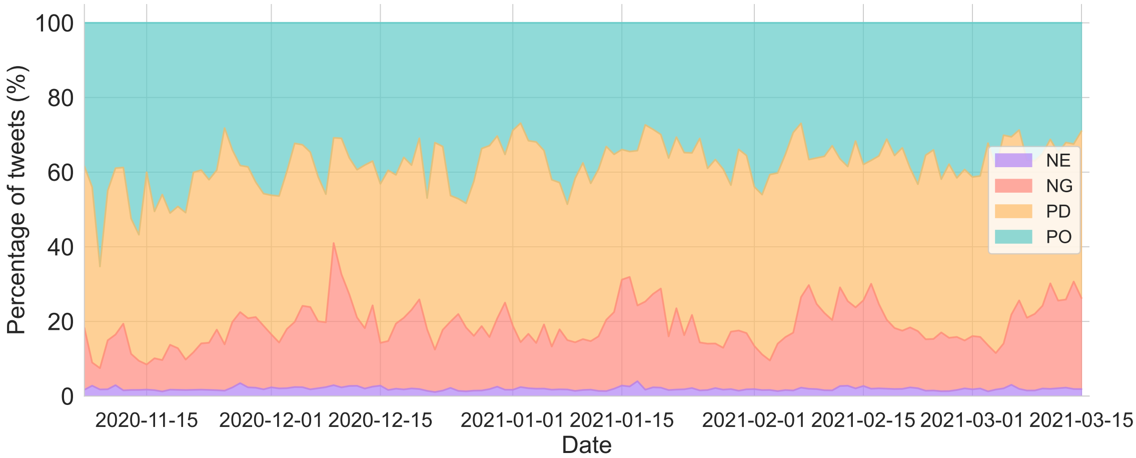


Figure 4.5: Temporal distribution of tweets with different vaccination attitude labels.

In monitoring vaccine hesitancy, more attention should be paid to the fluctuations of the negative attitudes. In the following, we take four peaks of the curve of negative tweets as examples and discuss the potential causes. We first plot word clouds to identify the most frequently used keywords and then search these keywords on the Internet to understand the events that may contribute to the changes. The first peak happened around December 9, 2020 due to negative news and misinformation about the vaccine efficacy. For instance, on December 9, 2020, two healthcare workers were reported to have experienced symptoms after receiving their first shots, and on December 11, Sanofi and GSK delayed COVID-19 vaccine. A wide spread piece of misinformation during this period is also identified, saying that 6 people had died after vaccinated and another 4 developed Bells Palsy. Another two peaks occurred around January 15, 2021 and February 10, 2021, respectively. We notice that both of these two peaks attribute to the propagation of a large volume of misinformation. Take two pieces of misinformation identified for the peak in January, 2021 as examples. One said that on January 14, the Norwegian Medicines Agency reported that a total of 29 people had suffered side effects, 13 of which were fatal. The other was about the death of an Indian healthcare worker after receiving Covid-19 vaccines. The last peak in March, 2021 is due to the negative news that multiple countries decided to suspend the use of the AstraZeneca vaccine due to the reported negative effects.

From the above discussion, we can see that with NLP deep learning models, public vaccination attitudes can be extracted. When social network data are available, we can track almost in real time the changes of vaccination attitudes and understand the potential causes. This may finally help the governments make corresponding effective intervention methods proactively.

4.3 Exploration of the GR-ego Dataset

In this section, we focus on the GR-ego dataset, which was collected to study public opinion and reactions to the COVID-19 pandemic in the Greater Region of Luxembourg and related countries. Our exploration aims to address two key aspects of this dataset: the potential correlation between the volume of tweets and daily cases of COVID-19, and the evolution of topic categories over time in each country and region.

First, we investigate whether there is a strong correlation between the volume of tweets and COVID-19 daily cases in the GR and related countries. If such a correlation exists, we will explore whether the volume of tweets helps to predict COVID-19 daily cases. This will involve applying various time series prediction methods to assess the predictive power of tweet volume on the number of daily cases. Second, we aim to examine changes over time in the categories of topics discussed in each country and region. In doing so, we can also compare changes in topic categories in the GR with those in other countries to determine similarities and differences in public concerns and responses throughout the pandemic.

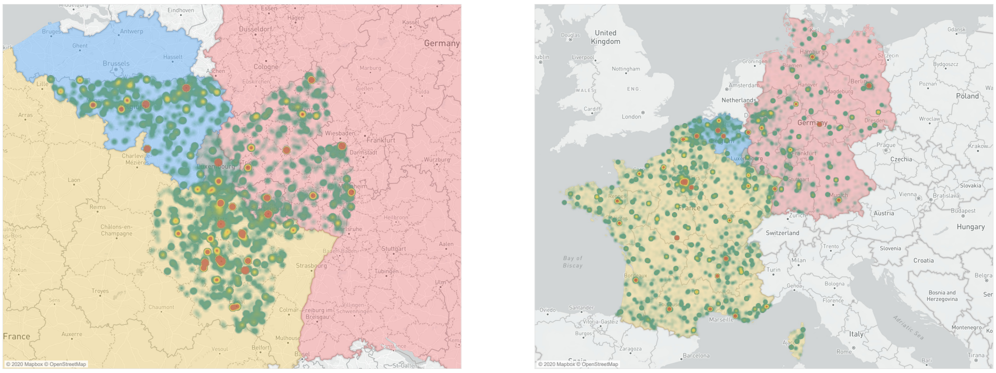
Through a comprehensive analysis and visualisation of the GR-ego dataset, as well as a comparison of topic categories and temporal changes, we provide valuable insights into public sentiment and behaviour during the COVID-19 pandemic. This analysis demonstrates the potential value of this dataset in addressing the challenges facing virtual ethnography. By understanding the intricacies of public opinion and discourse from the social media data, researchers and policy makers can have a better response to the concerns and needs of the population, ultimately improving the effectiveness of public health measures and policy implementation.

4.3.1 Data Description

In this section, we briefly describe the statistical information of the dataset and how we obtained information about the daily cases of COVID-19 in these regions and countries. Table 4.2 shows the summary of the collected tweet data of the GR, Luxembourg, France, Germany, Belgium and globally. Figure 4.6 contains two heatmaps of user locations in the GR and the related countries for a better understanding of this study.

COVID-19 data collection. The dataset published by the European Center for Disease Prevention and Control [2] allows us to obtain COVID-19 data including daily cases, deaths and locations for the country we selected. As there is no official COVID-19 data published for the GR, which is composed of Luxembourg, Wallonia in Belgium, Saarland and Rhineland-Palatinate in Germany and Lorraine in

[2] `http://www.granderegion.net/en/The-Greater-Region-at-a-Glance`



(a) The GR

(b) The related countries

Figure 4.6: User location heatmap of the GR and the related countries.

Region/Country	tweet volume	User volume
Twitter	51,966,639	15,551,266
The GR	35,329	7,894
Luxembourg	7,512	1,545
Belgium	119,467	31,446
France	1,050,312	288,009
Germany	430,688	87,796

Table 4.2: Summary of our COVID-19 Twitter dataset.

France, we add up all the data for the cities and regions mentioned above from the datasets[3] published by corresponding countries as the final the GR data when counting daily cases and deaths in the GR. It should be noted that as the number of daily new cases in France is not available at the regional level, and deaths, hospitalisations, departures data have been published only since March 18, 2020, data for Lorraine is counted as zero until March 18, 2020, and the sum of hospitalisations, hospital departures and deaths is considered as the total number of cases on that particular day.

4.3.2 Correlation between COVID-19 Daily Cases and Tweet Volume

To explore the correlation between tweet volume and COVID-19 daily cases in GR and the related countries, we introduce basic reproductive rate R_0 and effective reproductive rate $R(t)$ in epidemiology to slice the periods of the pandemic, and a spatio-temporal analysis of the correlation between tweet volume and daily cases in each period is conducted by Pearson Correlations (PC).

R(t)-based time division. R_0 is the expected number of cases arising directly from a single case in a population where all individuals are susceptible to infection [HD96] and $R(t)$ represents the average number of new infections caused by an infected

[3] `https://bit.ly/2ErDii7,https://bit.ly/3gaGGMm,https://bit.ly/33c8CM8`

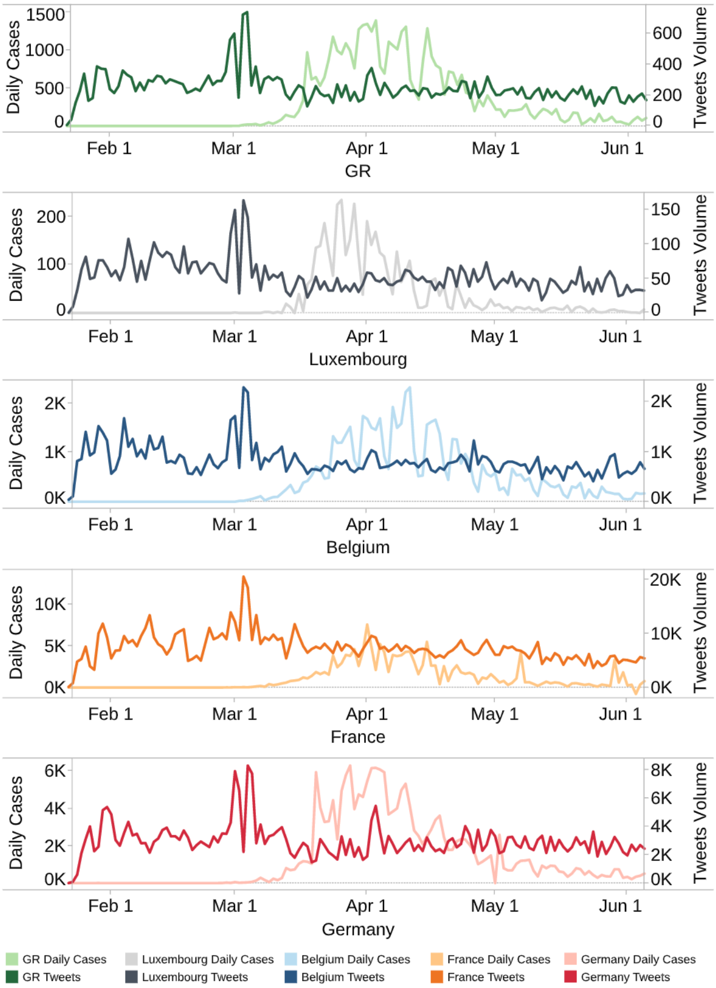


Figure 4.7: Daily tweet volume and COVID-19 new cases.

person at time t. If $R(t) > 1$, the number of cases will increase, e.g. at the beginning of an epidemic. When $R(t) = 1$, the disease is endemic, and when $R(t) < 1$, the number of cases will decrease. For the calculation of real-time $R(t)$, we use a Bayesian approach [BR08] with Gaussian noise to calculate the time-varying $R(t)$ based on daily new cases, which is also the official method for calculating $R(t)$ in Luxembourg [4]. While the study of calculating R_0 of COVID-19 have not reached a consensus conclusion [WLL20, SPXZ20, WWW$^+$20, RA20], we use R_0 estimated by WHO [5]. with $1.4 \leq R_0 \leq 2.5$, in this study. The results of time-varying $R(t)$ for the GR, Luxembourg, Belgium, France, and Germany are shown in Figure 4.9.

The relationship between R_0 and the $R(t)$ indicates the spreading ability of the virus. As the estimation of R_0 values is a range, we discuss R_0 here as a range as well. When $R(t) > max(R_0)$, it indicates that the virus is spreading at a higher rate than natural transmission, and the number of cases is about to reach a peak. When $min(R_0) \leq R_t \leq max(R_0)$, the virus spreads within the basic reproductive rate R_0 range, which implies that the effectiveness of the containment measures is not yet reflected in $R(t)$. In short, the virus is still spreading freely at its natural

[4]https://github.com/k-sys/covid-19

[5]https://www.who.int

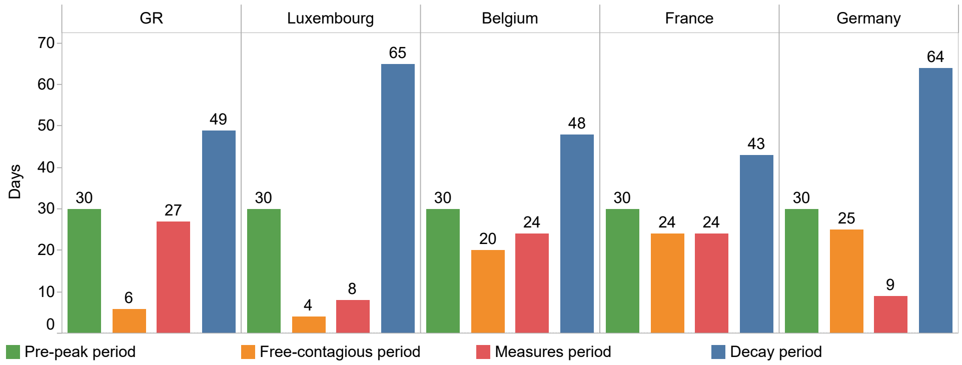


Figure 4.8: Total days for each pandemic period

transmission. When $1 \leq R_t < min(R_0)$, it means that the virus is spreading at a rate lower than R_0, the transmission is impeded, and the containment measures are in effect. When $R_t < 1$, the virus spreads slowly, and can eventually die out.

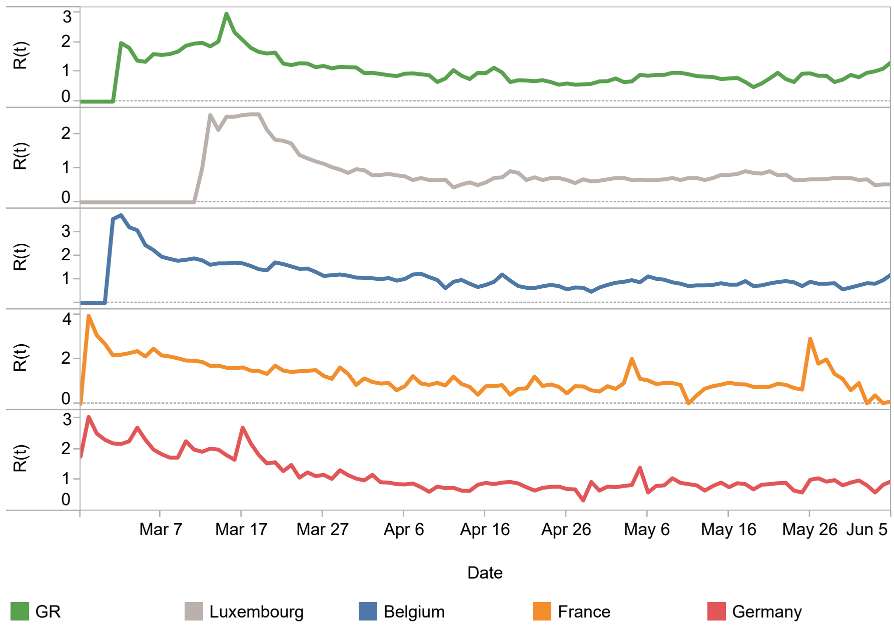


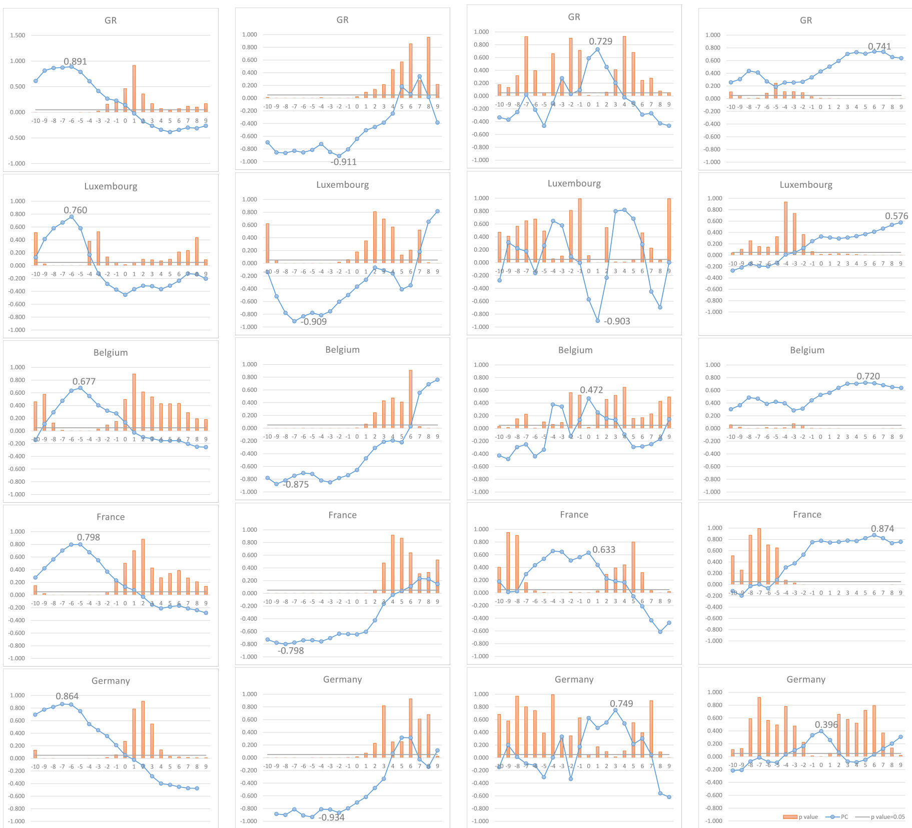
Figure 4.9: Effective reproductive rate ($R(t)$)

	Pre-peak	*Free-contagious*	*Measures* period	*Decay* period
The GR	2/14 - 3/15/2020	3/15 - 3/21/2020	3/21 - 4/17/2020	4/17 - 6/05/2020
Luxembourg	2/19 - 3/20/2020	3/20 - 3/24/2020	3/24 - 4/01/2020	4/01 - 6/05/2020
Belgium	2/04 - 3/05/2020	3/05 - 3/25/2020	3/25 - 4/18/2020	4/18 - 6/05/2020
France	2/05 - 3/06/2020	3/06 - 3/30/2020	3/30 - 4/23/2020	4/23 - 6/05/2020
Germany	1/29 - 2/28/2020	2/28 - 3/24/2020	3/24 - 4/02/2020	4/02 - 6/05/2020

Table 4.3: Time duration of the four pandemic periods for the GR, Luxembourg, Belgium, France and Germany

Here, we divide the pandemic into four periods based on the above analysis, which are: *Pre-peak* period (if $R(t)$ peaks for the first time on day t_0 and begins to decrease, with $R(t) < 2.5$ on day $t_1, (t_1 \geq t_0)$, then the pre-peak period is the 30-day period before t_1). *Free-contagious* period ($1.4 \leq R_t \leq 2.5$); *Measures* period ($1 \leq R(t) < 1.4$); *Decay* period ($R(t) < 1$). It should be noted that the second wave of the pandemic did not begin at the time when this study was conducted, so this division of intervals only applies to this time period, i.e., from 2020-01-22 to 2020-06-05. The precise time duration of these pandemic periods for each country and region is summarised in Table 4.3.

The exact numbers of days of each pandemic period are shown in Figure 4.8 for the region and countries. The *Free-contagious* period in Luxembourg and the GR is particularly shorter (4 & 6 days) compared to other countries (24-20 days). Being a relational city characterised by high mobility, it may be relatively difficult to control the pandemic. The reason why the GR and Luxembourg, has a shorter *Free-contagious* period instead, will be discussed in Section 4.3.3 in terms of the public concerns that reflected by tweet text.



(a) Pre-peak (b) Free-contagious (c) Measures (d) Decay

Figure 4.10: *PC* (Pearson's correlation) between tweet volume and COVID-19 daily cases with different lags.

Conclusion. To investigate whether there is a strong correlation between tweet volume and COVID-19 daily cases in the GR and related countries, and if so, whether tweet volume can help predict COVID-19 daily cases, we test the following hypotheses:

H1 There is a strong correlation between tweet volume and COVID-19 daily cases in the GR and related countries.

H2 Tweet volume can help predict COVID-19 daily cases.

We calculate the correlation between tweet volume and COVID-19 daily cases by *PC*, where a *PC* with a large absolute value means greater relation strength. The results are shown in Figure 4.10. A *lag* refers to the tweets occurring after the cases; a Lag = -5 days means that we match the daily cases with the tweet volume from five days earlier, in other words, a 5-days lead.

*Pre-peak period***.** As shown from Table 4.10, there is a clear trend of strong correlation ($PC > 0.8$, $p < 0.05$) with lags during the *Pre-peak* period, reaching its' maximum at -5 or -6 days, indicating that a correlation exists between tweet volume and COVID-19 daily cases and tweet volume can help predict COVID-19 daily cases in this period. This is highly consistent to the conclusions presented in the existing studies [SBB$^+$20, SZ12, JR20, YFR$^+$20].

*Free-contagious period***.** There is no clear trend of correlation with lags except the value of Luxembourg, indicating that tweet volume cannot help predict the daily cases in the *Free-contagious* period. The period only lasted for 4 days in Luxembourg, which is too small to make *PC* a reflection of the correlation. However, the *PC* values show a highly negative correlation between tweet volume and daily cases. This indicates that there is a short downward trend in the discussion of the pandemic after it reached its peak, even though the number of cases continued to rise rapidly. This result validates the conclusion of Smith et al. [SBPD16] from our dataset, who noted that public concerns of disease decline sharply after the peak even though the infection rates remain high. In other words, the public concerns of the pandemic decline after the *Pre-peak* period.

*Measures period***.** There is a clear trend of correlation with lags, tweet volume begins to level off, with a 0 or 1-day-lag moderate correlation ($0.8 > PC > 0.3$, $p < 0.05$) to the daily cases. Tweet volume cannot help predict daily cases here because it fluctuates with the number of cases on the current or previous day. It is worth noting that Pearson's coefficient is sensitive to outliers and is not robust. With too few dates included, a single outlier can change the direction of the coefficients. This period existed for only 8 days in Luxembourg, resulting in an anomaly value ($PC = -0.903$). It is assumed here that fluctuating changes in tweet volume during this period are influenced by local news and policies, and further discussion will take place in Section 4.3.3.

*Decay period***.** The correlations between tweet volume and daily cases occur in two ways. One is weakly correlated, the other reveals a correlation, but the trend of correlation with lags is insignificant. Both ways demonstrate that it is not possible to estimate daily cases with the help of tweet volume during this period.

In summary, with the Spatio-temporal analysis of the correlation between tweet volume and COVID-19 daily cases during the four periods of the pandemic, we reject the hypothesis that there is a strong correlation between tweet volume and

COVID-19 daily cases in the GR and related countries (H1) and tweet volume can help predict COVID-19 daily cases (H2). More accurately, H1 and H2 can only be confirmed during the *Pre-peak* period. In this period, regardless of the time at which $R(t)$ peaks, there is a 5-6 day lead between tweet volume and COVID-19 daily cases. Moreover, before the pandemic strikes, there is a high level of tweet volume regarding the pandemic. On the particularity of the GR, we find that the *Free-contagious* period in the GR and Luxembourg are exceedingly shorter (6 and 4 days, respectively), during the *Measures* period.

4.3.3 Topic Modelling and Classification of Tweets

In the previous section, we conduct an overarching preliminary analysis of tweet volume, but without the in-depth discussion of tweet text. In this section, we build a workflow to analyse tweet text as shown in Figure 4.11. This workflow includes tweet text pre-possessing, topical modelling, and classification of the generated topics, each part is described in details below. We perform topical modelling on the tweet text to extract the main topics discussed every day in each region and country. After extracting the tweet text topics, we generally followed the pipeline of previous studies [PMC$^+$20, KJ17]. However, the number of topics extracted in previous studies was relatively small, so the topics were classified by manually labelling. The number of topics we extract is relatively large. Hence, we take a supervised learning approach and train a classifier to distinguish these topics into 7 categories in order to observe and analyse the changes in the topics discussed in each region and country during different periods of the pandemic. In parallel, we observe and investigate the changes of topic categories, and focus on the differences that exist in the GR.

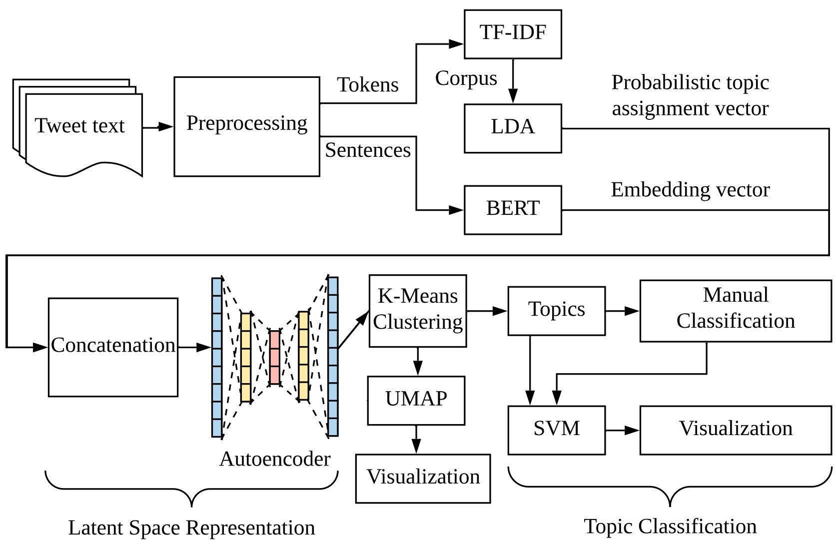


Figure 4.11: Workflow of topic modelling and classification.

Text pre-processing. Prior to topical modelling, the tweets data needs to be pre-processed. All text are lower-cased, while URLs that mention usernames and 'RT' are removed as well. Besides, punctuation and numbers are filtered out, typos are corrected by Symspel [6] and stop words are removed. Since the tweets are collected based on the keyword search, each tweet contains keywords such as "coronavirus", "koronavirus", "corona", "covid-19", and "covid". As these frequent subject-specific words are unlikely to assist for classification and topic modelling [SC03] and result in a large number of topics in the final result containing these words, rather than a more precise topic about COVID-19. In detail, if these general high-frequency keywords are not removed, these words will be ranked high in the results of topic modelling. As a consequence, this will make the final extracted topics not well represent the topics of the clusters of tweet text. Therefore, we considered these words as subject-specific stop words and remove them following Älgå et al.'s work [ÄEN20].

Topic modelling. Aiming to identify the latent topics of the tweets posted by the public in the GR and related countries, we adopt the general structure of contextual topic embedding method (CTE) [7] to extract daily topics and get a more accurate picture of topic trends. CTE mainly consists of two components, LDA and BERT, to extract different information from sentences to embedding.

LDA, a bag-of-words approach which is widely used to identify latent subject information in a large-scale archives or corpus has its drawback: it needs large corpus to train, ignores contextual information and performs mediocrely in handling short texts [YGLC13].

BERT utilises bidirectional transformers for pre-training on a large unlabelled text corpus, taking both left and right context into account simultaneously, which compensates for the shortcoming of LDA. Also, BERT is a method available for sentence embedding, thus we concatenate the generated tokens of each tweet text as input sentences for BERT to obtain sentence embedding vectors. CTE combines the sentence embedding vector generated by BERT with the probabilistic topic assignment vector generated by LDA with a hyper-parameter γ. After obtaining the concatenated vector in high-dimensional space, CTE uses an autoencoder to learn a low-dimensional latent space representation of the concatenated vector with more condensed information. Then k-means [WCR$^+$01] is implemented for clustering, and the number of clusters k, that is, the number of topics, reserved as a hyper-parameter. We extract the word frequency in each cluster, sort and then take the top ten as the representative topics of that cluster. In terms of visualisation, Uniform Manifold Approximation and Projection (UMAP) [MHM18] is used for low-dimensional latent space degradation, which is the state-of-the-art visualisation and dimension reduction algorithm.

The CTE rather than a single LDA model is chosen as our topical modelling approach due to the fact that LDA is designed for monolingual contents and lacks the structure necessary to generate effective multilingual topics [GSL$^+$16]. The GR, as a relational city, are multilingualism. CTE includes BERT, a sentence embedding model that can handle multi-language, can tackle this problem. Two adjustments are therefore made to the original CTE. For one, we adopt the BERT-based multilingual model as the pre-trained model in BERT [DCLT19] In addition, some words

[6] `https://github.com/mammothb/symspellpy`

[7] `https://github.com/Stveshawn/contextual_topic_identification`

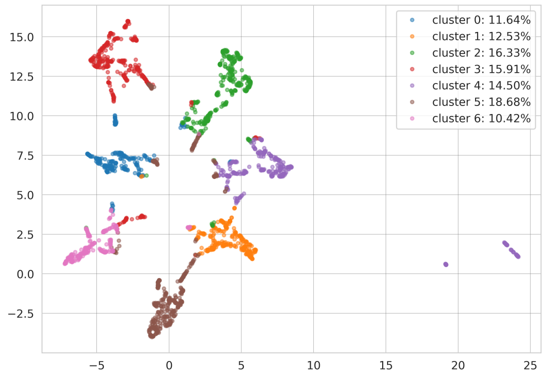


Figure 4.12: A sample of UMAP clustering results.

appear less frequent than in English which is predominantly spoken and are easily overlooked in LDA. Thus, we adopt the TF-IDF model to determine word relevance in the documents [Ram03]. We further feed the generated corpus by TF-IDF to LDA, instead of simple bag-of-words corpus.

Country	Coherence score	Silhouette score
The GR	0.432	0.893
Luxembourg	0.474	0.894
France	0.351	0.590
Belgium	0.377	0.864
Germany	0.336	0.655

Table 4.4: Average coherence score and average silhouette score of CTE.

Average coherence score [OGC$^+$15, NLGB10] and average silhouette score [AT08] are utilised as the metrics of CTE. We calculated an average coherence score by calculating the topic coherence for each topic individually and averaging them. The hyper-parameters are tuned to obtain the best results. The value of k is chosen from $\{1, 2, \ldots, 15\}$ and the value of γ is chosen from $\{0.1, 0.2, \ldots, 0.9\}$. The model arrive at the optimal with $k = 7$ and $\gamma = 0.5$.

The results are shown in Table 4.4 and a sample of clustering result from UMAP is shown in Figure 4.12. It can be observed from Table 4.4 and Figure 4.12 that the results generated by CTE are coherent and can be observed as well-separated clusters.

We extract topics by day for 137 days from the text of each country's tweets and region and get a total of 4795 topics. Since the essence of CTE is to cluster the tweet text's embedding vectors and extract the top ten words with the highest frequency in the tweet text corresponding to all vectors in each cluster as the final topic. The clusters containing too few tweets and their corresponding topics do not convey information well, so we remove the topics containing no more than 2 tweets from the clusters and end up with 4,763 topics. Then we randomly selected 51% (2,435 in total) of the topics from each country and region for manual labelling following the central idea existing work [LLLY04, LPN$^+$11]. We used three annotators to label these topics and only labels that are agreed by at least two annotators can be used as the final label.

Topic classification. We classified the 2,435 topics manually into the following 7 categories:

1. 'Wuhan & China': Topics about Wuhan and China.
2. 'Measures': Topics about basic information including symptoms, anti-contagion and treatment measures of COVID-19.
3. 'Local news': Topics about local COVID-19 news, including daily new cases, deaths, etc.
4. 'International news': Topics about international COVID-19 news
5. 'Policy and daily life': Topics about COVID-19 related policies encompass lockdown, closure of borders, limits on public gatherings and the impact of the policies on daily life.
6. 'Racism': Topics about racism.
7. 'Other': Other topics.

The division of these 7 categories is based on the classification of COVID-19 related Twitter topics analysis in existing studies [AAAH+20], and is determined empirically on the basis of common knowledge and the status quo.

These manually classified topics are utilised to train a Support Vector Machine (SVM) [CL11] for supervised classification. The reasons for training a classifier instead of manually labelling all the topics are, on the one hand, the classifion of all the topics manually is time-consuming, and, on the other hand, the classifier can be used in further studies.

Words of each topic are converted to word frequency vectors with TfidfVectorizer and country are encoded with Label Encoder [PVG+11]. The feature vector is consisted by these two elements. Since our manually labelled dataset is imbalance in classification, Synthetic Minority Oversampling Technique [CBHK02] is utilised for oversampling imbalanced the dataset and mitigate imbalances. The dataset is split, 80% of which is the training dataset and 20% the test dataset. Grid search with 10-fold cross-validation is deployed on training dataset to find the optimal hyper-parameter, and the final SVM model is obtained with the entire training set Table 4.5 shows the precision, recall, F1 score, support and Macro-average F-Score of the trained classifier for each topic category. Then, the obtained SVM model classifies the rest of topics. Table 4.6 shows the number of topics of each category for each country and region.

The categories with higher percentages are topics of Wuhan & China and policy and daily life. In general, the number of topics about policy and daily life is much higher in Luxembourg (56.6%) than in other countries (ave = 33.0%). France, on the other hand, shows a high level of interest in local news (30.2%), compared with other countries (9.4%). In terms of the overall data of the GR, however, it does not show particular differences compared with other countries. Note that as there may be cases where the cluster for a topic contains no more than two tweets, we treat such topics as the invalid topic and remove them. This leads to a different total number of topics in each country. Next, we introduce dates to plot the changes in categories over time.

Category	Precision	Recall	F1-score	support
1	0.89	0.77	0.82	163
2	0.92	0.93	0.93	166
3	0.80	0.79	0.80	155
4	0.74	0.86	0.80	155
5	0.73	0.68	0.71	149
6	0.99	1.00	0.99	157
7	0.97	1.00	0.98	142
Macro avg	0.86	0.86	0.86	1,087

Table 4.5: Metrics of the classification results.

Figure 4.13 shows the tweet volume contained in each category demonstrated in the form of percentage of the total tweet volume on that day (CR), with the darker red representing higher CR. The interval colored in white represents the period from 22 January to the start of *Pre-peak* period, other regions in different colours indicate, in sequence, *Pre-peak* period, *Free-contagious* period, *Measures* period, and *Decay* period. The black dotted line illustrates the date on which the first case appeared. The figure shows an interval between the date of the first case and the date of consecutive cases every day in the GR. The solid black line indicates the date that new cases appear every day since that date. For ease of discussion, we name the day as 'outbreak day' (OD).

Conclusion. In this part, we aim to explore the changes in the categories of topics discussed over time in each country and region, focusing on understanding the evolving discourse in the context of the COVID-19 pandemic.

Figure 4.13 reveals that initially, the main topic in all the countries and region is about China, but over time the categories of topics change. In France, Germany and Belgium, the appearance of the first case trigger only a small amount of discussions about the protective measures, and related discussions do not start to increase until OD. In other words, the public concerns in these region and countries do not really heed the protective measures until OD, when the virus is already spreading. The change in topic is at odds with the conclusion of Bento et al. [BNW^{+}20] that the announcements of the first case have the greatest impact on the public concerns for searching basic information about COVID-19 and its symptoms.

Moreover, the report of first case does not stimulate discussions about policies and daily life as well, and discussions about it do not emerge frequently until OD. This may be explained by the existence of a large interval between the date of the first case and OD (27.3 days on average) in France, Germany, and Belgium. During this interval, sporadic cases may not attract enough public concerns, and the public's concerns is still focused on China-related news.

Category	the GR	Luxembourg	Belgium	France	Germany	Total
1	245	168	287	202	315	1,217
2	64	34	48	65	41	252
3	99	44	109	285	110	647
4	134	77	114	52	167	544
5	353	525	370	250	295	1,793
6	23	7	23	31	15	99
7	41	72	15	60	23	211
Total	959	927	966	945	966	4,763

Table 4.6: Topic volume for each category/country (region)

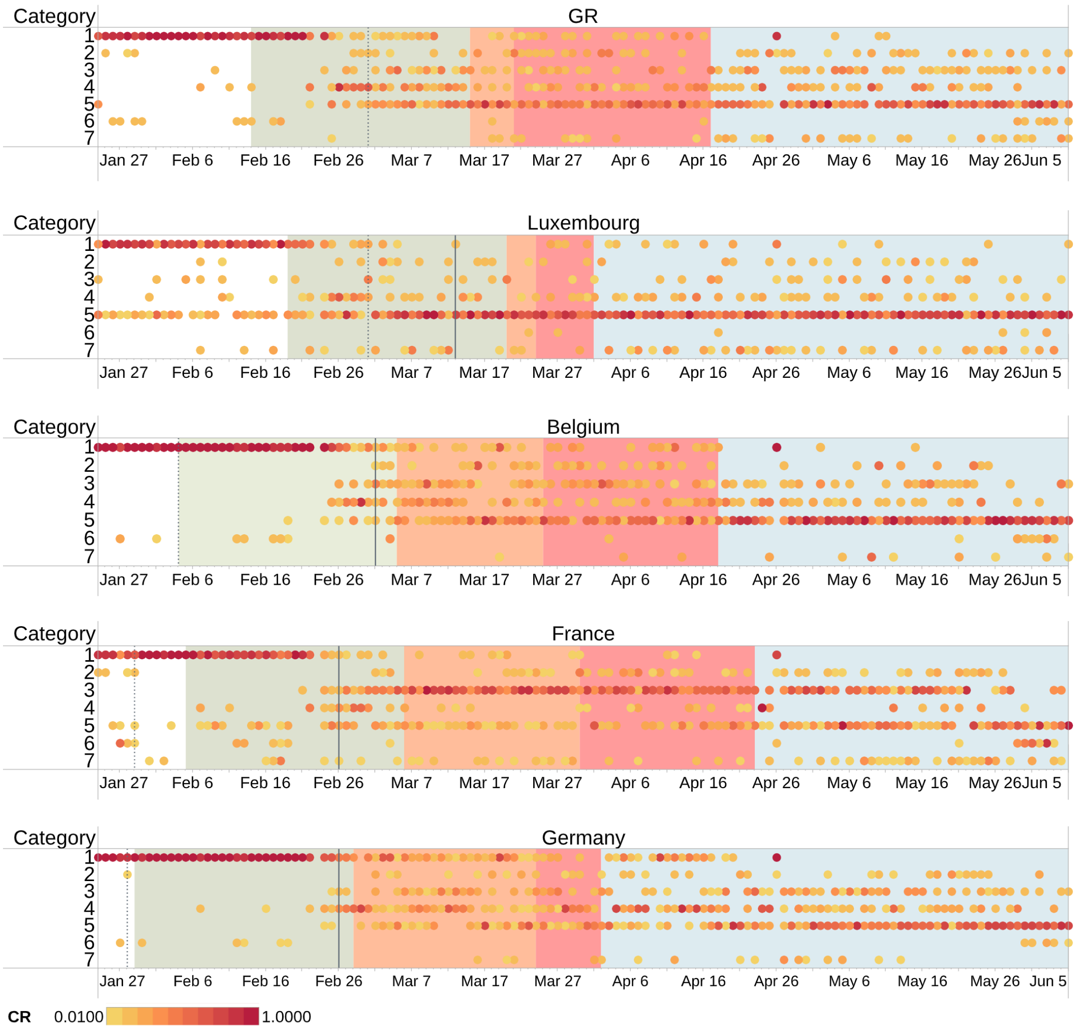


Figure 4.13: Topic categories in the GR and related countries.

1: Wuhan& China; 2: Measures; 3: Local news; 4: International news; 5: Policy and daily life; 6: Racism; 7: Other.

The situation is different in the GR, a relational city, and in Luxembourg, its centre. Figure 4.13 shows that the public in Luxembourg and the GR start to have discussions about measures $1-2$ days before the first case appears. Furthermore, during the *Pre-peak* period, the CR of measures is much higher in the GR (3.41%) and Luxembourg (7.62%) than in France (1.90%), Belgium (1.84%) and Germany (0.0%). It

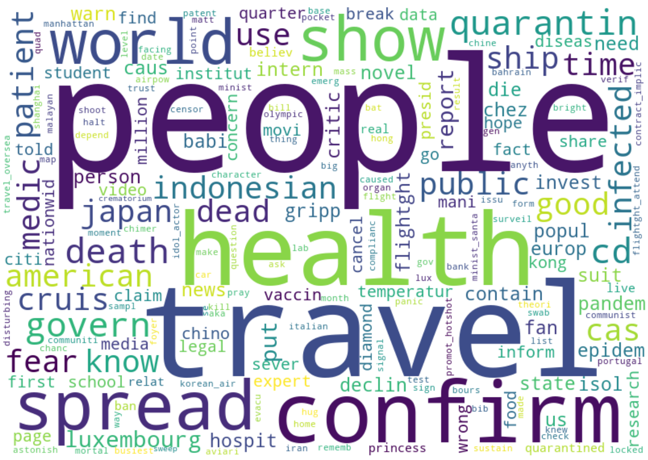


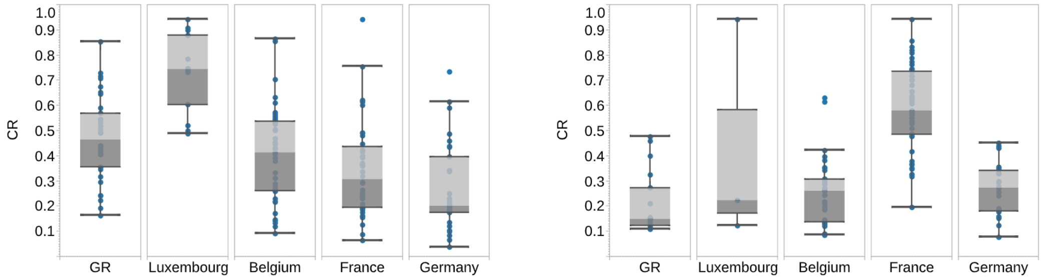
Figure 4.14: Word cloud of Luxembourg Tweets from 2020-01-22 to 2020-03-01.

should be noted that discussions of measures are not totally non-existent in Germany, but the tweet volume may be too small to be recognised as separate topics during the topic modelling process. By comparing the topics discussed in other countries of the same time, this may be explained by the late occurrence of the first case in Luxembourg and the GR, where the other three countries have already passed OD, the outbreak in other countries may have attracted public concerns in the GR and Luxembourg. Concurrently, the results indicate that the GR exhibits a high level of interest in policy and daily life with 47.1% of total tweet volume during the *Free-contagious* and the *Measures* period, while for Luxembourg, this rate is 66.1%. Figure 4.15(a) shows boxplots of the distribution of the CR on policy and daily life during the *Free-contagious* and the *Measures* period. This shows that the public is more responsive to policies as a region that relies on foreign labour and has high mobility than Belgium, France and Germany.

The reason why *Free-contagious* is a period more transient in Luxembourg and the GR compared with other regions is still unclear, but part of the reason may stem from the fact that the public concerns to the virus itself during *Pre-peak* period led to better responsiveness to the anti-contagion policies in these region and countries. Interestingly, in Luxembourg, the discussion about policies and daily life persisted before the first case is announced and increased immediately after then. A word cloud of the topics from 22 January to 1 March (date of the first case) of Luxembourg is depicted in Figure 4.14, this shows that the topics are mainly travel-related. This may be explained by the fact that the proportion of foreign residents in the Luxembourg region is 47.4% [8], and residents are more concerned about travel-related policies in Luxembourg and other countries.

In addition, Figure 4.15(b) illustrates that the *Free-contagious* and *Measures* periods coincided with the France municipal election, and thus the public concerns in local news among French is higher. In the end, during the *Decay* period, while there is a downward trend ($p < 0.05$) in the total daily tweet volume, there is a upward trend ($p < 0.05$) in the CR of policy and daily life, except in Luxembourg, where the rate is consistently high.

[8] https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12856



(a) Policy and daily life

(b) Local news

Figure 4.15: Distribution of proportion of tweets on 'policy and daily life' and 'local news' during Free-contagious and Measures period.

4.3.4 Conclusion and Discussion

In this chapter, we delve deeper into the two datasets, GR-EGO and EU-Vax, which were collected based on the data collection method described in the previous chapter. Our exploration aims to provide valuable insights into public sentiment and behavior during the COVID-19 pandemic by examining various aspects of these datasets, such as the relationship between tweet volume and COVID-19 cases, the evolution of topic categories, and the public's attitudes toward vaccination.

For the GR-EGO dataset, we examine the correlation between tweet volume and COVID-19 daily cases in the GR and related countries. Additionally, we explore the changes in the categories of topics discussed in each country and region over time. This helps us understand how public concerns and reactions evolved throughout the pandemic, and how they varied between different countries.

With the EU-Vax dataset, we focus on understanding public attitudes toward COVID-19 vaccination in Western Europe. We perform statistical analysis and visualization, and evaluate the performance of established NLP benchmarks. This allows us to track temporal changes in vaccination attitudes and identify sudden shifts in public sentiment that may warrant further attention.

Through the comprehensive analysis of both datasets, we aim to shed light on the challenges facing virtual ethnography and demonstrate the potential value of social media data in understanding and analyzing public sentiment and behavior during the pandemic. By providing these insights, we hope to contribute to the development of more effective public health measures and policy implementations, ultimately improving the response to public concerns and needs during and beyond the pandemic.

Part II

Social Characteristics Analysis

Chapter 5

Bridging Performance and Subjective Well-Being of Twitter Users

In the previous section, we focus on addressing the first research question: how to collect a sufficient amount of data to meet the requirements of a virtual ethnographic study. We describe the collection methods for the two datasets primarily used in the previous chapters and conducted a preliminary exploration of these datasets, validating their feasibility for meeting virtual ethnographic research needs.

Building on these datasets, we devote a chapter to the second research question which centers on the use of social characteristics: how to design a measurement to capture the actual bridging performance of social media users in terms of spreading COVID-19-related information? Specifically, we investigate how to identify influential Twitter users and their role in spreading credible information during the COVID-19 pandemic. To do so, we introduce two new measurements called "bridging performance" to assess the speed and breadth of information diffusion due to users' sharing activities. Through the analysis of our tweet collection, we uncover the mental anguish experienced by influential users during the pandemic. Using hierarchical multiple regression analysis, we establish a strong relationship between individual users' subjective well-being (SWB) and their bridging performance. In addition, we extend the concept of bridging performance from individuals to subgroups of users, allowing for more in-depth analysis based on users' multilingualism. We study the role of multilingual users in spreading COVID-19 messages and investigate their SWBs during the pandemic. We find that multilingual users not only suffered from a much lower SWB in the pandemic, but also experienced a more significant SWB drop.

5.1 Introduction

In response to the infodemic, governments and healthcare bodies have outlined a series of effective social media-based countermeasures to spread trustworthy information to overcome the negative impact of misinformation. With the purpose of promoting the speed and wideness of information spread, individual users with a large number of followers are invited to share messages [Zar20, BM21]. Meanwhile, various user subgroups, including healthcare professionals and social activists, also voluntarily take part in relaying information they consider useful with their social media accounts. All these users in fact act the role of bridges on social media delivering information to the public, although their *bridging performance* varies. In this thesis, we use the term 'bridging performance' as an analogy to evaluate how quickly and widely information can diffuse on social media because of the sharing of a user.

Subjective well-being (SWB) evaluates individuals' cognitive (e.g., life satisfaction) and affective (i.e., positive and negative) perceptions of everyday lives [JGS+20]. The global decrease in the public's SWB has been unanimously recognised since the outbreak of the COVID-19 pandemic. The SWBs of various sub-populations such as immigrants and healthcare workers [HLKX20, ENW+21] has been studied extensively. Such research manages to identify the vulnerable populations which deserve special attention. Despite the contribution of the influential users on social media in combating the infodemic, no work has been conducted to examine and analyse their psychological status. In this chapter, we perform the first attempt to study the negative impact of the COVID-19 pandemic on the subjective well-being of this specific group of people.

There are two challenges that have to be overcome in advance to conduct this analysis. First, no measurements exist to accurately quantify social media users' actual performance in promoting the spread of COVID-19 related information. The measurements, used in crisis communications and marketing, identify influential users based on social relations. Their effectiveness are found to deteriorate when capturing users' real bridging performance, especially in the currently ongoing pandemic [Str20]. Take healthcare professionals on Twitter as an example. Their professional endorsement significantly promotes the popularity of the tweets they shared [Str20]. However, they usually do not belong to the super tweeters who have thousands of followers. Our second challenge is how to obtain the SWB levels of a sufficiently large group of social media users whose bridging performance is also available.

In this chapter, we address the identified challenges by leveraging the information outbreak on social media triggered by the COVID-19 pandemic and the recent advances of deep learning in text mining. To address the first challenge, we start by proposing a new measurement based on *information cascades* instead of relying only on social relations between users to comprehensively quantify individual users' influences in diffusing COVID-19 related information. We also extend the measurement to quantify and compare the bridging performance of user subgroups. To address the second challenge, we take advantage of the success of deep learning in Natural Language Processing (NLP) and quantitatively evaluate individual users' SWB with the sentiments of their posts. Recent studies [JGS+20] have illustrated the advantages of social media posts in extracting well-beings, especially with data-driven methods. In this chapter, we follow the same data-driven methodology and adopt the state-of-the-art text embedding based on transformers. Compared to traditional machine learning models which rely on manually pre-defined features, it allows us to automatically learn representative features of textual posts.

We make use of Twitter as a representative social media platform considering its increasing popularity during the pandemic. It had become one major source for the public to acquire COVID-19 related information, especially during the first few months after the onset of the pandemic. In this Chapter, we use the GR-ego Dataset we collect in Chapter 3, and our contributions are summarised as follows:

- We propose a new measurement to capture the actual bridging performance of individual social media users in diffusing COVID-19 related information. Compared to existing social connection-based measurements, it is directly derived from information diffusion history. This measurement allows for identifying the accounts of influential health professionals and volunteers that are missed by existing ones.
- Through deep learning-based text embedding methods, we implement a classification model which can accurately extract the sentiments expressed in social media messages. With the sentiments of posts, we quantitatively estimate the SWB of individual users, and discover that influential individual users are more affected in their SWB during pandemics.
- Through the hierarchical multiple regression model, we reveal for the first time that users' SWB has a strong negative relationship with their bridging performance in COVID-19 information diffusion, but a weak relationship with their social connections.
- We identify the insufficiency of our bridging performance measurement for individual users in comparing the bridging performance between user subgroups. We thus extend it to the level of user subgroups. With the new measure, we successfully re-confirm the bridging role of multilingual users in information diffusion during the pandemic, and reveal the more drastic adverse impact of the pandemic on their SWB. This complements the previous studies before the pandemic claiming that monolingual users have lower SWB due to their language boundaries [Tra95, PA08].

Our research provides policy makers with an effective method to identify influential users and user subgroups in the fight against infodemic. In addition, we highlight the need to pay particular attention to the mental health of people who are actively involved in transmitting information such as multilingual users, in the case of COVID-19 or other potential future large-scale infectious diseases of similar type.

5.2 Related Work

5.2.1 Impact of the Pandemic on Subjective Well-being

The negative impact of the COVID-19 pandemic on mental health and well-being has been acknowledged as a global health concern. Statistics have been reported on various subgroups of people with different demographic characteristics. More than 4 in 10 adults in the US suffered from symptoms of anxiety and depressive disorders in January 2021 compared to 1 in 10 before the pandemic[1]. Females and young people are found to be more negatively respond to the loneliness caused by social distancing [TO21, OCM+21]. In Japan, the monthly suicide ratio increased by 37% for women and 49% for children and adolescents while the average ratio increased by 16% [TO21]. The vulnerable subgroups such as migrants [QSZ+20, YNF+20] and refugees [EJB20] have also been found deserve more attention due to their relatively bad economic conditions. As an indispensable part of mental health, subjective well-being has also drawn public attention. Multiple factors have been found associated with its decline after the onset of the pandemic such as country-specific pandemic severity [FFG22] and satisfaction with health [HGS22].

In this chapter, we concentrate on the SWB of a specific group of people who are actively involved in information spread on social media and have not been studied in the literature. Furthermore, we will examine whether their contribution to COVID-related information diffusion can be used as an effective predictor of their SWB.

5.2.2 Measuring Bridging Performance

A considerable amount of literature has been published quantifying users' bridging performance based on social connections to identify amplifiers in social media. We can divide the measurements into two types. The first type of measurements implicitly assume that influential users are likely to hold certain topology properties on social networks such as large degrees, strong betweenness centrality or community centrality [Fre78, ZL17, GCCH19]. Degrees only capture the number of users' local social connections without considering their overall position in the whole social network. Betweenness centrality [ZL17] measures the importance of a user connecting other users in the network while community centrality [GCCH19] measures a user's importance in connecting communities. Measurements are also proposed to fuse multiple topology properties such as fusing degree, ego-betweenness centrality and eigenvector centrality into an overall evaluation [HLZ+14]. It has been noticed that such measurements are usually inefficient to compute in sparse networks where node degrees do not follow power law distributions [FG22]. The second type of measurements assume that influential users tend to be more likely reachable from other users through random walks. PageRank [PBMW99] and its variant TwitterRank [WLJH10] among the representative benchmarks of this type of measurements. PageRank is calculated only with network structures while TwitterRank additionally takes into account topic similarities between users. All the two categories of measurements have been widely applied in practice, from public health crisis communication [MBS+20] to online marketing [LLC11]. However, recent studies have pointed out that they may not truly capture users' actual bridging performance in information diffusion during a specific public healthy crisis [MBS+20].

[1]https://www.cdc.gov/nchs/data/nhis/earlyrelease/ERmentalhealth-508.pdf

One of our goals in this chapter is to define new measurements that can capture social media users' bridging performance in information diffusion during global public health crises like COVID-19 in real-world social networks.

5.2.3 Subjective Well-being Extraction

Subjective well-being is used to measure how people subjectively rate their lives both in the present and in the near future [DOL03]. Many methods have been used to assess subjective well-being, from traditional self-reporting methods [DELG85] to the recent ones exploiting social media [YS16].

Studies have cross-validated the trustworthiness of social media as a complementary data source for public opinions [NCP+20, JGS+20]. Chen et al. [CCZP22c] extracted vaccine hesitancy from Twitter during the COVID-19 pandemic and empirically illustrated its consistency with social surveys across regions and along with time. With the Gallup-Sharecare Well-Being Index survey,[2] a classic reference used to investigate public SWB, SWB extracted from social media has been shown as a reliable indicator of SWB [JGS+20]. Twitter-based studies usually calculate SWB as the overall scores of positive or negative emotions (i.e., sentiment or valence) [JGS+20]. Sentiment analysis has developed from the original lexicon-based approaches [BL99] to the data-driven ones which ensure better performance [JGS+20]. We adopt the recent advances of the latter approaches, and make use of the pre-trained XLM-RoBERTa [OL20], a variant of RoBERTa [LOG+19], to automatically learn the linguistic representation of textual posts. As a deep learning model, RoBERTa and its variants have been shown to overwhelm traditional machine learning models in capturing the linguistic patterns of multilingual texts [BCCEAN20].

5.3 Data Processing

5.3.1 Cascade Computation

A cascade records the process of the diffusion of a message. It stores all activated users and the time when they are activated. In our dataset, a user is activated in diffusing a message when he/she retweets the message. In this chapter, we adopt the widely accepted *cascade tree* to represent the cascade of a message [WSL+17], the detailed definition is given in Section 2.3.

For our study, we follow the method in [KOU+12] to construct tweet cascades. Recall that when a tweet's status is '*Retweeted*', the ID number of the original tweet is also recorded. We first create a set of original tweets with all the ones labelled in our meta data as '*Original*'. Second, for each original tweet, we collect the IDs of users that have retweeted the message. At last, we generate the cascade for every original tweet based on the following relations in our GR-ego social network and their retweeting time stamps. We eliminate cascades with only two users where

[2] www.gallup.com/175196/gallup-healthways-index-methodology.aspx

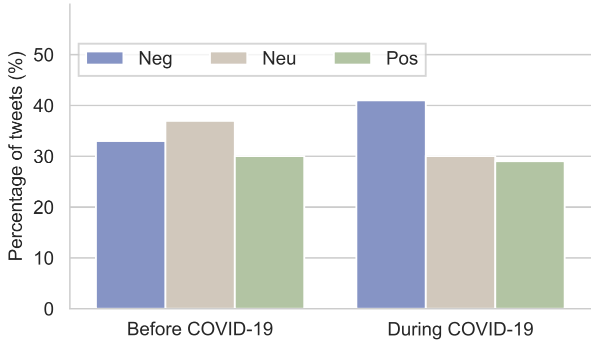


Figure 5.1: Sentiment distribution of users' timeline tweets.

messages are just retweeted once. In total, 614,926 cascades are built and the average size of these cascades is 7.13.

5.3.2 Sentiment Analysis

Previous works [ZJZ20] leverage user-provided mood (e.g., angry, excited) or status to extract users' sentiment (i.e., positive or negative) and use them to approximately estimate affective subjective well-being. However, such information is not available on Twitter. We refer to the sentiments expressed in textual posts to extract users' SWB. In this chapter, we treat sentiment extraction as a tri-polarity sentiment analysis for short texts, and classify a tweet as *negative*, *neutral* or *positive*. In order to deal with the multilingualism of our dataset, we benefit from the advantages of deep learning in sentiment analysis [BCCEAN20], and build an end-to-end deep learning model to conduct the classification. Our model is composed of three components. The first component uses a pre-trained multilingual language model, i.e., XLM-RoBERTa [OL20], to calculate the representation of tweets. The representations are then sent to the second component, a fully-connected ReLU layer with dropout. The last component is a linear layer added on the top of the second component's outputs with sigmoid as the activation function. We use cross-entropy as the loss function and optimise it with the Adam optimiser.

Model training and testing. We train our model on the *SemEval-2017 Task 4A* dataset [RFN17], which has been used for sentiment analysis on COVID-19 related messages [DLP+20]. The dataset contains 49,686 messages which are annotated with one of the three labels, i.e., positive, negative and neutral. We shuffle the dataset and take the first 80% for training and the rest 20% for testing. We assign other training parameters following the common principles in existing works. We run 10 epochs with the maximum string length as 128 and dropout ratio as 0.5. When tested with macro-average F1 score and accuracy metrics, we achieve an accuracy of 70.09% and macro-average F1 score of 71.31%.

Despite its effectiveness in classifying *SemEval-2017 Task 4A* data, in order to check whether such performance will persist on our GR-ego dataset, we construct a new testing dataset. This dataset consists of 500 messages, 100 for each of the top 5 most popular languages. We hire two annotators to manually label the selected tweets and the annotated labels are consistent between them with Cohen's Kappa coefficient $k = 0.93$. When applied on this annotated dataset, our trained model achieves a similar accuracy of about 87%.

Analysing our GR-ego dataset. Before applying our sentiment classification model on our GR-ego dataset, we clean tweet contents by removing all URLs, and mentioned usernames. Figure 5.1 summarises the statistics obtained from user timeline tweets before and during the pandemic. The numbers of users' timeline tweets are consistent with previous studies. For instance, users tend to become more negative after the outbreak of the COVID-19 pandemic [ENW+21, HLKX20].

5.3.3 Measuring SWB

We extend the definition proposed in [ZJZ20] to measure the level of subjective well-being of users based on the sentiment expressed in their past tweets. Specifically, we extend it from bi-polarity labels, i.e., negative and positive affection, to tri-polarity with neutral sentiment by multiplying a scaling factor to simulate the trustworthiness of the bi-polarity SWB.

Definition 4 (Social media Subjective well-being value (SWB))**.** *We use $N_p(u)$, $N_{neg}(u)$ and $N_{neu}(u)$ to denote the number of positive, negative and neutral posts of a user u, respectively. The subjective well-being value of u, denoted by $swb(u)$, is calculated as:*

$$\frac{N_p(u) - N_{neg}(u)}{N_p(u) + N_{neg}(u)} \cdot \left(\frac{N_p(u) + N_{neg}(u)}{N_p(u) + N_{neg}(u) + N_{neu}(u)}\right)^{\frac{1}{2}}.$$

If all messages are neutral, then $swb(u)$ is 0.

Discussion. Note that i) consistent with [ZJZ20], we focus on affective SWB (i.e., positive and negative) in this chapter, while ignoring its cognitive dimension; ii) users' SWB is evaluated based on their original messages: originally posted tweets and quotations; iii) for tweets with quotations to other messages, only the texts are considered without the quoted messages. As retweets may not explicitly include users' subjective opinions, we exclude them from the SWB calculation.

5.4 Bridging Performance of Individual Users in Information Diffusion and its Relation with SWB

We devote this section to measuring the bridging performance of individual users in the diffusion of COVID-19 related information. Through experimental validation and manual analysis, we validate the effectiveness of our proposed measurement in identifying influential social media users.

5.4.1 Measuring Individual User Bridging Performance

We evaluate individual users' overall performance in the diffusion of observed COVID-19 related messages. As a user can participate in diffusing multiple messages, we first focus on her/his importance in the diffusion of one single message and then combine her/his importance in all messages into one single measurement. Intuitively, we consider a user *more important* in diffusing a message when his/her retweeting behaviour activates more users, or leads to a given number of activated users with fewer subsequent relays, e.g., retweets in Twitter. In other words, a more important user promotes the popularity of the information or accelerates its transmission.

Given a cascade path $S = (u_1, u_2, \ldots, u_n)$, we use $S^*(u_i)$ $(1 \leq i < n)$ to denote the subsequence composed of the nodes after u_i (including u_i), i.e., $(u_i, u_{i+1}, \ldots, u_n)$. For any u that does not exist in S, we have $S^*(u) = \varepsilon$ where ε represents an empty sequence and its length $|\varepsilon| = 0$. In the following, we define how to quantify a user's contribution in the diffusion of a given message as a transmitter.

Definition 5 (Cascade bridging value)**.** *Given a cascade tree C and a user u ($u \neq r(C)$), the cascade bridging value of u in C is calculated as:*

$$\alpha_C(u) = \left(\sum_{S \in C} \frac{\mid S^*(u) \mid}{\mid S \mid} \right) / |C|.$$

Note that our purpose is to evaluate the importance of users as transmitters of messages. Therefore, the concept of cascade bridging value is not applicable to root users, i.e., the message originators.

Example 1. *In Figure 2.1(b), u_3 participated in two cascade paths, i.e., $S_1 = (u_1, u_3, u_4)$ and $S_2 = (u_1, u_3, u_6)$. Thus, $S_1^* = (u_3, u_4)$ and $S_2^* = (u_3, u_6)$. We then have $\alpha_C(u_3) = \frac{2/3+2/3}{3} \approx 0.44$.*

In Definition 5, we do not simply use the proportion of users activated by a user in a cascade to evaluate her/his bridging performance. This is because that only allows for capturing the number of activated users. The speed of the diffusion will be ignored. Take u_2 in Figure 2.1(b) as an example. According to our definition, $\alpha_C(u_2) = 0.25$ which is smaller than $\alpha_C(u_3)$. This is due to the fact that u_2 activated two users through two re-transmissions while u_3 only used one. However, if we only consider the proportion of activated users, the values of these two users will be the same.

With a user's bridging value calculated in each cascade, we define *user bridging magnitude* to evaluate her/his overall importance in the diffusion of the observed messages. Intuitively, we first add up the cascade bridging values of a user in all his/her participated cascades and then normalise the sum by the maximum number of cascades participated by a user. This method captures not only the bridging value of a user in each participated cascade, but also the number of cascades she/he participated in. This indicates that a more active user in sharing COVID-19 related information is considered more important in information diffusion.

Table 5.1: Comparison of bridging performance with benchmarks.

	in-degree	PageRank	TwitterRank	Betweenness centrality	Community centrality	UBM
Avg. #activated user/minute	0.042	0.057	0.064	0.043	0.056	0.104
Avg. #activated users	13.99	16.84	17.68	15.54	17.00	23.81
%impacted user	32.17	52.54	57.44	43.44	56.54	71.66

Definition 6 (User bridging magnitude (UBM))**.** *Let $\mathcal{C}$ be a set of cascades on a social network and $\mathcal{U}$ be the set of users that participate in at least one cascade in $\mathcal{C}$. A user u's user bridging magnitude (UBM) is calculated as:*

$$\omega_{\mathcal{C}}(u) = \frac{\sum_{C \in \mathcal{C}} \alpha_C(u)}{\max_{u' \in \mathcal{U}} |\{C \in \mathcal{C} | \alpha_C(u') > 0\}|}.$$

With this measurement, we can compare the bridging performance of any two users, and learn which one plays a more important role in information diffusion.

5.4.2 Validation of UBM

As discussed above, we make use of observed cascades instead of network typologies to estimate the capacity of individual users in promoting the diffusion of COVID-19 related information. In order to illustrate its advantages against existing measurements, we evaluate our UBM measurement from two perspectives. With our dataset, we empirically validate whether the influential users identified by UBM can improve the speed and popularity of information diffusion. Then we examine the profiles of the identified influential users and cross-validate whether their composition is consistent with previous social studies in the literature.

Experimental results. We compare the effectiveness of our UBM to five widely used topology-based measurements in the literature, i.e., in-degree, PageRank [PBMW99], TwitterRank [WLJH10], betweenness centrality and community centrality [Fre78]. We randomly split the set of cascades into two subsets. The first subset accounts for 80% of the cascades and is used to calculate the bridging performance of all users. Then we select the top 20% users with the highest bridging performance in every adopted measurement and use the other subset to compare their actual influences in information diffusion. We adopt three measurements to quantitatively assess the effectiveness of UBM and the benchmarks. We use the *average number of activated users per minute* to evaluate the efficiency of the information diffusion. The more users activated in a minute, the faster information can be spread when it is shared by the influential users. The *average number of activated users* counts the users who received the information after the retweeting behaviour of an identified influential user. It is meant to evaluate the expected wideness of the spread once an influential user retweets a message. The *percentage of impacted users* gives the proportion of users that have ever received a message due to the sharing behaviours of identified influential users. This measurement is to compare the overall accumulated influence of all the selected influential users. We show the results of UBM and other

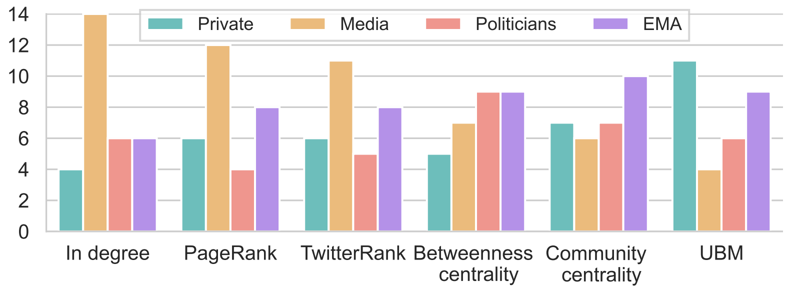


Figure 5.2: Profile distribution of the top 30 accounts with highest bridging performance

benchmark measurements in Table 5.1. We can observe that it takes less time on average for the influential users identified according to UBM to activate an additional user, with 0.104 users activated a minute due to their retweets. With 23.81 users activated, UBM allows for finding the users whose retweeting action can reach more than 35% users than those identified by the benchmarks. In the end, the top 20% influential users identified by UBM spread their shared information to 71% users in our dataset, which overwhelms that of the best benchmark by about 15%. From the above analysis in terms of the three measurements, we can see that our UBM can successfully identify influential users whose sharing on social media manages to promote the wideness and speed of the diffusion of COVID-19 information.

Manual analysis. In order to understand the profiles of the calculated influential users by the measurements, we select the top 30 users with the highest bridging performance of each measurement. We identify four types of user profiles: *private*, *media*, *politicians* and *emergency management agencies* (EMA). Figure 5.2 shows the distributions of their profiles. We can observe that the distributions vary due to the different semantics of social connections captured by the measurements. For instance, due to the large numbers of followers, Twitter accounts managed by traditional media are favoured by in-degree. This obviously underestimates the importance of accounts such as those of EMAs in publishing pandemic updates. With reachability and importance in connecting users and communities considered, more accounts of politicians and EMAs stand out. The proportion of private accounts also starts to increase. When UBM is applied, the percentage of private accounts becomes dominant. A closer check discovers that 10 out of the 11 private accounts belong to health professionals and celebrities. This is consistent with the literature [HHW+21] which highlights the importance of health professionals and individuals in broadcasting useful messages about preventive measures and healthcare suggestions in the pandemic.

5.4.3 Impact of the Pandemic on SWB of Influential Users

With the proposed SWB measurement in Definition 4, we study how users' subjective well-beings change due to the outbreak of the COVID-19 pandemic. As our target is the SWBs of natural persons, the accounts of organisations and bots should be excluded from our analysis. We make use of existing methods/tools to identify these two types of accounts. We detected about 12.04 organisation accounts

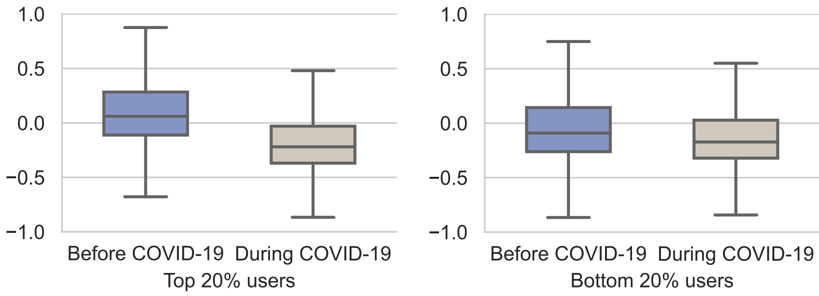


Figure 5.3: SWB changes after the outbreak of the pandemic.

which are about 8.16% of the selected active users with the methods proposed in [WHaPAG+19]. We use Botometer [SVY+20] to detect bot accounts and only 131 users are classified as bots. In total, we removed 1,333 users (as 2 users are identified as both bot and organisation) from our collected dataset.

We calculate the UBM values of the remaining users and order them in descending order. Then, to compare the response of these two groups to the pandemic, we select the top 20% users and the bottom 20% users. For each group, we calculate users' SWBs according to their posts before the pandemic and after the pandemic to capture the changes. Note that we only consider the users with more than 5 posts in each time period.

In Figure 5.3, we show the SWB distributions of the two user groups. On average, the users with high UBM have positive SWB of 0.13 before the pandemic while the users with low UBM are negative. The SWB of both user groups decreases after the pandemic but the SWB of the top 20% users drops more significantly. Specifically, their SWB falls by 0.38, which is two times as much as that of the bottom 20% users. The lowest value of the top 20% users' SWB slightly decreases after the pandemic, while the lowest value of the bottom 20% of users does not change significantly. Note that the minimum values here do not include outliers that lie outside the box whiskers. This indicates that the top 20% users become even more negative than the bottom 20% users, in terms of mean and minimum values. To sum up, the pandemic causes more negative mental impacts on the social media users who play a more important bridging role in transmitting COVID-19 related information.

5.4.4 Relation between SWB and Bridging Performance

We conduct the first attempt to study if a user's bridging performance has a relationship with the SWB changes of the users actively participating the diffusion of COVID-19 related information. In addition to UBM and the five benchmark measurements used in Section 5.4.2, we consider two additional variables: *out-degree* and *activity*. Out-degree is used to check whether the number of accounts a user follows correlates with SWB changes. The activity variable evaluates how active a user is engaged in the online discourse and is quantified by the number of messages he/she posted. In order to isolate the impacts of these variables, we adopt the method of *hierarchical multiple regression* [TFU07]. The intuitive idea is to check whether the variables of interest can explain the SWB variance after accounting for some variables.

To check the validity of applying hierarchical multiple regression, we conduct first-line tests to ensure a sufficiently large sample size and independence between variables. We identify the variables corresponding to community centrality and TwitterRank fail to satisfy the multi-collinearity requirement. We thus ignore them in our analysis. The ratio of the number of variables to the sample size is 1:1,917, which is well below the requirement of 1:15 [TFU07]. This indicates the sample size is adequate. We iteratively input the variables into the model with three stages. The results are shown in Table 5.2. In the first stage, we input the variables related to network structures, i.e., in-degree, out-degree, Pagerank and Betweenness centrality. The combination of the variables can explain 4.80% of the SWB variance ($F = 4.672, p < 0.05$). Note that an F-value of greater than 4 indicates the linear equation can explain the relation between SWB and the variables. This demonstrates that there exists a positive relationship between the topology-based variables and SWB, but this relationship is rather weak. A closer check on the t-values show that out-degree is irrelevant to SWB and the rest three variables are weakly related. In the second stage, we add the variable of activity to the model. After controlling all the variables of the first stage, we observe that user activity does not significantly contribute to the model with t-value of 0.396. This suggests that user activity is not a predictor of SWB. In the third stage, we introduce UBM to the model. The addition of UBM, with the variables in the previous two stages controlled, reduces the R value from -0.335 to -0.603. UBM contributes significantly to the overall model with $F = 167.32$ ($p < 0.001$) and increases the predicted SWB variance by 25.1%.

Together with the t-value of -11.684 ($p < 0.001$), we can see there exists a strong negative relation between UBM and SWB, and UBM is a strong predictor for SWB.

Discussion. To conclude, the results illustrate that UBM is strongly related to SWB, while in-degree, Pagerank and betweenness centrality are weakly related. This difference further shows that UBM can more accurately capture users' behaviour changes after the outbreak of the pandemic while topology features remain similar to those before the pandemic. This may be explained by the recent studies [HLKX20] that once considered as a change in life after the pandemic outbreak, this extra bridging responsibility in diffusing COVID-19 related messages is likely to associate with lower life satisfaction.

5.5 Comparing the Bridging Performance of User Subgroups

With the above discussion, we have shown how to identify individual influential users. In practice, *subgroup analysis* is also an important analysis methodology to compare how people respond differently to an intervention or event. Samples are divided into multiple subsets, e.g., according to their demographic characteristics [HGS22]. Regarding bridging performance, our purpose can be to understand the different roles played by various user subgroups. Straightforward methods consist in aggregating users' UBM in each group into the overall bridging performance

Table 5.2: Hierarchical multiple regression model examining variance in SWB explained by independent variables, $*p < 0.05$; $**\,p < 0.001$

Variable	*B*	*SEB*	*b*	*t*	*R*	R^2	ΔR^2
Stage 1					-0.219	0.048	0.048
In-degree	0.231	0.102	0.150	2.265*			
Out-degree	0.892	0.676	0.038	1.320			
Pagerank	0.307	0.187	0.122	1.642*			
Betweenness centrality	-3.218	0.800	-0.198	-4.023**			
Stage 2					-0.335	0.112	0.064
In-degree	0.222	0.107	0.181	2.075*			
Out-degree	0.060	0.044	0.077	1.364			
Pagerank	0.348	0.166	0.178	2.096*			
Betweenness centrality	-3.264	0.774	-0.176	-4.217**			
Activity	0.059	0.149	0.089	0.396			
Stage 3					-0.603	0.363	**0.251**
In-degree	0.016	0.107	0.169	0.150*			
Out-degree	0.593	0.481	0.080	1.233			
Pagerank	0.197	0.116	0.170	1.701*			
Betweenness centrality	-1.171	0.511	-0.182	-2.292**			
Activity	0.039	0.169	0.064	0.231			
UBM	-2.255	0.193	-0.182	-11.684**			

of the subgroup. In this section, we discover the deficiency of such methods and propose a new measurement to fix it. In order to validate the effectiveness of our measurement, we conduct a subgroup analysis based on users' *multilingualism* which successfully confirms previous studies about the potential bridging role of multilingual users in social networks. In addition, we analyse the SWB drops of multilingual and monolingual users after the onset of the pandemic and observe the same relation between SWB and the bridging performance among subgroups.

5.5.1 Measuring User Subgroup Bridging Performance

Our UBM measurement focuses on the level of users and evaluates their overall performance across all observed information cascades. It does not compare the relative bridging performance between different user subgroups within the cascades. We take the following example to clarify this deficiency.

Example 2. *Suppose u_2 in Figure 2.1(b) is the only multilingual user. According to Definition 5, we learn that $\alpha_C(u_2) = 0.25$, and as a multilingual user, u_2 plays a more important role in diffusing the message than all monolingual users except for u_3 with $\alpha_C(u_3) = 0.44$. In this example, without a unified standard, we still cannot determine which group of users play a more important role in this cascade.*

We propose a new measurement to compare the bridging performance of different user subgroups. Suppose the set of users $\mathcal{U}$ is divided into multiple subsets, i.e., $\mathcal{S} \subset 2^{\mathcal{U}}$ and for any $S_1, S_2 \in \mathcal{S}$, we have $S_1 \cap S_2 = \emptyset$. Note that $2^{\mathcal{U}}$ denotes the power set of $\mathcal{U}$.

Given a cascade C, we calculate an integrated value through a function γ from the bridging values of the users in each subgroup, denoted by α_C^S for any $S \in \mathcal{S}$. The integrated value can be the *mean*, *median* or *maximum*. Formally,

$$\alpha_C^S \quad = \gamma(\{\alpha_C(u) \mid u \in S\}).$$

The integration function γ should be determined according to real application scenarios. We consider a user subgroup S playing a more important bridging role in a particular cascade C than S' if $\alpha_C^S > \alpha_C^{S'}$. We use the notion *subgroup bridging magnitude* (SBM) to quantify the importance of a user subgroup as a whole in information diffusion.

Definition 7 (Subgroup bridging magnitude (SBM))**.** *Let $\mathcal{C}_S = \{C \in \mathcal{C} | \exists_{u \in S} u \in C\}$ be the set of cascades involving at least one user in subgroup S. The subgroup bridging magnitude (SBM) of the user subgroup S is calculated as follows:*

$$\rho^S = \frac{\sum_{C \in \mathcal{C}} \mathbb{1}(\alpha_C^S = \max_{S' \in \mathcal{S}} \alpha_C^{S'})}{\max_{S' \in \mathcal{S}} |\mathcal{C}_{S'}|}$$

where $\mathbb{1}(\cdot)$ is an indicator function which returns 1 *when the given proposition is true and 0, otherwise.*

5.5.2 Validation of SBM

Due to the privacy policy of Twitter, we cannot obtain users' personal identifying information including any demographic profiles such as age and gender. In order to divide users into subgroups, we make use of one inherent characteristic of our tweet dataset: *multilingualism*. Figure 5.4 presents the distribution of tweets in the top 15 most used languages. We can see that the collected tweets are composed in very diverse languages. The distribution of languages is consistent with that of GR inhabitants' nationalities and the corresponding official languages.[3] Intuitively, according the languages used in users' tweets, we have two user subgroups: *monolingual* and *multilingual*. We will conduct the following analysis based on this division.

Defining multilingualism. No consistent definition exists for the multilingualism of a person. We follow a conservative approach to determine multilingual users based on language usage frequency. We only consider active users who posted more than 5 messages to ensure sufficient evidences. If a user posted or retweeted tweets in more than two languages, we select the language with most tweets as his/her main language. If the messages of the main language make up less than 60%, we consider the user as *multilingual*. Otherwise, the user is considered as *monolingual*. This conservative criteria helps exclude most monolingual users who just infrequently or accidentally retweet or cite information in languages other than their mother tongue. In our dataset, about 37% users are labelled as multilingual.

[3]Key figures on the GR available at https://www.grande-region.lu/portal/.

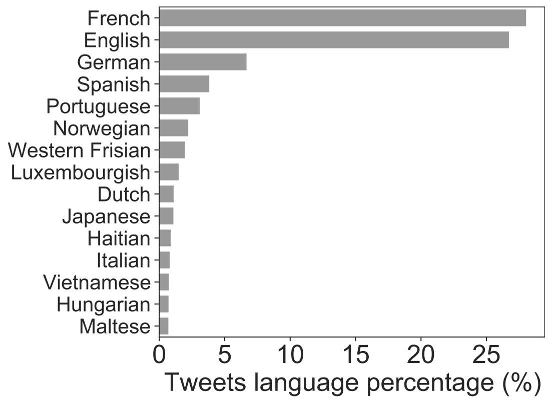


Figure 5.4: Distribution of languages of GR tweets.

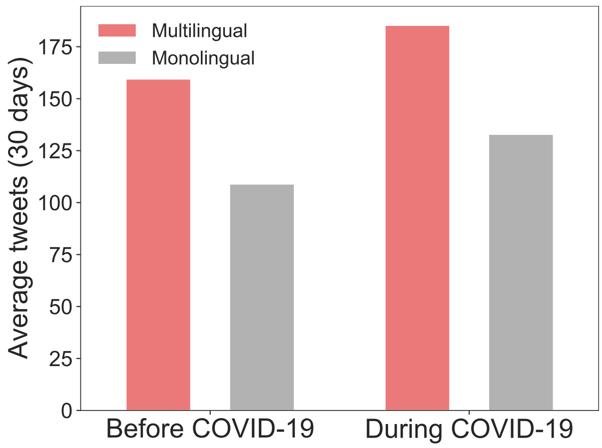


Figure 5.5: Average number of monthly tweets per user.

In Figure 5.5, we show the average number of monthly tweets posted by a user before and during the pandemic. We can see that users became more active on Twitter after the outbreak of the pandemic and multilingual users are more willing to participate in the online discourse. These observations are consistent with existing studies [Hal14], implying the reliability of our multilingualism definition.

Experimental evaluation. We use the same measurements evaluating UBM in Section 5.4.2 to quantify the speed and popularity of COVID-19 related information diffusion after the retweeting of both subgroups. The results are summarised in Table 5.3. We can see that on average, every minute the retweets of multilingual users can activate about 2.7 times as many users as monolingual users. Moreover, each multilingual user can activate 3.84 users, which is 50% more than monolingual users. Although multilingual users account for only 37% of the collected users, they can impact in total almost 92% of the users as a subgroup, which is about 10% more than monolingual users.

From the above discussion, we can see that multilingual users actually play a more important bridging role in diffusing COVID-19 related information during the pandemic. This also reconfirms the similar findings of the existing studies which highlight the bridging role of multilingual users in network connectivity [Hal14] and

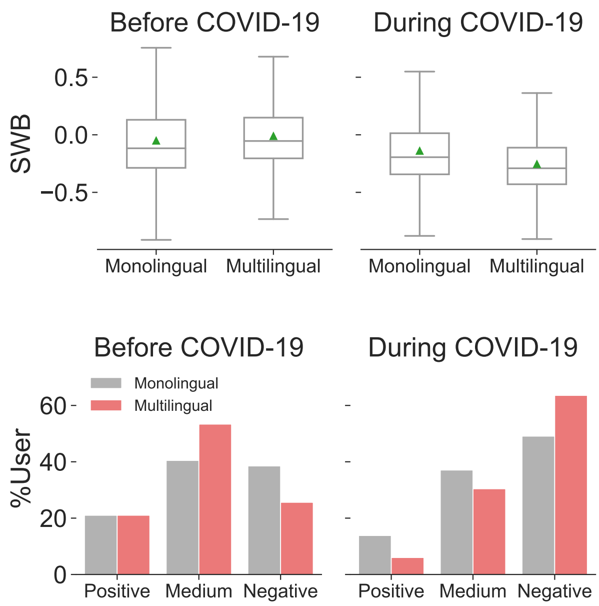


Figure 5.6: Distribution of SWB values before and during COVID-19.

between communities speaking different languages [EG12]. With regard to information diffusion, it has been studied that non-native English speakers have a higher influence than native English users [KWWO14] and multilingual users play a special role in cross-lingual diffusion[AGJ$^+$20].

Both the experimental evaluation and the literature illustrate again the validity of our dataset. In the following, we will check whether our SBM measurement can capture this special role of multilingual users in our dataset.

Table 5.3: Multilingual and monolingual users bridging performance comparison.

	Monolingual	Multilingual
Avg. #activated user/minute	0.021	0.057
Avg. #activated users	2.66	3.84
%impacted user	83.11	91.69

In Table 5.4, we list the results about the SBM values calculated with our GR-ego dataset. Note that the third column depicts the number of cascades in which the corresponding subgroup has a larger bridging value than the other subgroup. In our analysis, we instantiate the integration function γ with *maximum*, *median* and *mean*. The obvious observation from Table 5.4 is that according to SBM, multilingual users have a cascade bridging magnitude of 0.78 on average under all the three integration functions, which is more than two times larger than that of monolingual users. The above analysis shows that our SBM measurement can successfully capture the specific bridging role of multilingual users.

Table 5.4: The SBM of multilingual and monolingual users: multilingual users perform dominantly better with respect to all the three integration functions.

Operation	**User group**	**#cascade**	**SBM ρ**
Maximum	multilingual	20,862	**0.7843**
	monolingual	5,739	0.2011
Median	multilingual	20,843	**0.7835**
	monolingual	5,758	0.1994
Mean	multilingual	20,849	**0.7838**
	monolingual	5,752	0.2013

5.5.3 Analysing SWB Changes of User Subgroups

The correlation discovered in the previous section implies that multilingual users would suffer lower SWB due to their bridging role in diffusing COVID-19 related information. We proceed to validate this inference by comparing the SWB of multilingual users before and after the pandemic outbreak.

Figure 5.6 presents the SWB values of multilingual and monolingual users before and after the onset of the pandemic. From the box plot on the left, we can see that multilingual users on average behaved more positively than monolingual users before the pandemic. This is consistent with previous studies conducted in different language regions that multilingualism is generally associated with better subjective well-being [PA08, Tra95]. Similar to recent COVID-19 related works [HLKX20], we can observe that the outbreak of the pandemic lowers the SWBs of all users but multilingual users' SWB dropped more drastically and became even lower than monolingual users' SWB. The right part of Figure 5.6 shows the distribution of users in each SWB category. We clearly see that 25.8% of the multilingual users are consistently negative before the pandemic, which is about one third less than that of monolingual users (37.7%). During the pandemic, in both the multilingual and monolingual groups, a large number of users changed from the positive category to the negative category due to the adverse impact of the pandemic. Positive multilingual users decrease from 20% to 5.4% while positive monolingual users drop to 12.8%. Negative multilingual users increase by 200% and negative monolingual users just increase by 26%.

Discussion. From the above analysis, we conclude that the outbreak of the pandemic imposes a more adverse impact on multilingual users and they reacted more negatively than monolingual users. This finding complements the existing pre-pandemic studies. It is claimed that monolingual users suffer from a lower SWB in an international environment due to their language barriers. Our results show that during a global pandemic like COVID-19, the influence of language barriers does not constitute a factor for SWB with the same level of importance as that before the pandemic.

5.6 Conclusion and Limitation

In this chapter, our focus is on the social characteristics of users using social media, particularly their sharing behaviour and performance in promoting the popularity of trustworthy information. By proposing two new measurements, we quantify the bridging performance of both individual users and subgroups of users. With Twitter data collected from an international region, we successfully show the influential users and subgroups suffer from more decrease in their subjective well-being. With our measurement on individual users, we conduct the first research to reveal the strong negative relationship between a user's bridging performance in information diffusion and his/her SWB during the pandemic. With our bridging performance measurement on subgroups, we re-confirm the bridging role of multilingual users in diffusing information on social media and illustrate the negative relation between multilingual users' SWB and their bridging performance. This finding complements existing studies about multilingual people's subjective well-being before the pandemic in the sense that the impact of language barriers on SWB becomes less significant during a global health crisis like the COVID-19 pandemic. Our research provides a cautious reference to public health bodies that some individual users and subgroups can be mobilised to help spread health information, but special attention should be paid to their psychological health.

Limitations and our future work. This chapter has a few limitations that deserve further discussion. First, we only focused on the affective dimension of subjective well-being while noticing its multi-dimensional nature. This allows us to follow previous SWB studies to convert the calculation of SWB to sentiment analysis, but does not comprehensively evaluate users' cognitive well-being, such as life satisfaction. In our following research, we will attempt to leverage more advanced AI models to investigate cognitive aspects such as *happy* and *angry*. Second, extracting SWB from users' online disclosure inevitably incurs bias compared to social surveys although it supports analysis of an unprecedented large number of users. Third, socio-demographic information of users is not taken into account in this chapter. It is known that SWB varies among different socio-demographic groups, and such variation may have an impact on our results. Currently, deep learning based models exist for socio-demographic inference. In our future work, we will extract users' socio-demographic information such as age, gender and income to ascertain whether the regression results will change due to the variations of socio-demographic information. Last, we notice that the region we targeted at may introduce additional bias in our results. As a continuous work, we will extend our study to a region of multiple European countries and cross-validate our findings with other published results in social science.

Ethical considerations. This work is based completely on public data and does not contain private information of individuals. As we explained in Chapter 3, our dataset is built in accordance with the FAIR data principles [WDA+16] and Twitter Developer Agreement and Policy and related policies. Meanwhile, there have been a significant amount of studies on measuring users' subjective well-being through social media data. It has become a consensus that following the terms of service of social media networks is adequate to respect users' privacy in research [FLŞ+20]. To conclude, we have no ethical violation in the collection and interpretation of data in our study.

Part III

Information Diffusion Prediction

Chapter 6

Exploring Spillover Effects

As emphasised in the first chapter, this thesis is focused on exploring three social media-based responses to combat the infodemic. In the previous chapter, we focus on the role of influential users in diffusing reliable information on platforms like Twitter. As we move into this chapter, our focus shifts to another essential element of combating the infodemic: controlling the diffusion of misinformation and fake news. We aim to answer the research question: How can we accurately predict the popularity of a piece of information? We investigate the spillover effects of information exposure on users' decisions to participate in the diffusion of specific information, concentrating on COVID-19 preventive measures. By employing the dataset we gathered, we confirm the presence of these spillover effects and examine their potential influence in the context of misinformation management. Building on this understanding, we propose enhancements to three cascade prediction methods based on GNNs to optimise the prediction of information diffusion dynamics. Our experiments demonstrate that integrating the identified spillover effects substantially improves the performance of cutting-edge GNNs methods in predicting the popularity of both preventive measure messages and other COVID-19-related information.

6.1 Introduction

In the combat against the infodemic, alongside fostering the diffusion of trustworthy information, another widely recognised approach is known as *cascade prediction*. The goal of cascade prediction is to gauge the popularity of messages based on their early adopters. Precise prediction can aid the public in identifying information that merits special attention and help healthcare departments pinpoint misinformation that necessitates swift response to mitigate its negative impact. Research on cascade prediction has been sustained, with a large number of prediction models developed. Earlier models rely on hand-crafted features extracted from demographic profiles of early adopters [CAD+14, CJY+13] or social graphs composed of early adopters and their relationships [Moo97]. The recent advances of deep learning lead to models that can automatically learn useful features, encoded as a low-dimensional representation of available evidences that can be intuitively interpreted as most related features [BLG16, GPG+17]. In particular, the application of GNNs allows for capturing the features of nodes' neighbourhoods and simulating information cascading over social networks [CSG+20].

In spite of the various diffusion patterns exploited, previous studies have not considered the *spillover effect* of a user's exposed information on his/her behaviour of forwarding certain types of messages and becoming part of their diffusion, which we call *info-exposure spillover effect* for short. Spillover effects have become a commonly adopted theory in studying the impact of certain information on the opinions and behaviour changes of information consumers. For example, studies of political attitudes have found that exposure to scandals about some candidates may have negative spillover effects on the public's trust in other politicians [vSH20, Lee18]. We say a user is *exposed* to a message if the user posts the message or perceives it from his/her friends on social media. In this chapter, we adopt the original definition of behaviour spillover effect which intuitively means *"the observable and causal effect that a change in one behaviour has on a different, subsequent behaviour"* [GW19]. According to this definition, the info-exposure spillover effect studied in this chapter can be interpreted as the impact of the information a user perceived from the social media on his/her behaviour of forwarding a COVID-19 related post received from his/her friends.

We hypothesise the existence of this info-exposure spillover effect according to the previous studies related to the COVID-19 pandemic. Park et al. [PPC20] demonstrated that information with medically oriented thematic framework has a wider spillover effect on COVID-19 issues in a Twitter context. Racist information is found to have spillover effects on the mistrust of medical system [AMH20] and thus harm public trust in the information released by these systems.

In this chapter, we focus on the diffusion of messages related to COVID-19 preventive measures considering their importance in slowing down virus transmission and protecting public health. After the outbreak of the pandemic, the topics of information to which social media users are exposed have experienced subtle changes. Some of these changes may subsequently lead to the changes of their intention to forward messages concerning preventive measures. For example, tweets about unemployment or loneliness may make a user who reads them perceive the severity of the pandemic and thus becomes more likely to retweet tweets about staying at home.

With the GR-geo dataset, we successfully validated the existence of the info-exposure spillover effect of users' exposed messages on their decisions to retweet messages related to preventive measures. Specifically, we take into account all the messages exposed to users, regardless of whether they were related to COVID-19 or not. We observed that although all messages present certain a level of spillover effects on retweeting preventive messages, those related to COVID-19 have stronger impacts. This motivates us to extend existing state-of-the-art cascade prediction models by taking into account info-exposure spillover effects. Through comprehensive experimental evaluation on our dataset, we show that our extended models can increase the cascade prediction performance up to 23.84% in COVID-19 messages related to preventive measures. In order to attest whether info-exposure spillover effects also exist for other messages, we also run the extended models to predict the size of cascades of general messages concerning COVID-19 but not related to preventive measures. The results show an obvious increase in accuracy due to the use of the info-exposure spillover effect.

6.2 Related Work

Cascade prediction. Cascade prediction becomes attractive after studies reveal that some key properties of information cascades can be predicted [CAD+14, YCW+15]. In general, the cascade prediction methods can be divided into two classes: macro-level prediction and micro-level prediction. Micro-level prediction aims to predict users who will be activated during the information diffusion, while macro-level cascade prediction directly calculates the final size of targeted cascades.

The idea of most micro-level methods are based on the Independent Cascade model (IC) [KKT03], which calculates the probability of influence between every pair of users [GLK12]. These methods rely on a number of assumptions that overly simplify the real situation such as the complete observation of diffusion processes [ML10]. Although Deepinf [QTM+18] uses an end-to-end deep learning method to overcome such assumptions, micro-level methods generally do not perform well in predicting cascade future size as they require simulating the entire diffusion process. In this paper, as our target is popularity prediction, we opt for macro-level methods.

Macro-level prediction methods can be divided into three categories as a result of technological evolution, i.e., statistical prediction model, machine learning-based methods and deep learning-based methods. The development of macro-level prediction started with statistical models like SEISMIC [ZEH+15] and Weibull [YCW+15]. Then, the advancements of machine learning led to methods using manually designed features extracted from text content, temporal and demographic information, and network structure [YCW+15, CAD+14, CJY+13]. Deep learning-based methods overcome the deficiency of machine learning-based methods of constructing manual features and capture effective features automatically. DeepCas [LMGM17] and DeepHawkes [CSC+17] use Recurrent Neural Networks (RNNs) to capture cascading sequences in place of manually designed features. However, RNNs are limited in capturing structural information. This limitation is addressed by graph neural networks (GNNs) [SGT+08]. Intuitively, GNNs update the representation of each node by recursively aggregating the representations of its neighbours. In this way, the iterated node representation summarises both structural and representation information in neighbourhoods. CasCN [CZZ+19] utilises a dynamic Graph Convolutional Network (GCN) to learn the structural information of the cascade. CoupledGNN [CSG+20] (CGNN) effectively addresses cascade prediction with two GNNs, capturing the cascading effect which indicates that the activation of one user will successively trigger its neighbours.

Although deep learning-based methods have achieved relatively good results in cascade prediction, little research has been conducted to incorporate textual content into cascade prediction. Users' textual posts, as an important part of social media, may contain information that are related to users' decision to participate in diffusion of certain messages. Thus, we narrow the focus in this article to macro-level cascade prediction by extending the existing models to leverage online textual content on social media.

Spillover effects. The spillover effect has been widely used to study the impact of information on the information consumers' opinion and behaviour [KK15, Sch11, Lee18, vSH20]. Spillover effects can be interpreted and explained in various ways. We identify two main typologies in the literature, namely *behavioural spillover effects* and *affective spillover effects*.

The former interprets spillover effects as implicit ideas people build up that two things are connected, regardless of whether they are in the same context or across different contexts [Lee18]. For instance, Sikorski explained the damage of the public's trust in politicians following scandals of candidates as a behavioural spillover effect [vSH20]. Other examples include the impact of religious activities on political orientation [Pet92], and imposition of extra congestion charges on environmental behaviour changes in situations irrelevant to traffic [KK15]. The latter studies how affective responses (e.g., emotions such as happiness and anger) triggered by certain information affect human behaviour, usually based on the 'feelings-as-information' model [Sch11], Schwarz et al. found that anger triggered by other information may have negative effects on people's political attitudes [SB07]. Yegiyan discovered that the emotional feelings caused by film clips shown before commercial advertisements may affect audience's product preference [Yeg15].

Based on these previous studies, we make our hypothesis that during the COVID-19 pandemic the information exposed to an individual may have spillover impacts on his/her behaviour of retweeting messages. In our validation (see Section 6.4), we consider both behavioural or affective spillover effects. To capture our info-exposure spillover effect, we do not explicitly distinguish these two typologies and profit from the power of deep learning to automatically learn the features of exposed messages that have spillover effects.

6.3 Preliminaries

6.3.1 Problem Definition

In this section, we give the formal definition of the popularity prediction problem studied in this chapter which takes into account both social relations and online textual contents.

We use graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to denote the social network where $\mathcal{V}$ is the set of nodes representing users and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges indicating the relationships between users. For each $v \in \mathcal{V}$, given a time period, we use $\mathcal{M}_v$ to denote the messages posted by the user corresponding to v, and $\mathcal{M}$ to denote the set of all messages, i.e., $\mathcal{M} = \cup_{v \in \mathcal{V}} \mathcal{M}_v$. In the rest of the chapter, we will misuse the notions of users and nodes whenever it is clear from the context.

Table 6.1: Keyword lists for filtering tweets related to preventive measures and selected topics.

Topic	Abbre.	Keywords
Preventive measure	PM	stayathome, mask, masque, maske, wash hand, social distancing, socialdistancing, staysafe, lockdown
Unemployment	U	job, jobsearch, unemployment, employment, career, resume, recruitment, recession, economy, economic emploi, stelle, employ, arbeitslos, chômeurs
Panic buying	PB	panicbuying, panicshopping, panicbuyers, toiletpaper, handsanitizer, coronashopping
School closures	SC	schoolclos, closenypublicschool, closenycschools, suny, cuny, homeschool, noschool, shutdownschools
Stop Asian hate	SAH	stopasianhate, stopaapihate, stopasianhatecrimes, asian, aapi, asianlivesmatter, asiansareguman, antiasianhate
Black life matters	BLM	blacklifematters, blacklivesmatter, atlantaprotest, blm, changethesystem, justiceforgeorgefloyd
Loneliness	L	lonely, loneliness, alone, solitaire, solitude, seul, einsam, einsamkeit, allein

When a message m is firstly posted by a user, it will be perceived by the user's followers who might adopt the message and relay the message. This cascading process will continue on the social network until no further sharing occurs. We denote the observed diffusion cascade of m at time t by $C_m^t = \{v_1, v_2, \ldots, v_{n_t^m}\}$ ($v_i \in \mathcal{V}$ for $1 \leq i \leq n_t^m$), i.e., the set of users who had adopted m before t. Note that n_t^m is the number of the adopters of m at t. Compared to the previous works, we take into account the online textual messages posted by users in addition to the social network. This leads to the following definition.

Definition 8 (Online textual content-aware cascade prediction)**.** *Given the cascade of message m at time t (i.e., C_m^t), social network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the messages posted by users in $\mathcal{V}$, i.e., $\forall_{v \in \mathcal{V}} \mathcal{M}_v$, the problem is to predict the final popularity of m at time ∞, i.e., n_∞^m.*

As mentioned previously, we focus on the diffusion of the messages related to COVID-19 preventive measures although we will also show the effectiveness of our extended models in predicting the popularity of other general messages.

To integrate the online textual messages, i.e., $\mathcal{M}$, in solving the problem, we will make use of the info-exposure spillover effects of messages exposed to users on their decision to relay preventive measure-related messages.

6.4 Spillover Effects in COVID-19 Preventive Measure Information Diffusion

In this section, we validate our hypothesis that the information exposed to a user has spillover effects on his/her behaviour of retweeting a message related to COVID-19 preventive measures. We first briefly describe the measurement used for quantifying the hypothesised info-exposure spillover effect. Then we give the detailed experimental analysis designed to validate its existence in the diffusion of COVID-19 preventive measure-related messages.

6.4.1 Data Pre-processing

We concentrated on the initial wave of the pandemic, choosing from the GR-geo dataset, which amassed 18,523,099 tweets from all users in the social graph between 22nd January and 18th July 2020. We divide the tweets into COVID-19 related and COVID-19 unrelated based on the keywords provided by Chen et al. [CLF20]. In our collected tweets, the COVID-19 related tweets account for 26.19%.

We construct cascades from our tweet dataset and the social graph built previously based on the definition in Section 6.3.1. A total of 7,485,895 cascades are built and we remove those cascades with fewer than 3 users, the same as the existing works [LMGM17, CSG$^+$20]. Eventually, 89.14% of the cascades are kept and we end up with 6,672,926 cascades. The average size of these cascades is 4.31. We use $\mathcal{C}$ to denote the set of all the selected cascades. From $\mathcal{C}$, we construct the set of cascades corresponding to messages related to preventive measures, denoted by $\mathcal{C}_{PM}$, based on the keywords listed in

6.4.2 Measuring Info-exposure Spillover Effect

We design our validation based on the experimental investigation method commonly used for spillover effect validation [CLCR$^+$19, vSH20]. The idea is to investigate whether users exposed to different information will behave differently in retweeting a message related to preventive measures. In other words, we will check whether certain exposed information will change the likelihood that users retweet messages related to preventive measures.

Info-exposure spillover effect validation framework. We construct groups of users according to the information they are exposed to. Each group is composed of users who are exposed to a certain composition of information. One of these groups is set as the control group. The selection of the control group depends on the purpose of the experiment. The proportion of users in each group retweeting preventive measure messages is used to measure the likelihood of adopting preventive measure messages, which we call the *adoption likelihood*. By comparing the measurement of a group with that of the control group, we can then quantitatively evaluate the magnitude of the info-exposure spillover effect of the information exposed to this user group on adopting preventive measure messages, which we call the *info-exposure spillover elasticity*.

Formally, Let $\mathcal{D}$ be a set of groups of nodes in $\mathcal{G}$, i.e., $\mathcal{D} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ where $\forall_{\mathcal{V}_i \in \mathcal{D}} \mathcal{V}_i \subset \mathcal{V}$. Suppose $\mathcal{V}_c \in \mathcal{D}$ be the selected control group. For each user group $\mathcal{V}_i \in \mathcal{D}$, we identify the users who ever retweeted at least one preventive measure message in $\mathcal{M}_{PM}$, and then construct the set of identified users $\mathcal{V}_i^{PM}$. The *adoption likelihood* for users in $\mathcal{V}_i$ is calculated as

$$\alpha_{\mathcal{V}_i} = \frac{\mid \mathcal{V}_i^{PM} \mid}{\mid \mathcal{V}_i \mid}.$$

With these notations, we can define the info-exposure spillover elasticity as follows:

Table 6.2: Validation of info-exposure spillover effect of single topics.

Topic type	**Topic**	**Exposed**		**Unexposed**		**Elasticity** ε
		#user	α	**#user**	α	
COVID related	Unemployment (U)	4,238	0.67	17,101	0.25	1.69
	Panic buying (PB)	6,119	0.39	15,220	0.31	0.25
	School closures (SC)	6,460	0.61	14,879	0.21	1.87
COVID unrelated	Stop Asian hate (SAH)	6,740	0.72	14,599	0.28	1.53
	Black life matters (BLM)	9,041	0.48	122,98	0.41	0.16
	Loneliness (L)	5,343	0.79	15,996	0.30	1.63

Definition 9 (Info-exposure spillover elasticity)**.** *The* elasticity *of the info-exposure spillover effect of a user group $\mathcal{V}_i$ in the user group set $\mathcal{D}$ is calculated as*

$$\varepsilon^{\mathcal{D}}_{\mathcal{V}_i} = \frac{\alpha_{\mathcal{V}_i} - \alpha_{\mathcal{V}_c}}{\alpha_{\mathcal{V}_c}}.$$

Positive elasticity indicates the information commonly exposed to the users in $\mathcal{V}_i$ increases the likelihood of retweeting a preventive measure message while negative elasticity indicates the opposite.

6.4.3 Experimental Validation of Info-exposure Spillover Effect

We verify through our collected data that being exposed to certain information may affect users' behaviour of retweeting messages related to preventive measures. It is not tractable to analyse all the contents that are mentioned or discussed in tweets. Therefore, inspired by previous research [MU20, SSR20], we classify tweets from the level of topics and select six frequently studied ones in the literature [MU20, SSR20] as the representatives. Among these topics, three are related to COVID-19, i.e., *Unemployment*, *Panic buying* and *School closures*, while the other three studied in previous Twitter-based studies are general and not directly related to the pandemic, namely, *Stop Asian hate*, *Black life matters* and *Loneliness* [KL20, KCT+20]. We extract corresponding tweets in each topic with the keywords listed in Table 6.1. According to our manual check, the keywords ensure a good coverage rate of the tweets in the selected topics. In total, the messages covered by these topics take up 18.17% of our collected tweets excluding those related to preventive measure.

For the purpose of being comprehensive, we conduct our experimental validation from two perspectives. We first evaluate the spillover effect of messages of a single topic on the behaviour of retweeting a preventive measure message. Second, we investigate the spillover effect of messages in various compositions of topics.

Spillover effects of information of single topic. We build six sets of user groups each of which corresponds to a selected topic, i.e., $\mathcal{D}_{\mathsf{U}}$, $\mathcal{D}_{\mathsf{PB}}$, $\mathcal{D}_{\mathsf{SC}}$, $\mathcal{D}_{\mathsf{SAH}}$, $\mathcal{D}_{\mathsf{BLM}}$, $\mathcal{D}_{\mathsf{L}}$. Each set has only two groups. One consists of users that have been exposed to messages of the corresponding topic while the other group is composed of users who have not been exposed. We will take the one unexposed to the topic as the control group. In Table 6.2, we show the number of users exposed and unexposed in each group set, the adoption likelihood and the final info-exposure spillover effect elasticity.

We have three main observations. First, the exposure to each topic of messages will increase the likelihood of users to retweet a preventive measure message. On average, the adoption likelihood of exposed groups equals to 0.58 while the unexposed group only has an activation likelihood of 0.28. The average elasticity is 1.19, which indicates that the activation likelihood doubles for the users exposed to the topics on average. Second, the increase of adoption likelihood for exposed users differs among the topics of exposed information. For instance, the exposure to information related to *Panic buying* and *Black life matters* just increases the elasticity by 0.25 and 0.16, respectively, which are much smaller than the other topics. We manually examine messages in the topic *Black life matters* and *Stop Asian hate* to understand the difference. We notice that users exposed to the messages about racist have more diverse attitudes towards prevention measures. This is consistent with previous studies [EAC+20]. For example, some users argue that the protest breaks the social distancing policy and exacerbates the virus transmission, while some others hold the view that the impact of COVID-19 is overstated and the lockdown policy worsens racial discrimination. The above two observations apply in both COVID related topics and COVID unrelated topics. Third, exposure to messages unrelated to COVID imposes weaker spillover effect than those related to COVID. On average, the average elasticity of the COVID-19 unrelated topics is 14.76% smaller than that of the COVID-19 related topics.

From the above analysis, we can conclude that i) exposure to certain topics of information, regardless of whether they are related to COVID-19, will impose positive spillover effects on users' likelihood to retweet preventive measure messages; ii) the scale of spillover effect differs according to the topics of exposed messages.

Spillover effects of information of compositions of topics. In the previous analysis, we focus on the spillover effect of single topics and ignore the changes when multiple topics of information are exposed to users simultaneously. We construct a user group set $\mathcal{D}_{comp}$ of 22 groups, Of which 15 groups correspond to the users who are only exposed to messages of every pair of the 6 topics, 6 are composed of users only exposed to tweets of one of the selected topics. The last group contains the users exposed to no messages in all the topics and is chosen as the control group. Note that we do not consider the compositions of more than 2 topics in $\mathcal{D}_{comp}$ because we observe in our analysis that exposure to messages of any three topics leads to an adoption likelihood of at least 0.79. This indicates the improvement of info-spillover effect will be marginal when users are exposed to messages of more topics.

Figure 6.1 shows the adoption likelihood of user groups exposed to the topic compositions in $\mathcal{D}_{comp}$ except for the control group. We can see that exposure to more selected topics increases the likelihood of retweeting a preventive measure message. Exposure to an additional topic significantly increases the adoption likelihood. The most significant increase occurs to the topic of *Panic buying*. The addition of any other topic except for the topic BLM increases the adoption likelihood by at least two times. When exposed to none of the topics, the activation likelihood for the

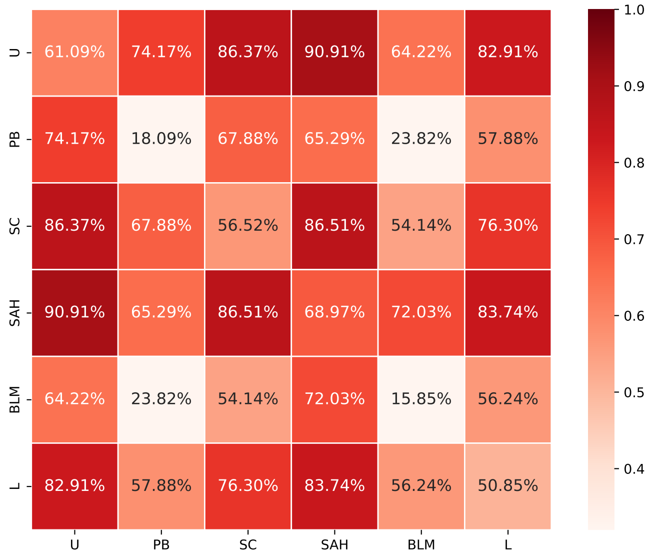


Figure 6.1: Activation likelihood when exposed to compositions of topics.

users drops below 5%.

Discussion From the above analysis, we empirically validated the existence of the info-exposure spillover effects. Specifically, certain information exposed to users indeed increases the likelihood of users to retweet preventive measure messages. In addition, we also illustrated that the magnitudes of this spillover effect depend on the content of tweets exposed. In the following, we will leverage deep learning to automatically capture the contents of tweets exposed to users that impose strong info-exposure spillover effects, and thus improve the accuracy of cascade prediction.

6.5 Predicting Popularity with Spillover Effects

We use the framework of GNNs to learn the magnitudes of the info-exposure spillover effect of a user's exposed information on his/her behaviour of retweeting preventive measure messages. Recall that the information exposed to a user comes from two sources: the messages posted by their friends and his/her own posts. We need to combine these two sources in a specific manner and calculate an overall representation for each user that can be used in the following cascade prediction. This explains our selection of GNNs. When a user's past posts are encoded as a vector and attached to the corresponding node as node attributes, the message passing scheme of GNNs will conduct the combination. The combination may even involve the messages from users that are not incident but within a certain number of hops. In this section, we describe how we calculate nodes' attributes with the encoding of users' past posts, and then detail how we extend various GNN-based models to integrate the identified info-exposure spillover effect into cascade prediction.

6.5.1 Calculating Initial Node Attributes

Given a cascade of m at time t, i.e., C_m^t, we calculate the initial attribute of a node $v \in \mathcal{V}$, denoted by h_v^0, by concatenating the following three components:

1. the representation vector of the messages posted by the user before t, denoted by δ_v;
2. the activation status of the user according to the given cascade C_m^t, denoted by s_v;
3. the node embedding of the user's corresponding node in the network, denoted by e_v.

Formally, we have $h_v^0 = s_v \| \delta_v \| e_v$ where $\cdot \| \cdot$ is the concatenation operator.

Past message encoding δ_v. For each user v, we collect her/his past messages posted or retweeted before t. We have learnt in Section 6.4 that exposure to COVID related messages may impose stronger spillover effects than those unrelated to COVID. We distinguish these two types of information to capture the difference. For each type, we collect the last λ textual messages before t in $\mathcal{M}_v$, and thus construct two lists of messages ordered by their posting time, i.e., $(m_1^{rel}, m_2^{rel}, \ldots, m_\lambda^{rel})$ and $(m_1^{unrel}, m_2^{unrel}, \ldots, m_\lambda^{unrel})$ for the COVID related and unrelated, respectively. Note that λ is a pre-defined hyper-parameter that should be tuned manually. For embedding, RoBERTa [LOG+19] is a language pre-trained transformer to encode short texts in multiple languages into a vector of real numbers with a pre-defined length. In this chapter, we use a widely used multilingual pre-trained RoBERTa variant: XLM-RoBERTa [CKG+20]. For each message, we calculate its embedding with the default XLM-RoBERTa model and obtain the corresponding lists of message representation vectors. The resulted lists are represented as $(z_1^{rel}, z_2^{rel}, \ldots, z_\lambda^{rel})$ and $(z_1^{unrel}, z_2^{unrel}, \ldots, z_\lambda^{unrel})$.

Many methods exist to combine these embeddings and obtain δ_v while considering their relative temporal importance, e.g., Hawkes process and Gated Recurrent Unit (GRU). In this chapter, according to our experimental evaluation (see Section 6.6.4) we adopt the content-aware temporal encoding (TE) which assigns fixed importance to messages based on their temporal order. Formally,

$$\begin{aligned} \phi_v^{rel} &= \textstyle\sum_{i \leq \lambda} a_i \cdot z_i^{rel}, \\ \phi_v^{unrel} &= \textstyle\sum_{i \leq \lambda} a_i \cdot z_i^{unrel}. \end{aligned}$$

Note that the messages related to COVID and those unrelated share the same temporal importance settings. According to our manually probe, using two different importance settings does not give notable improvement, and increases the model complexity.

In order to capture the different contributions of messages related to COVID and those unrelated, we introduce a weight parameter ρ ($0 \leq \rho \leq 1$) and compute the integrated past message embedding δ_v as follows:

$$\delta_v = \phi_v^{rel} \cdot \rho + \phi_v^{unrel} \cdot (1 - \rho).$$

Table 6.3: Brief description of selected GNN variants.

Model	Aggregate(*)	Combine(*)
GCN	$\boldsymbol{a}_v^\ell = \frac{\sum_{u \in \mathcal{N}(v) \cup \{v\}} \boldsymbol{h}_u^{\ell-1}}{\lvert\mathcal{N}(v) \cup \{v\}}$	$\boldsymbol{h}_v^\ell = \mathsf{LeakyReLu}\left(\boldsymbol{W}^\ell \boldsymbol{a}_v^\ell\right)$
GAT	$\boldsymbol{a}_v^\ell = \sum_{u \in \mathcal{N}(v) \cup \{v\}} \boldsymbol{\beta}_{uv}^\ell \boldsymbol{h}_u^{\ell-1}$ $\boldsymbol{\beta}_{uv}^\ell = \frac{\exp\left(\mathsf{LeakyRelu}(\gamma^T[\boldsymbol{W}\boldsymbol{h}_u^{\ell-1} \Vert \boldsymbol{W}\boldsymbol{h}_v^{\ell-1}])\right)}{\sum_{u' \in \mathcal{N}(v) \cup \{v\}} \exp\left(\mathsf{LeakyRelu}(\gamma^T[\boldsymbol{W}\boldsymbol{h}_{u'}^{\ell-1} \Vert \boldsymbol{W}\boldsymbol{h}_v^{\ell-1}])\right)}$	$\boldsymbol{h}_v^\ell = \mathsf{LeakyReLu}\left(\boldsymbol{W}^\ell \boldsymbol{a}_v^\ell\right)$
CoupledGNN	$\boldsymbol{a}_v^\ell = \sum_{u \in \mathcal{N}(v)} \mathsf{InfluGate}\left(\boldsymbol{r}_u^{\ell-1}, \boldsymbol{r}_v^{\ell-1}\right) \boldsymbol{s}_u^{\ell-1} + p_v$ $\mathsf{influGate}\left(\boldsymbol{r}_u^\ell, \boldsymbol{r}_v^\ell\right) = \boldsymbol{\beta}^\ell \left[\boldsymbol{W}^\ell \boldsymbol{r}_u^\ell \Vert \boldsymbol{W}^\ell \boldsymbol{r}_v^\ell\right]$	$\boldsymbol{s}_v^{\ell+1} = \begin{cases} 1 & v \in C_m^T \\ \sigma(\boldsymbol{\mu}_s^\ell \boldsymbol{s}_v^\ell + \boldsymbol{\mu}_a^\ell \boldsymbol{a}_v^\ell) & v \notin C_m^T \end{cases}$

Activation status $\boldsymbol{s}_v$ & Node embedding $\boldsymbol{e}_v$. The user activation status s_v is set to 1 if $v \in C_m^t$ and 0, otherwise. The node embedding captures the structural properties of the user's neighbourhoods in the graph. Following existing studies [LMGM17, CSG+20], we use DeepWalk without further fine-tuning to learn the structural embedding for each user.

6.5.2 Instantiating GNNs with the Info-exposure Spillover Effect

We implement three variants of GNNs to integrate the info-exposure spillover effect we identified in the previous section, i.e., Graph Convolutional Networks (GCN) [KW17], Graph Attention Network [VCC+18] and CoupledGNN [CSG+20]. GCN is a semi-supervised learning algorithm for graph representation and GAT is a variant of GCN which introduces the attention mechanism to distinguish the significance of neighbours. These two variants are not designed specifically for cascade prediction. The calculated node representations are usually used for the downstream tasks such as link prediction and node classification. CoupledGNN [CSG+20] is a model developed for cascade prediction, and can stand for the state-of-the-art. It has overwhelming performance over existing models by simulating the cascading effect of information diffusion on social network, the phenomenon that users are activated due to the influence from their activated neighbours. By extending these models, our purpose is to illustrate the effectiveness of info-exposure spillover effects in improving the accuracy of the predicted popularity of COVID-19 preventive measure messages. In addition, our extended models can provide useful references for future cascade prediction models to integrate info-exposure spillover effects.

The definitions of the function $\mathsf{Aggregate}(*)$ and $\mathsf{Combine}(*)$ of GCN, GAT and CoupledGNN are briefly given in Table 6.3. GAT and GCN share the same combination function. For GCN, we use the mean of the representation vectors of both the nodes and their one-hop neighbours as the aggregated value at each layer while GAT uses the weighted average.

We describe CoupledGNN in more details due to its relatively large difference from the conventional GNN framework and explain how to simulate the cascading effect in information diffusion. For the full description, we refer the readers to the original paper [CSG+20]. It deploys two GNNs. One GNN captures the activation statuses of users during the information diffusion at each layer, e.g., the activation status of user v at the ℓ-th layer s_v^ℓ. The other GNN aims to simulate how the influence of users changes along with the activation status and the influences of their neighbours, i.e., r_u^ℓ. A neighbour u's influence to activate user v in the next layer $\ell+1$ is calculated by the function $\mathsf{influGate}(\boldsymbol{r}_u^\ell, \boldsymbol{r}_v^\ell)$. Then the aggregation function is the weighted average of all the neighbours' activation statuses with the default activation probability p_v added. The combination function is based on the weighted average of its status on the previous layer and the aggregated representation. With the activation status output by the last layer (e.g., k), the popularity of the message diffused in C_t^m is calculated as $\tilde{n}_\infty^m = \sum_{v\in\mathcal{V}} s_v$. In the following, we will describe how we extend each selected model to capture the info-exposure spillover effect.

SE-GCN & SE-GAT. We can interpret the output of the k-th layer of a k-layered GCN or GAT as the summary of the information exposed to every user. Then we use an activation function to capture the info-exposure spillover effect. Specifically, the function takes as input the output of the GCN or GAT and the representation of the message diffused in the given cascade, and outputs the predicted final activation statuses of the nodes. Let m be the message being diffused and z_m be the embedding vector of m calculated by the RoBERTa model. Let $\tilde{s}_v^\infty$ be the predicted activation status of node v. Our activation function is defined as:

$$\tilde{s}_v^\infty = \begin{cases} \mathsf{activate}\left(\boldsymbol{W_h h_v^k} \| \boldsymbol{W_z z_m}\right) & v \notin C_m^t \\ 1 & v \in C_m^t \end{cases}$$

where function $\mathsf{activate}$ is implemented as a 3-layer neural network in this chapter and $\boldsymbol{W_h}$ and $\boldsymbol{W_z}$ are two matrices to be learned. We add this function as a downstream component after the last layer of the GCN and GAT.

SE-CGNN. Recall that CoupledGNN uses the function $\mathsf{InfluGate}$ to simulate the process of a user to be activated by their neighbours. The influence vector, e.g., r_u of user u, contains user u's posted messages and the messages from u's neighbourhood. Therefore, it can be considered as a summary of the information perceived by a user v from u if v follows u in Twitter. Based on this intuition, we extend CoupledGNN by reformulating the function $\mathsf{InfluGate}(*)$ to capture the the info-exposure spillover effect:

$$\mathsf{influGate}\left(\boldsymbol{r}_u^\ell, \boldsymbol{r}_v^\ell\right) = \boldsymbol{\beta}^\ell \left[\boldsymbol{W}^\ell \boldsymbol{r}_u^\ell \| \boldsymbol{W}^\ell \boldsymbol{r}_v^\ell \| \boldsymbol{W_z z_m}\right].$$

6.5.3 Objective Function

We use the same objective function as [CSG+20] which is the mean relative square error (MRSE). Let $\mathcal{M}_\mathcal{C}$ be the set of diffused messages corresponding to the cascades in $\mathcal{C}$ whose final sizes are to be predicted. Then MRSE can be defined as follows:

$$L_{MRSE} = \frac{1}{|\mathcal{M}_\mathcal{C}|} \sum_{m\in\mathcal{M}_\mathcal{C}} \left(\frac{\tilde{n}_\infty^m - n_\infty^m}{n_\infty^m}\right)^2$$

This loss function is regularised to avoid over-fitting and accelerate the convergence speed, i.e., $L = L_{MRSE} + L_{Reg}$ where $L_{Reg} = \theta \sum_{p \in \mathcal{P}} \|p\|_2 + \lambda L_{user}$. Note that $\mathcal{P}$ denotes the set of parameters and L_{user} is the cross-entropy

$$\frac{1}{|\mathcal{M}_C|} \sum_{m \in \mathcal{M}_C} \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \left(s^{\infty}_{v,m} \log \tilde{s}^{\infty}_{v,m} + \left(1 - s^{\infty}_{v,m}\right) \log \left(1 - \tilde{s}^{\infty}_{v,m}\right) \right)$$

where $s^{\infty}_{v,m}$ is the final activation status of v in the cascade of message m and $\tilde{s}^{\infty}_{v,m}$ is v's status predicted by the model under evaluation.

6.5.4 Computational Complexity

In general, all our extended models inherent the complexity of the original models. According to a recent survey, the theoretical computation complexity of the message passing schemes such as GCN [KW17] and GAT [VCC+18] is $\mathcal{O}(|\mathcal{E}|)$ [WPC+21] where $|\mathcal{E}|$ is the number of edges of the graph $\mathcal{G}$. This is because, in these methods, the computation of each node v's representation involves messages that come from its adjacent nodes. The models that based on GCN and GAT, proposed previously, i.e., SE-GCN and SE-GAT, also work in the same way, and thus have the complexity of $\mathcal{O}(|\mathcal{E}|)$. Similarly, SE-CGNN, have the same computational complexity as CGNN, i.e., $\mathcal{O}(p|\mathcal{V}| + q|\mathcal{E}|)$ [CSG+20], where p and q are the constants determined by the batch sizes, and $|\mathcal{V}|$ is the number of nodes in $\mathcal{G}$.

6.6 Experimental Evaluation

6.6.1 Evaluation Measurements

We adopt the measurements in [CSG+20] to evaluate and compare the prediction performance of our extended models and the bench-markings models in our experiments. Specifically, in addition to the mean relative square error (MRSE) introduced in the previous section, we also use mean absolute percentage error (MAPE) and wrong percentage error (WroPerc). MAPE measures the average deviation between the predicted popularity and the true values, while WroPerc measures the percentage of cascades that are incorrectly predicted with a given error tolerance ϵ. Formally, they can be defined as follows:

$$\begin{aligned} MAPE &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{|\tilde{n}^m_\infty - n^m_\infty|}{n^m_\infty}, \\ WroPerc &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{I}\left[\frac{|\tilde{n}^m_\infty - n^m_\infty|}{n^m_\infty} \geq \varepsilon \right]. \end{aligned}$$

Note that $\mathbb{I}(*)$ is an indication function which outputs 1 when the input proposition is true or 0 otherwise, and the threshold ε is set as 0.5 in our experiments. For all the three measurements, smaller values indicate better performance.

6.6.2 Baseline Methods

In addition to CoupledGNN, we use the following models as baselines.

Feature-based method. This is a linear regression model with L2 regularisation with features. For better comparison, we adopt the same features used in the past studies [CSG+20, LMGM17].

SEISMIC [ZEH+15]. SEISMIC uses the Hawkes self-activation point process to estimate or approximate the impact of cascading effect with their average number of followers.

DeepCas [LMGM17]. DeepCas is an end-to-end deep learning method for information cascades prediction. It utilises the structure of the cascade graphs for prediction. An attention mechanism is designed to assemble a cascade graph representation from a set of random walk paths.

DeepHawkes [CSC+17]. DeepHawkes is also an end-to-end deep learning method for information cascades prediction. It combines user embedding vectors and cascades encoding by RNNs, and then uses the Hawkes process to model and predict information cascade.

CasCN [CZZ+19]. CasCN for cascade modelling and prediction is achieved by splitting the cascade graph into a series of sequential sub-cascades and then employing GCN to learn the structural information of the cascades.

GCN and GAT. We construct these two models from our SE-GCN and SE-GAT models by removing the representation vectors of messages. In other words, these two models only rely on network structure to predict the sizes of final cascades.

We implement several variants of our extended models, i.e., SE-GCN, SE-GAT and SE-CGNN according to the methods used to integrate users' past messages with their temporal significance considered. We considers three other methods in addition to the TE methods adopted in our model, namely, Mean, Hawkes and GRU. Note that regarding Hawkes and GRU, we use their basic versions. The method Mean calculates the average embedding vectors of the past messages for both ϕ_v^{rel} and ϕ_v^{unrel}. In order to distinguish these variants, we append the corresponding methods at the end of the model names. For instance, SE-CGNN-TE corresponds to the implementation of the model presented in Section 6.5, and SE-CGNN-Hawkes replaces the TE method in SE-CGNN-TE with the Hawkes process.

6.6.3 Implementation Details

As the output of the RoBERTa for a sentence is a high-dimensional and sparse vector. We apply linear transformation to map its output to a relatively low-dimensional space. The dimension of the final text embedding used is set as 128. For all models including the bench-marking models, we tune their hyper-parameters to guarantee their performance on validation sets. The L2-coefficients are chosen from $0.5, 0.1, 0.05, \ldots, 10^{-8}$. For all neural network models, the learning rate is chosen from $0.1, 0.05, \ldots, 10^{-5}$. The coefficient in the loss function is set to be 0.5, and the mini-batch size is chosen from 15, 10, 5. The number of GNN layers k is selected from 5, 4, 3, 2. As for DeepCas, the number of walk sequences and the walk length are set as 100 and 8, respectively. For SEISMIC, we follow the parameters from the original study. Moreover, we randomly select 80%, 10%, 10% of the set of cascade instances for training, validation and testing, respectively.

Considering the diffusion time of the messages in our collected data, we set the observation time window as 3 hours and construct the set of observed cascades, i.e., $\mathcal{C}$, by removing users in our cascades that were activated after the first 3 hours. The number of past messages λ is critical in enforcing the quality of prediction. As a result, we undertake an empirical investigation to identify the impact of λ on the final performance. We present the MRSE with different values of λ when the SE-CGNN-TE is used in Figure 6.2. We observe that λ does have an important impact on prediction results. We set λ as 3 with which our model achieves the best performance.

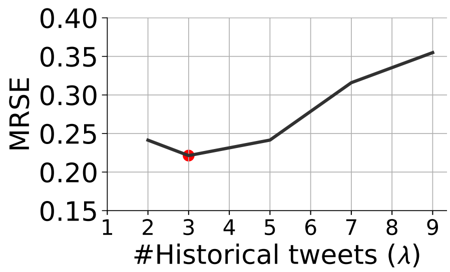


Figure 6.2: Parameter tuning for λ.

As we repeatedly emphasised, our original goal is to predict the popularity of messages on social media which are related to COVID preventive measures. In order to comprehensively evaluate the effectiveness of the info-exposure spillover effect, in addition to the cascades of preventive measure messages $\mathcal{C}_{PM}$, we also apply all the models on another two sets of cascades. One is the set of all cascades $\mathcal{C}$. The other is the set of cascades that are not related to preventive measures, i.e., $\overline{\mathcal{C}}_{PM} = \mathcal{C}/\mathcal{C}_{PM}$, the complement of $\mathcal{C}_{PM}$ in $\mathcal{C}$.

6.6.4 Experimental Results

We show the performance of all the above mentioned models in Table 6.4 in the form of the three selected measurements. In general, we can observe three obvious differences when the info-exposure spillover effect is introduced in cascade prediction.

First, compared to the original models, our extended models significantly improve their performance not only for the preventive measure messages, but also for all the three types of messages. The most significant improvement occurs to SE-CGNN-TE and reaches 23% in the WroPerc measurement for the preventive measure messages and over 10% for the messages unrelated to preventive measures. This is due to the fact that CoupledGNN simulates the cascading effects iteratively and this allows for applying the info-exposure spillover effect on activating individual users in a finer granularity. From the above analysis, we can conclude that the use of the info-exposure spillover effect can effectively improve the performance of existing cascade prediction models. It should be integrated into future models by design.

Table 6.4: Cascade prediction performance of our extended models and baselines.

Models	$\mathcal{C}$			$\mathcal{C}_{PM}$			$\overline{\mathcal{C}}_{PM}$		
	MRSE	**MAPE**	**WroPerc**	**MRSE**	**MAPE**	**WroPerc**	**MRSE**	**MAPE**	**WroPerc**
Feature-based	0.3611	0.4018	41.31%	0.4403	0.4049	46.08%	0.3704	0.4151	41.56%
SEISMIC	0.5580	0.5104	56.35%	0.5899	0.5265	55.88%	0.5419	0.5083	56.14%
DeepCas	0.2837	0.3959	37.71%	0.2847	0.3724	38.67%	0.2872	0.4010	37.31%
DeepHawkes	0.3278	0.4089	37.10%	0.3297	0.4092	37.94%	0.3213	0.3948	36.78%
CasCN	0.3097	0.4300	37.12%	0.3017	0.4166	40.39%	0.3098	0.4106	37.58%
GCN	0.3144	0.4217	38.88%	0.3179	0.4238	41.76%	0.3110	0.4200	38.69%
SE-GCN-Mean	0.2826	0.4056	36.76%	0.2757	0.3990	35.86%	0.2899	0.4178	36.82%
SE-GCN-Hawkes	0.2826	0.4056	36.76%	0.2708	0.3961	35.44%	0.2887	0.4126	36.89%
SE-GCN-GRU	0.2875	0.4085	36.87%	0.2712	0.3974	35.43%	0.2871	0.4124	36.92%
SE-GCN-TE	0.2802	0.4050	36.15%	0.2702	0.3932	35.20%	0.2819	0.4109	36.15%
GAT	0.3072	0.4211	39.19%	0.3014	0.4268	40.01%	0.3101	0.438	39.85%
SE-GAT-Mean	0.2862	0.4124	37.58%	0.2721	0.4001	35.31%	0.2903	0.4175	38.64%
SE-GAT-Hawkes	0.2790	0.4078	37.45%	0.2654	0.3986	35.30%	0.29353	0.4154	37.83%
SE-GAT-GRU	0.2762	0.4055	37.05%	0.2680	0.3964	35.58%	0.2961	0.4153	37.47%
SE-GAT-TE	0.2744	0.4014	37.56%	0.2673	0.3990	35.16%	0.2896	0.4177	38.06%
CoupledGNN	0.2678	0.3861	35.19%	0.2769	0.3920	34.44%	0.2601	0.3812	34.70%
SE-CGNN-Mean	0.2414	0.3610	34.17%	0.2587	0.3801	30.13%	0.2561	0.3608	33.22%
SE-CGNN-Hawkes	0.2240	0.3432	31.10%	0.2085	0.3171	27.44%	0.2271	0.3478	31.35%
SE-CGNN-GRU	0.2283	0.3469	32.28%	0.2174	0.3164	28.65%	0.2411	0.3625	33.04%
SE-CGNN-TE	**0.2131**	**0.3358**	**30.63%**	**0.2031**	**0.3073**	**27.78%**	**0.2262**	**0.3437**	**31.56%**

Second, we observe that the extended models can more accurately predict the popularity of COVID-19 preventive measure messages than the other messages, which is the opposite for the baseline models. For the baseline models, their performance on $\mathcal{C}$ and $\overline{\mathcal{C}}_{PM}$ are almost the same but becomes worse on $\mathcal{C}_{PM}$. The feature-based model has the worst performance which decreases by over 11% compared to that in predicting the sizes of the other two sets of cascades. However, when the identified info-exposure effect is used in our extended models, the popularity of preventive measure messages can be predicted with better accuracy. SE-CGNN-TE can improve the performance by about 13.8% for preventive measurement messages compared to those unrelated to preventive measures. This observation validated empirically that the exposure to information generated during the COVID-19 pandemic has strong spillover effects on retweeting messages about how to prevent the transmission of the COVID virus.

Third, the consideration of the temporal importance of past tweets does bring about further improvement, and our selected TE method overwhelms the other widely used ones. The method Mean which ignores the temporal significance of past messages produces the worst predictions. Hawkes and GRU have similar performances. Compared to them, our TE method leads to an improvement of about 0.02 in all three types of cascades.

6.6.5 Compare SE-CGNN-TE with its Variants

Recall that we distinguish the messages related to COVID-19 and those unrelated in integrating the embedding vectors of users' past messages into the initial node attributes (see Section 6.5.1). We use a parameter ρ to learn the relative importance of the message related to COVID-19. We conduct additional experiments to justify our selection. Specifically, we implement another three variants of our SE-CGNN-TE model. The first one, named by SE-CGNN-TE-REL, only takes the last

Table 6.5: The performance comparison of methods of past message integration.

Models	$\mathcal{C}$			$\mathcal{C}_{PM}$			$\overline{\mathcal{C}}_{PM}$		
	MRSE	**MAPE**	**WroPerc**	**MRSE**	**MAPE**	**WroPerc**	**MRSE**	**MAPE**	**WroPerc**
SE-CGNN-TE-REL	**0.2171**	**0.3367**	**32.23%**	**0.2210**	**0.3294**	**28.46%**	**0.2312**	**0.3526**	32.96%
SE-CGNN-TE-UNREL	0.2357	0.3572	33.83%	0.2442	0.3690	30.74%	0.2484	0.3567	33.01%
SE-CGNN-TE-ALL	0.2208	0.3406	32.53%	0.2318	0.3172	28.96%	0.2470	0.3534	**32.80%**
SE-CGNN-TE	**0.2146**	**0.3351**	**30.45%**	**0.2024**	**0.3062**	**27.41%**	**0.2268**	**0.3421**	**31.20%**

λ messages that are related to COVID-19 as a user's past messages. Similarly, the second variant, named by SE-CGNN-TE-UNREL, only consider those unrelated to COVID-19. The last SE-CGNN-TE-ALL variant ignores the difference and straightforwardly consider the last λ messages regardless of their types. The same as our previous experiments, we train these three variants and run them on the three sets of testing cascades, i.e., $\mathcal{C}$, $\mathcal{C}_{PM}$, and $\overline{\mathcal{C}}_{PM}$. The results are shown in Table 6.5. We also include the results of SE-CGNN-TE for comparison.

In general, we have two main observations. First, we observe that among the three variants, the one with only messages related to COVID generates the best performance while the one only utilising those unrelated to COVID performs the worst. This also confirms our findings in Section 6.4.3 that COVID related messages tend to impose stronger spillover effects on retweeting preventive measure messages. This performance difference also indicates this finding may also apply on other messages which are not relevant to preventive measures. Second, the integration method used in SE-CGNN-TE can effectively improve the performance. This improvement may come from two sources. On one hand, our selected method actually uses 2λ past messages. This implies that more information can help increase the prediction accuracy. On the other hand, a balance between these two types of information can be reached during the model training.

6.7 Conclusion & Discussion

In this chapter, we concentrate on the problem of cascade prediction for COVID-19 information about preventive measures on online social media platforms. Compared to previous works, we take into account the phenomenon that the exposure to different information will influence social media users' behaviour of participating in information diffusion during the pandemic, which we call *info-exposure spillover effect*. With a dataset we collected from Twitter, we successfully validate its existence. In particular, both COVID-19 related and unrelated messages may have spillover effects on the spread of COVID-19 messages promoting preventive measures. Meanwhile, the COVID related messages tend to impose stronger spillover effects. We then apply the identified spillover effects in predicting the popularity of preventive measure messages. Specifically, we build three new models by making use of the recent advances of graph representation techniques, i.e., graph neural networks (GNN). In addition, we utilise a temporal encoding method to capture the important variance caused by message posting time. With extensive experiments, we show that our new models outperform baselines not only for preventive measure messages but for all messages. This illustrates that the use of info-exposure spillover effect can effectively improve the performance of cascade prediction, and it should be recommended to be considered in designing future cascade prediction models. Specifically, we through this chapter showcase a general method, that can be referred to validate the existence of spillover effects of other types of information on the changes of information consumers' behaviours which are not restricted to retweeting. Moreover, other applications can also benefit from our work. For instance, social media posts have been used to extract effective indicators, e.g., numbers of daily posts and their sentiments [MSB+18], in predicting the price of cryptocurrencies such as Bitcoin. Our extended models can help accurately forecast the popularity of Bitcoin influencers' social media posts, e.g., Elon Musk [TRPP21], which can be integrated into existing models to further improve the accuracy of predicted prices. As our future work, we will consider other types of information in addition to users' textual posts and propose new methods to integrate them in cascade prediction.

We identify three main limitations that have not been well addressed in our current research. First, our empirical validation of the info-exposure spillover effect only focused on messages on Twitter related to preventive measures and conducted from the level of selected topics. Although in our experiment the overwhelming performance of our extended models on other general messages could partially validate its existence, finer-grained and more comprehensive analysis will be desired and we will take this as our future work. Second, our cascade prediction models are extended from existing GNN models.

It will be interesting to design a new end-to-end GNN model which is specifically adapted to the identified spillover effects of users' adopted information. Last, we only distinguish the significant difference between messages related and unrelated to COVID while ignoring the other linguistic features of individual messages.

Chapter 7

Exploring Topic-specific Susceptibility and Influence

In the preceding chapter, we addressed the research question of how to accurately predict information popularity. To this end, we developed a model that specifically assesses the impact of message exposure on a user's decision to participate in the diffusion of specific information. However, this approach overlooks the fact that when a user decides to retweet a message, it is not only influenced by the message content and active friends, but also by the user's susceptibility, i.e., the likelihood that the user will be influenced by others.

To fully consider all three factors in cascade prediction, we propose a new model in this chapter. We introduce a novel deep learning cascade prediction model, CasSIM, which accomplishes two objectives of great interest simultaneously: popularity prediction and end-user prediction. Unlike existing approaches based on cascade representations, CasSIM models the information diffusion process by studying the dual role of users in information dissemination, successfully capturing all three factors in cascade prediction. Moreover, the implementation of GNNs enables CasSIM to capture the dynamics of susceptibility and influence in the information diffusion process. We evaluate the effectiveness of CasSIM using three real-life datasets, and the results show that CasSIM outperforms state-of-the-art methods in terms of prevalence and end-user prediction.

7.1 Introduction

On social media, people are sharing billions of posts, news and videos with their friends or followers everyday. These sharing behaviours lead to rapid diffusion of unprecedented amounts of information [CZZ+19] in the form of *cascades*. The prevalence of information cascades exposes people to information of their interest faster, and meanwhile also amplifies the damage of false information such as rumours [GPGC21]. The COVID-19 pandemic gives us a chance to rediscover the importance of social media not only as networking platforms but also as an information source which can actually interfere with our everyday decisions [XZZ+23]. Thus, it is crucial to understand and forecast cascade dynamics to effectively promote useful messages, e.g., for viral marketing [WSLC15], and proactively control the impact of misinformation [LW18, SHL17]. The problem of *cascade prediction* aims to achieve two objectives along this direction: *popularity prediction* and *adopter prediction*. We say that a user *adopts* a message and becomes an active adopter if the user shares the message from one of his/her friends. With an observation of early adopters, the goal of popularity prediction is to predict the number of final active adopters while adopter prediction is to forecast who will adopt the message at a future time point.

Cascade prediction has garnered attention from both industry and academy over the past decade [CAD+14, YCW+15] and the solutions have evolved from the methods based on *diffusion models* [PMV20] to those based on *cascade representation* [CZZ+19]. Diffusion model-based methods characterise the interpersonal influences between users and simulate the diffusion process through social relations. These methods are not scalable for large networks due to the repetitive simulations of diffusion models. Moreover, they rely on some unrealistic assumptions such as independent cascades and uniform influence probabilities between users [PMV20]. Therefore, despite their explainability, this class of methods are suboptimal for cascade prediction. By contrast, the methods based on cascade representation characterise the features of observed early cascades instead of modelling diffusion processes. Machine learning models are employed for downstream predictions. These methods have become the state-of-the-art due to their overwhelming prediction performance, especially with the recent success of deep learning. Compared to earlier methods using hand-crafted predictive features, deep learning allows for automatic extraction of cascade representations which capture the heterogeneous types of information embedded in cascades [XZZ+23]. For instance, the application of recurrent neural networks (RNN) and graph node embedding simultaneously capture the temporal rankings of early adopters and the structural properties of their neighbours in social graphs [YTS+19]. Despite their promising performance, deep learning methods confront a few inherent challenges as repeatedly emphasised in the literature such as the imbalanced distribution of cascades [TLH+21] and cascade graph dynamics [SRZ+22]. Moreover, except for FOREST [YTS+19], they are designed either for popularity prediction or for microscopic prediction which infers the *next adopter*. Without modelling diffusion processes, they are thus suboptimal for predicting final active adopters. **Our contributions.** In this chapter, we aim to combine the advantages of the two classes of previous studies and apply deep learning to model the diffusion process of information on social media. This approach will allow us to get rid of the inherent challenges in embedding observed cascades, and efficiently achieve the two objectives using a single method. The key to diffusion process modelling is to capture the interpersonal influences between a user and his/her friends before adopting a message. Cao et al. [CSG+20] conducted the first attempt CoupledGNN by modelling the cascading effect only with users' influences. One shortcoming of this method is that it ignores the double roles simultaneously played by users in information diffusion: distributors and receivers which have been widely accepted in the literature [PMV20, WSLC15].

In this chapter, our goal is to explore users' profiles of these two roles in cascade prediction. Specifically, we would like to address the following perspectives in our new diffusion process modelling which have not been well studied in the literature:

- A user's decision to forward a message should result from three factors: *message content*, *influences* of active friends and *susceptibility* of the user. Intuitively, a user's influence measures his/her ability to convince another user to share his/her message while susceptibility measures how likely the user gets influenced by other users [PMV20, WSLC15].
- Users' influence and susceptibility are not only user-specific but also *topic-specific*. This phenomenon has not been discussed before in the literature. Social media users, especially on platforms featured by microblogs such as Twitter and Weibo, usually have multiple topics of interests and different sharing patterns. Suppose a sports news reporter with a hobby of pop music. He will be more influential for sports-related tweets than those about music. As an information receiver, the reporter will be more cautious to spread sports news compared to music-related tweets.
- Influences and susceptibilities are *context-dependent* [WSLC15]. In other words, they spread through social relations during the diffusion process. A user will become more susceptible to a message when he/she sees that message shared by a larger number of users. Similarly, when more users have adopted the message a user shared, then the user becomes more influential to his/her friends due to the accumulated trust in the message.

To the best of our knowledge, we are the first to integrate users' topic-specific and context-dependent susceptibility and influence into cascade prediction. We start by validating our hypothesis that users' influence and susceptibility are topic-specific with our collected Twitter data. Then we propose a new deep learning cascade prediction model, which leverages the social network structure and simulates the propagation of messages from early adopters through social relations. The model can be effectively trained to achieve the two cascade prediction objectives at the same time. In this model, we explicitly embed users' susceptibility and influence profiles as two representation vectors. With GNNs [KW17], we model the activation of users according to topic-specific susceptibilities and influences and the dynamics of susceptibilities and influences. Through comprehensive experiments with three real-life datasets, we show our model outperforms state-of-the-art baselines in both popularity and adopter prediction with almost all measurements.

7.2 Problem Definition

Let $\mathcal{M}$ be a set of messages. We use the term "*message*" to refer to a piece of information that can be disseminated over social media. It can be a tweet on Twitter or an image on Instagram. In this chapter, we focus on textual messages and our approach can be straightforwardly extended to other message types if their representations can be effectively calculated and adapted. For any message $m \in \mathcal{M}$, we have the set of active users that had adopted this message up to the time t_0 after the message was first posted, denoted by $C_m^{t_0}$. The observation time t_0 depends on the requirements of downstream applications as well as the popularity of social media platforms. It can be of hours on Twitter and Weibo, and years for citation networks. We use $G = (\mathcal{V}, \mathcal{E})$ to denote be social graph recording the social relations between users. Specifically, $\mathcal{V}$ is the set of nodes that represent the set of users and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set edges indicating the social relations. The network can be directed or undirected depending on social media platforms. For instance, the following relation on Twitter is directed while friendship on Facebook is undirected.

Popularity prediction. The problem of *popularity prediction* is to predict the final number of active users, i.e., $n_m^\infty = |C_m^\infty|$. In practical applications, the final time can be determined as a given fixed time period t or by approximating the cessation of growth or slow growth of the cascade. This is also commonly called the *macroscopic cascade prediction* in the literature [YCW+15, YTS+19]. Formally, given a set of messages $\mathcal{M}$ and their observed cascades $\{C_m^t | m \in \mathcal{M}\}$, the problem of popularity prediction can be solved by minimising the following mean relative square error (MRSE) loss:

$$\mathcal{L}_{pop} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Big(\frac{\widetilde{n}_m^\infty - n_m^\infty}{n_m^\infty} \Big)^2 \tag{7.1}$$

where $\tilde{n}_m^\infty = f_{\Theta,G}(C_m^{t_0})$. Note that $f_{\Theta,G} : \mathcal{V}^{\mathcal{P}} \to \mathbb{Z}$ is the regression function customised to graph G and parameterised by the set of trainable parameters Θ where $\mathcal{V}^{\mathcal{P}}$ denotes the powerset of $\mathcal{V}$. It takes an observed cascade as input and outputs the predicted final size of the cascade.

Final adopter prediction. The goal is to predict the set of users who will forward the target message. This is different from the *microscopic cascade prediction* in the literature [YCW+15, YTS+19] which aims to predict the next active user according to the observed ones. The problem of final adopter prediction can be solved by minimising the following loss function:

$$\mathcal{L}_{\mathbf{adp}} = -\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Big(\sum_{v \in C_m^\infty} \log q_{\Theta,G}(C_m^{t_o}, v) + \sum_{v \notin C_m^\infty} \big(1 - \log q_{\Theta,G}(C_m^{t_o}, v)\big) \Big) \tag{7.2}$$

where $q_{\Theta,\mathcal{G}} : \mathcal{V}^{\mathcal{P}} \times \mathcal{V} \to [0,1]$ is the trainable function customised to social graph G and parameterised by Θ that predicts the probability that a specific user will adopt the message. In the end, we can select the users with probabilities larger than a predefined threshold as the output set of final adopters. An alternative is to output the top $\tilde{n}_m^\infty$ users with the largest activation probabilities.

7.3 Topic-specific Susceptibility and Influence

In this section, we will validate our hypotheses that a user's susceptibility and influence vary according to the topic of the message. This hypothesis actually contains an implicit claim that users adopt messages of multiple topics on social media. In other words, users have their own topic preferences. We start with validating this claim and then examine the dependence of susceptibility and influence on topics.

Table 7.1: The statistics of our Twitter dataset.

Social network	#node		$5,808,938$
	#edge		$12,511,698$
	average degree		2.15
Timeline tweets	Period 1 (1/3/2020-30/6/2020)	#tweet	7,855,186
		#tweet per user	627.51
	Period 2 (1/3/2021-30/6/2021)	#tweet	3,591,664
		#tweet per user	303.76

7.3.1 Twitter Data Collection

In this section, we gather the datasets utilised in this chapter, drawing from the EU-VAX dataset as a foundation. We then collect the followers and followees for each originator of a tweet in the dataset, successfully constructing the social network. Specifically, if user v follows user v', an edge is created from v to v'. To eliminate isolated users and ensure connectivity between users, we calculate the largest weakly connected subgraph of the social network as the final set of users.

We conduct the timeline tweet streaming for the remaining users in two time periods, each spanning three months. One period starts from March 1st, 2020, while the other starts from March 1st, 2021. By examining these two periods separated by one year, we can assess the consistency of our empirical analysis over time. As our purpose is to examine users' sharing behaviours in information diffusion, we only keep those tweets that each user has retweeted. We summarise the statistics of the final social networks and tweets in Table 7.1.

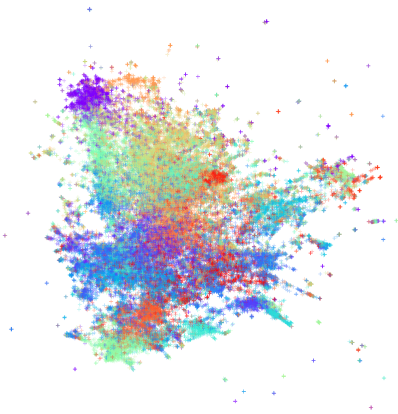
In this chapter, we focus on the texts of retweeted messages and thus remove all other contents such as '@', hyperlinks, and 'RT', which stands for 'retweet'. For quoted tweets, we only consider the quoted tweets and ignore the comments added by users. In our analysis, we only consider the users with more than 5 retweets in our dataset to ensure the reliability of our analysis.

7.3.2 Users' Topic Preferences

We validate our observation that users simultaneously participate in discourses of multiple topics on social media. We take the three months' retweets in each period as our observations and extract the topics expressed in them. Then we calculate the distribution of a user's retweets on these topics as his/her topic preferences.

Topic modelling. *Topic modelling*, as an important task in natural language processing, has upgraded from traditional LDA method [BNJ01] to machining learning methods [GC06]. In [ZFCN22], it has been shown that the combination of high-quality text embeddings and clustering methods can more efficiently learn topics of the same quality as complex neural network models. In this chapter, we adopt the most effective combination in [ZFCN22], i.e., RoBERTa+UMAP+K-Means, to cluster

(a) Period 1



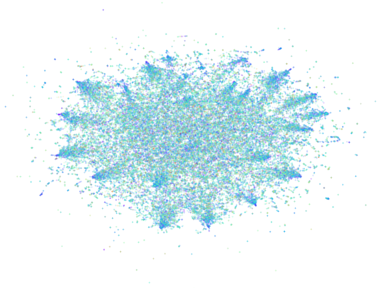
(b) Period 2

Figure 7.1: Clustering retweets into topics.

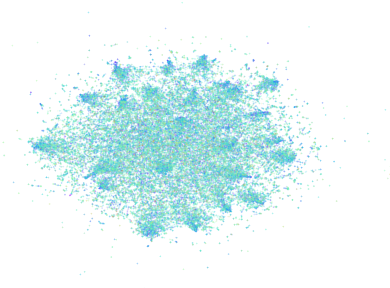
tweets with similar topics. RoBERTA [LOG+19] is a pre-trained transformer-based text embedding method and UMAP [MH18] is used to conduct dimension reduction of text embeddings while K-Means is one of the most widely used classic clustering methods. Besides textual tweets, the number of topics is required as an input parameter.

We classify the collected tweets in each period with the selected topic modelling method. After several trials, we select 25 topics due to the relatively higher quality of the output clusters. In the end, we have 25 clusters of retweets, the i-th of whcih is denoted by $\mathcal{S}_i$. In Figure 7.1, we depict retweets as data points and layout them according to their text embedding vectors mapped to a 2-D space with UMAP [MH18] in the two selected periods. The colours indicate their clusters. With the widely accepted measurements: C_V and NPMI, we measure the coherence values which are 0.649 and 0.138 for the first period, and 0.704 and 0.142 for the second period. According to general criteria, these numbers indicate a more than good topic coherence. We extract the representative keywords with their TFIDF rankings, and manually examine the topics of the clusters. We find that in general the tweets in these clusters are about specific topics such as the death and infection number of COVID-19, Black Life Matters movement and the COVID-19 policies.

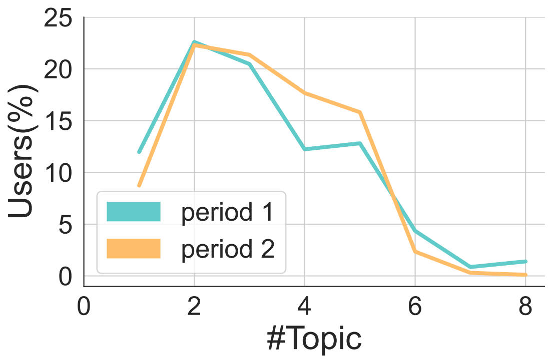
User topic preference. We represent the topic preferences of user v by a vector $\mathbf{v}$ by counting the proportions of his/her retweets in each topic. Formally, let $\mathcal{S}_v$ be the set of retweets of user v, then the j-th element of $\mathbf{v}$ is calculated as $\frac{|\mathcal{S}_v \cap \mathcal{S}_i|}{|\mathcal{S}_v|}$. In Figure 7.2(a) and 7.2(b) we layout users as data points according to their preference vectors mapped to 2 dimensions in the two periods. We can see that users' vectors scatter all over the space. Another observation is that users cluster naturally, which reflects the groups of users with similar interests. We consider a user is interested in the j-th topic if the j-th element of his representation vector is over 0.08 which is double the value of a null model. Figure 7.2(c) shows the distributions of the number of topics users prefer in the two periods. We observe that about 86% users actively participate in at least 2 topics. On average, each user is interested in 3 topics. According to the above discussion, we can conclude that users are interested in multiple topics.



(a) Visualisation (Period 1)



(b) Visualisation (Period 2)



(c) Distribution of the number of topics of interest.

Figure 7.2: User topic preferences and distribution.

7.3.3 Topic-specific Susceptibility & Influence

Whether a user retweets a message is determined by his/her susceptibility and the influences received from his/her followers who have shared the message. We hypothesise that the interplay between susceptibility and influences is not only user-specific but also *topic-specific*. Many methods have been proposed to learn the latent representations for users' susceptibility and influences according to past observed cascades [PMV20, WSLC15]. However, we cannot validate our hypothesis by directly comparing the representations extracted from past cascades of different topics. This is because the learning processes on different topics are independent. Therefore, the learned representations do not belong to the same space and are not comparable. Therefore, we select an intuitive approach based on a heuristic utilised in the literature [BLG16] that if users' susceptibility and influence are topic-specific, we will have two observations:

1. As an information receiver, a user will have different patterns regarding sharing messages from his/her followees between topics;

2. As an information distributor, a user's followers will have different patterns regarding sharing messages retweeted or posted by the user.

If these two differences are present in our dataset, we can infer that the interaction of a user's susceptibility and influence varies between topics. After a user shares a message, the user will have an influence on each follower's decision whether to share the message. According to this intuition, we use the frequency with which a follower forwards messages after the user's sharing to measure the strength of the interplay between the user's influence and the follower's susceptibility. In the following, we first present our measurements for a user's *susceptibility pattern* as a receiver and his/her *influence pattern* as a distributor, and then discuss our analysis on our dataset.

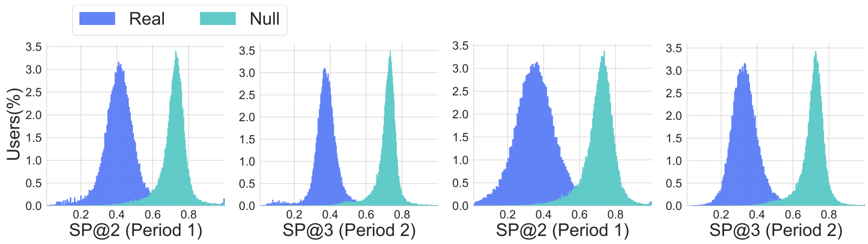


Figure 7.3: Distributions of susceptibility pattern *SP@K*.

Measuring susceptibility pattern. Intuitively, given a topic, we use the relative frequencies with which a user forwards messages from his/her followees to quantitatively capture the user's sharing pattern as an information receiver. Suppose a user v with the set of his/her followees $\mathcal{U}_v^+ = \{v' \in \mathcal{V} | (v, v') \in \mathcal{E}\}$. We assume a pre-defined order between the followees of user v and use v_i to denote the i-th followee. Let $\mathcal{S}_{v,j}$ be the set of tweets that u retweeted about the j-th topic, and $time(m, v)$ to denote the time when m is posted or retweeted by user v. The susceptibility vector of v is denoted by $\mathbf{s}_{v,j} \in \mathbb{Z}^{|\mathcal{U}_v^+|}$ whose i-th element is the number of messages retweeted by the i-th followee in $\mathcal{U}_v^+$ before v retweets the same message, i.e, $|\{m \in \mathcal{S}_{v,j} \cap \mathcal{S}_{v_i} | t(m, u) > t(m, v_i)\}|$.

As we discussed previously, a user is only interested in a certain number of topics. As a result, we consider the K most preferred topics of user u to measure his/her *susceptibility pattern* (denoted by *SP@K*) by averaging their mutual similarities. Let $\mathcal{T}_K^v$ is the set of the indices of users' top K favourite topics. Formally, we calculate *SP@K* with the average similarity of all the pairs of top topics:

$$SP@K(v) = \frac{2}{K\cdot(K-1)} \sum_{j,k \in \mathcal{T}_K^v \wedge j<k} \frac{\mathbf{s}_{v,j} \cdot \mathbf{s}_{v,k}}{\|\mathbf{s}_{v,j}\| \cdot \|\mathbf{s}_{v,k}\|}.$$

Measuring influence pattern. We use the frequencies which a user's followers share his/her retweeted messages to quantify the influence patterns of the user as an information distributor. Suppose a user v with the set of followers $\mathcal{U}_v^- = \{v' \in \mathcal{V} | (v', v) \in \mathcal{E}\}$ ranked according to a pre-defined order. Let $\mathbf{h}_{v,j}$ denote the influence vector of user v of the j-th topic. Then the i-th element is the number of retweets conducted by the follower v_i in $\mathcal{U}_v^-$ after seeing the same message posted or retweeted by user v, i.e., $|\{m \in \mathcal{S}_{v,j} \cap \mathcal{S}_{u_i} | t(m, v) < t(m, v_i)\}|$. Similar to the definition of topic-dependent susceptibility, we also consider the top K favourite topics of user v, i.e., $\mathcal{T}_K^v$. Then the *influence pattern* of user v is defined as follows:

$$IP@K(v) = \frac{2}{K\cdot(K-1)} \sum_{j,k \in \mathcal{T}_K^v \wedge j<k} \frac{\mathbf{h}_{v,j} \cdot \mathbf{h}_{v,k}}{\|\mathbf{h}_{v,j}\| \cdot \|\mathbf{h}_{v,k}\|}.$$

The domains of *SP@K* and *IP@K* are between 0 and 1. A lower value indicates the interplay between users susceptibilities and influences varies more between topics.

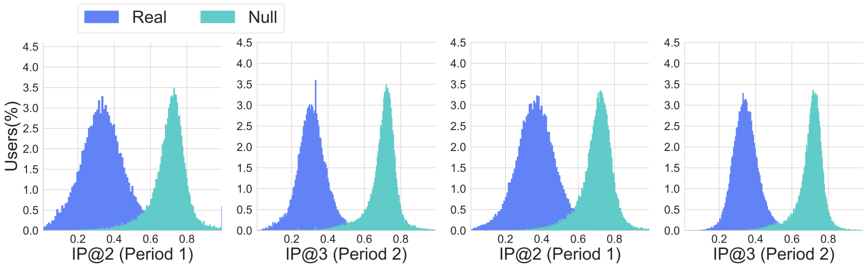


Figure 7.4: Distributions of influence pattern *IP@K*.

Experimental analysis. We re-use the topics extracted in Section 7.3.2 to analyse the topic dependence of susceptibility and influence. In Figure 7.3 and 7.4, we show the distributions of *SP@K* and *IP@K* values over the users when K is set to 2 and 3, respectively. For either measurement, we construct a null model as a reference to capture the distributions when the topic-specific phenomenon is absent. Take susceptibility patterns as an example. For any user v and each topic (e.g., the j-th topic), we construct a null vector $\mathbf{s}'_{v,j}$. Its k-th element is a uniformly sampled random number between 0 and $|\mathcal{S}_{v_k,j}|$ representing the number of messages of user $v_k \in \mathcal{U}_u^+$ that have been retweeted by user v.

A general observation is that users' susceptibility and influence patterns roughly follow normal distributions. The curves of the distributions become narrower and shift right when larger K values are set. This is natural that more topics considered will lead to smaller average mutual similarity. We can see with all the selected K values, users have smaller values for both susceptibility pattern and influence pattern than the null models. On average, users' *IP@K* and *SP@K* fall into the range between 0.3 and 0.4 which is only half of those when susceptibility and influence are not topic-specific. The difference indicates that users' sharing behaviours and their influences on friends differ between the topics of their interest.

7.4 Our CasSIM model

The propagation of a message can be interpreted as a process of multiple sequential generations. In any generation, each user first updates his/her influence and susceptibility according to the current activation states of users. Then the user decides whether to forward the targeted message according to his/her updated susceptibility and the influences of his/her friends who have forwarded the message. Inspired by CoupledGNN [CSG+20], we use multi-layered GNNs to model this iterative process. We depict our framework in Figure 7.5. At each layer, three sequential tasks are accomplished. The first task is to update susceptibility and influence by aggregating the profiles of social network neighbours. This task actually simulates the spread of influence and susceptibility and thus captures their context-dependence property. The second task is to calculate the topic-specific influences and susceptibility according to the user's topic preferences and the target message's content.

The last task is to update each user's activation state by aggregating the interplay between his/her susceptibility and the influences of all the active friends.

For each user $v \in \mathcal{V}$, we use $State_v \in [0, 1]$ to store his/her activation state indicating the probability that user v is activated. Furthermore, for each generation ℓ, user v is associated with three embedding vectors $\mathbf{r}_v^{(\ell)}$, $\mathbf{h}_v^{(\ell)}$ and $\mathbf{p}_v$ indicating his/her susceptibility, influence and topic preferences, respectively.

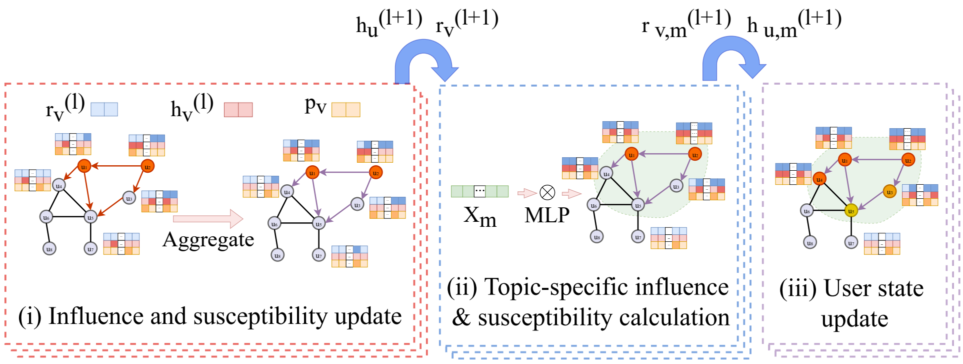


Figure 7.5: Framework of the CasSIM model.

7.4.1 Influence and Susceptibility Update

As users' influence and susceptibility propagate through social relations, we make use of a GNN to first aggregate the profiles from their friends and then combine the aggregation with their own profiles. We start by describing the update of susceptibility vectors.

We use the idea of graph attention networks [VCC+18] to take into account the various contributions of friends to the update. Let $\mathcal{N}(v)$ be the neighbours of user v, i.e., $\{v' \in \mathcal{V} | (v, v') \in \mathcal{E}\}$. Formally, the aggregated susceptibility of user v can be calculated as follows:

$$\mathbf{a}_{v,s}^{(\ell)} = \mathbf{W}^{(\ell)} \sum_{v' \in \mathcal{N}(v)} \mathbf{h}_{v'}^{(\ell)} \cdot \mathrm{StateGate}(State_{v'}^{(\ell)}) \cdot \phi_{v,v'}^{(\ell)} \tag{7.3}$$

where $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_r^{(\ell)} \times d_r^{(\ell+1)}}$ is the weight matrix and $d_r^{(\ell)}$ defines the dimension of a user's susceptibility vector at the ℓ-th layer, i.e., $\mathbf{r}_v^{(\ell)}$. The function StateGate() is the state-gating mechanism [CSG+20] to reflect the non-linearity of activation states. In our implementation, we use a 2-layered MLP. The attention $\phi_{v,v'}^{(\ell)}$ calculates the contribution of the influence of user v's neighbour v'. This attention is determined not only by v''s susceptibility but also by v's influence vector. Formally, the attention is calculated as follows:

$$\phi_{v,v'}^{(\ell)} = \frac{\exp(\mathbf{e}_{v,v'}^{(\ell)})}{\sum_{v'' \in \mathcal{N}(v)} \exp(\mathbf{e}_{v,v''}^{(\ell)})} \tag{7.4}$$

where $\mathbf{e}_{u,u'}^{(\ell)} = \boldsymbol{\psi}_s^{(\ell)}(\mathbf{W}_r^{(\ell)} \mathbf{r}_{v'}^{(\ell)} \parallel \mathbf{W}_h^{(\ell)} \mathbf{h}_v^{(\ell)})$. Note that $\parallel$ is the concatenation function of two vectors, and $\boldsymbol{\psi}_s \in \mathbb{R}^{d_h^{(\ell)} + d_r^{(\ell)}}$ where $d_h^{(\ell)}$ is the dimension of influence vectors at layer ℓ. Moreover, $\mathbf{W}_h^{(\ell)} \in \mathbb{R}^{d_h^{(\ell)} \times d_h^{(\ell)}}$ and $\mathbf{W}_r^{(\ell)} \in \mathbb{R}^{d_r^{(\ell)} \times d_r^{(\ell)}}$ are two weight matrices.

In the end, we combine the aggregated susceptibility of neighbours with the user's own susceptibility:

$$\boldsymbol{r}_v^{(\ell+1)} = \mathrm{relu}\left(\mathbf{W}^{(\ell)}\mathbf{h}_v^{(\ell)} + \boldsymbol{a}_{v,s}^{(\ell)}\right) \tag{7.5}$$

where $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_r^{(\ell+1)} \times d_s^{(\ell+1)}}$ and relu is the non-linear activation function.

The update of user v's influence is similar to that of his/her susceptibility. We first aggregate the influence of his/her friends according to their activation states with attention networks. Formally, the aggregated influence $\mathbf{a}_{u,h}^{(\ell)}$ is calculated as follows:

$$\mathbf{a}_{v,h}^{(\ell)} = \mathbf{W}^{(\ell)} \sum_{v' \in \mathcal{N}(v)} \mathbf{h}_{v'}^{(\ell)} \cdot \mathrm{StateGate}(State_v^{(\ell)}) \cdot \lambda_{v,v'}^{(\ell)} \tag{7.6}$$

where $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_h^{(\ell)} \times d_h^{(\ell+1)}}$ is the weight matrix. We calculate the attention $\lambda_{v,v'}^{(\ell)}$ as:

$$\lambda_{v,v'}^{(\ell)} = \frac{\exp(\mathbf{o}_{v,v'}^{(\ell)})}{\sum_{v'' \in \mathcal{N}(v)} \exp(\mathbf{o}_{v,v''}^{(\ell)})}, \tag{7.7}$$

where $\mathbf{o}_{v,v'}^{(\ell)} = \boldsymbol{\psi}_h^{(\ell)}(\mathbf{W}_h^{(\ell)}\mathbf{h}_{v'}^{(\ell)} \parallel \mathbf{W}_r^{(\ell)}\mathbf{r}_v^{(\ell)})$. Note that $\boldsymbol{\psi}_h \in \mathbb{R}^{d_h^{(\ell)}+d_r^{(\ell)}}$ and $\mathbf{W}_h^{(\ell)} \in \mathbb{R}^{d_h^{(\ell)} \times d_h^{(\ell)}}$ and $\mathbf{W}_r^{(\ell)} \in \mathbb{R}^{d_r^{(\ell)} \times d_r^{(\ell)}}$ are two weight matrices which are different from those used in updating susceptibility. User's influence vector at the layer $\ell + 1$ is calculated as follows:

$$\mathbf{h}_u^{(\ell+1)} = \mathrm{relu}(\mathbf{W}^{(\ell)}\mathbf{h}_v^{(\ell)} + \boldsymbol{a}_{v,h}^{(\ell)}) \tag{7.8}$$

7.4.2 Calculating Topic-specific Influence & Susceptibility

We show how to customise a user's susceptibility and influence according to the topic of the message under diffusion in order to capture their topic-specific property. Suppose $m \in \mathcal{M}$ is the message being propagated. We take user $v \notin C_m^{t_0}$ at ℓ-th generation as an example and illustrate how to convert the vectors $\mathbf{r}_v^{(\ell)}$ and $\mathbf{h}_v^{(\ell)}$ into $\mathbf{r}_{v,m}^{(\ell)}$ and $\mathbf{h}_{v,m}^{(\ell)}$. We use $\mathbf{x}_m \in \mathbb{R}^{d_x}$ to denote the embedding vector of message m. As emphasised previously, in this chapter, we concentrate on messages in the form of texts and the model can be extended to integrate other formats, such as images if their representations can be effectively calculated. In our model, we use the pre-trained RoBERTa model [LOG+19] to calculate the embedding vectors of textual messages.

As empirically validated in the previous section, most users have multiple topics of interest on social media and their preferences vary between topics. Although the focus of the topics may shift over time as pointed out in [YLZ+20], users' interests remain relatively stable. For instance, a sports news reporter may switch from reporting a local football team to national teams due to the opening of the FIFA world cup, but the topic still remains around football. Users' topic preferences are extracted from their past sharing behaviours. We use $\mathbf{p}_v \in \mathbb{R}^{d_p}$ to denote the embedding vector for his/her topic preferences. Intuitively, given a targeted message m, we capture its related topic by referring to users' past topic preferences and utilise an MLP module to calculate the adjustments that should be exposed to the user's susceptibility and influence vectors. Starting with susceptibility, we calculate the corresponding topic-specific susceptibility vector as follows:

$$\mathbf{r}_{v,m}^{(\ell)} = \mathbf{W}^{(\ell)}\big(\mathrm{MLP}(\mathbf{p}_v \parallel \mathbf{x}_m) \circ \mathbf{r}_v^{(\ell)}\big) \tag{7.9}$$

where $\circ$ represents the dot product of two vectors. In addition, MLP is a multi-layer perceptron with an input vector of dimension $d_p + d_x$ and outputs a vector of dimension $d_r^{(\ell)}$. The weight matrix $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_{r,m}^{(\ell)} \times d_r^{(\ell)}}$ conducts the linear transformation of the susceptibility vector to dimension $d_{r,m}^{(\ell)}$. Similarly we have the user's topic-specific influence calculated as follows:

$$\mathbf{h}_{v,m}^{(\ell)} = \mathbf{W}^{(\ell)}\big(\mathrm{MLP}(\mathbf{p}_v \parallel \mathbf{x}_m) \circ \mathbf{h}_v^{(\ell)}\big). \tag{7.10}$$

Note that $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d_{h,m}^{(\ell)} \times d_h^{(\ell)}}$ and the output of the MLP has the dimension of $d_h^{(\ell)}$.

7.4.3 User State Update

With users' topic-specific susceptibility and influence, we can model their interplay which changes their activation states. The influences of each user's active neighbours are first aggregated as the total amount of topic-specific influences exposed to the user. Then we use an MLP module to capture the likelihood of the user adopting the message only according to the exposed influences, denoted by $\gamma_v^{(\ell)}$. WE user

$$\gamma_v^{(\ell)} = \mathrm{sigmoid}\Big(\mathrm{MLP}\Big(\Big(\sum_{v' \in \mathcal{N}(v)} \mathbf{h}_{u,x}^{(\ell)} \cdot State_{v'}^{(\ell)}\Big) \parallel \mathbf{r}_{v,m}^{(\ell)}\Big) + \beta_v\Big). \tag{7.11}$$

where $\beta_u \in \mathbb{R}$ is a self activation parameter. Intuitively, the probability is dependent on the user's topic preferences. In our model, we use a one-layer MLP followed by a sigmoid function to capture this dependence. In the end, we combine the above activation probability with the user's current activation status into the user's new activation state:

$$State_v^{(\ell+1)} = \begin{cases} 1, & \text{if } v \in C_m^{t_0} \\ \mathrm{sigmoid}\big(\mu_1^{(\ell)} State_v^{(\ell)} + \mu_2^{(\ell)} \gamma_v^{(\ell)}\big), & \text{if } v \notin C_m^{t_0}. \end{cases} \tag{7.12}$$

Note that $\mu_1^{(\ell)}, \mu_2^{(\ell)} \in \mathbb{R}$ are two weight parameters which are to be trained. The initial state, i.e., $State_v^{(0)}$, is set to 1 if $v \in C_m^{t_0}$ or 0, otherwise. In the end, we calculate the final size of the cascade $\tilde{n}_m^{\infty}$ as $\sum_{v \in \mathcal{V}} State_v$.

7.4.4 User Profiling

From the above discussion, we can see that our model uses three input vectors for each user v at the 0-th layer: $\mathbf{p}_v^{(0)}$, $\mathbf{r}_v^{(0)}$ and $\mathbf{h}_v^{(0)}$. A few methods have been proposed in the literature to learn users' susceptibility and influence embedding from users' sharing history [WSLC15, PMV20]. In this chapter, we pre-train a simple but effective model to prepare the three types of initial vectors. Suppose we have the cascades for the past messages in $\mathcal{M}_{hist}$. We interpret them as the ultimate states of users in the corresponding information diffusion processes. In other words, for each $m \in \mathcal{M}_{hist}$, we have the final cascade C_m^∞. We set $State_v^\infty = 1$ if $v \in C_m^\infty$ and 0 otherwise. We calculate the activation state for each user $v \in C_m^\infty$ based on his/her topic-specific susceptibility and his/her active friends' influence, which is denoted by $\widetilde{State}_{v,m}$. Formally, it is calculated as follows:

$$\widetilde{State}_{v,m} = \text{sigmoid}\Big(\mathbf{ff} \cdot \sum_{v' \in \mathcal{N}(v) \cap C_m^\infty} \big(\text{MLP}(\mathbf{p}_v^{(0)} \parallel \mathbf{x}_m) \circ (\mathbf{h}_{v'}^{(0)} \parallel \mathbf{r}_v^{(0)})\big)\Big) \tag{7.13}$$

where $\mathbf{ff} \in \mathbb{R}^{d_r^{(0)} + d_h^{(0)}}$ and MLP outputs a vector of dimension $d_r^{(0)} + d_h^{(0)}$. In the end, $\mathbf{p}_v^{(0)}$, $\mathbf{r}_v^{(0)}$ and $\mathbf{h}_v^{(0)}$ are trained by minimising the following objective function:

$$\mathcal{L}_{initial} = -\frac{1}{|\mathcal{M}_{hist}|} \sum_{m \in \mathcal{M}_{hist}} \sum_{v' \in C_m^\infty} \log(\widetilde{State}_{v,m}). \tag{7.14}$$

There may exist users who do not participate in any cascades. For these users, we set the neutral vectors $\mathbf{0}$ to these users as their three profile vectors.

7.4.5 Model Training

In order to achieve the two objectives of cascade prediction: popularity and final adopter prediction, we aggregate the two corresponding objective functions into our final loss function to guide the parameter optimisation: $\mathcal{L} = \theta_1 \mathcal{L}_{adp} + \theta_2 \mathcal{L}_{pop} + \theta_3 \mathcal{L}_{reg}$ where θ_1, θ_2 and θ_3 are hyper-parameters. The $\mathcal{L}_{reg}$ is added for the purpose of regularisation as a L2 norm of all the model parameters.

7.5 Experimental Evaluation

7.5.1 Datasets

Table 7.2: Statistics of Sina, AMINER, and Twitter datasets.

	Social network		Ave. # user per cascade			#cascades
	#node	#egdes	Observation 1	Observation 2	Observation 3	
Sina	1,776,950	308,489,739	28.36	34.79	38.22	34,897
AMINER	131,415	842,542	14.71	15.71	18.80	30,106
Twitter	5,808,938	12,511,698	27.64	30.99	33.76	81,331

We leverage three real-life datasets in our experiments: Sina Weibo, AMINER and Twitter. Sina Weibo and AMINER are publicly available and widely exploited in the validation of previous works related to cascade prediction [LMGM17, CSG+20]. The Twitter dataset is an extension of our collection described in Section 7.3. Each dataset consists of two components: a social graph and a text dataset consisting of diffused messages.

We select these datasets to ensure a comprehensive evaluation that covers as many practical scenarios as possible. Weibo and Twitter represent the social media platforms characterised by microblogs. The users of the Weibo dataset are more densely connected. This dense social graph will benefit cascade prediction models with a more complete view of the sources of influences. AMINER is a citation network instead of social media storing the citation relations between academic authors. We use AMINER to test whether our CasSIM model can also predict the cascades in more general settings. Moreover, in order to check the performance of our model for different lengths of observation periods, i.e., t_0, for each dataset, we construct three sets of cascades by cutting the cascades according to three given time periods. For Twitter and Weibo, due to their fast propagation speed, the observation periods are set to 1 hour, 2 hours and 3 hours. For AMINER, we set them to 1 year, 2 years and 3 years. We briefly describe these three datasets as follows. The detailed statistics are summarised in Table 7.2.

Sina Weibo. The dataset is collected by Zhang et al. [ZLT$^+$13] from Sina Weibo, a popular Chinese social media platform and spans from September 28th, 2012 to October 29th, 2012. The social graph contains 1,776,950 users and 308,489,739 following relations between users. For each user, the dataset includes 1,000 additional most recent microblogs. The text data include 23,755,810 retweeted microblogs based on which we constructed 300,000 cascades. We only keep the cascades with more than 15 users. These cascades have an average size of 31.16.

AMINER. AMINER are constructed based on data collected from the DBLP computer science bibliography[1] which contain research papers' abstracts, titles and authors [TZY$^+$08]. The social graph describes authors and their citation relations between 1992 and 2002. If an author v publishes a paper citing one paper authored by v', then an edge from v to v' is added to the social graph. The social graph is composed of 131,425 authors and 8,466,859 citation relations. A cascade corresponds to a paper and tracks the researchers who ever co-authored the paper or cited the paper.

Twitter. We collected this dataset with the same social graphs used in Section 7.3. Specifically, in addition to the six months' tweets, this dataset contains the retweets spanning from March 1st, 2020 to October 30th, 2021 for almost two years. Our experiments use the cascades with a size larger than 15 users. In total, we have 81,331 cascades with an average size of 36.45.

7.5.2 Baselines

Considering the number of methods for macroscopic and microscopic prediction, we only select the *representative* methods for comparison. A method is representative if it is typical for a class of methods or it claims strong performances. Due to the overwhelming performance, we focus on deep learning methods in our evaluation.

[1]https://dblp.org/

DeepCas [LMGM17]. DeepCas is the first end-to-end representation learning-based method for popularity prediction. With the social graph, it represents cascades as cascade graphs and uses Bi-GRU to embed a cascade graph with random walk paths.

DeepHawkes [CSC$^+$17]. DeepHawkes integrates deep learning into Hawkes process for popularity prediction. It treats cascades as a temporal series of events. It combines user embedding vectors and cascades encoding by RNNs, and utilises the Hawkes process to model and predict cascade popularity.

CasCN [CZZ$^+$19]. Compared to models like DeepCas sampling random diffusion paths, CasCN samples sub-graphs from cascade graphs into a series of sequential sub-cascades. Then it propose a dynamic GCN model [BZSL14] to learn the representation of cascades for popularity prediction.

CoupledGNN [CSG$^+$20]. CoupledGNN leverages two GNNs to simulate the cascading effects of information diffusion. One is to update users' activation states and the other captures the spread of users' influence. Difference from our CasSIM model, it ignores the dual-role properties of users and the topic-specific properties of users' influences. In spite of being designed for popularity prediction, the explicit simulation of cascading effect allows it to predict the final adopters as well.

FOREST [YTS$^+$19]. FOREST makes use of an enforcement learning framework to endow a microscopic prediction model the ability of popularity prediction. As a result, it can predict both popularity and final adopters. The proposed microscopic model uses GRU neural networks to embed the representation of the last early adopter and predict the next adopter with this vector as well as the structural embedding.

CasFlow [XZZ$^+$23]. CasFlow uses graph signal processing, graph representation techniques and variational auto-encoder to capture node-level and cascade-level diffusion uncertainty. It is designed for popularity prediction.

TempCas [TLH$^+$21]. TempCas is designed for popularity prediction and claims to perform better than previous methods. It takes the temporal changes of cascade graphs into account and calculates the representation of a cascade graph based on a series of temporal snapshots of the cascade graphs. Variants of RNN, i.e., LSTM and GRU, are used to embed the time-series snapshots.

DyHGCN [YLZ$^+$20]. DyHGCN is designed for microscopic prediction, where it extracts structural information from both social graphs and cascade graphs to generate dynamic user embeddings through a heterogeneous GCN.

7.5.3 Experimental Settings

Evaluation Measurements. We use three widely adopted measurements to evaluate the prediction performances regarding popularity. MSLE (Mean square log-transformed error) is a standard evaluation metric [CZZ+19] and defined as:

$$MSLE = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} (\log n_m^{\infty} - \log \tilde{n}_m^{\infty})^2.$$

We use mean absolute percentage error (MAPE) and wrong percentage error (WroPerc) to evaluate prediction performance in terms of relative errors. MAPE measures the average relative errors and is defined as:

$$MAPE = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{|\tilde{n}_{\infty}^{m} - n_{\infty}^{m}|}{n_{\infty}^{m}}.$$

WroPerc measures the average percentage of cascades that are poorly predicted and is defined as:

$$WroPerc = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{1} \left(\frac{|\tilde{n}_{\infty}^{m} - n_{\infty}^{m}|}{n_{\infty}^{m}} \geq \varepsilon \right).$$

We set the threshold to 0.5 in our experiments. Note that $\mathbb{1}(*)$ is an indication function which outputs 1 when the input proposition is true or 0 otherwise. For each measurement, a lower value indicates better prediction performance.

With regard to evaluating the prediction performance of final adopters, we use the standard metrics: precision, recall and F1 score.

Hyperparameter settings. For each of the three datasets, we randomly split it into training, validation and testing sets according to the ratio 8:1:1. For the text embedding model RoBERTa, we utilise the implementation XLM-RoBERTa [CKG+20]. We set the maximum size of input strings to 128, and the length of text embedding is 768. For all models including the bench-marking models, we tune their hyper-parameters to obtain the best performance on validation sets. Early stopping is employed for tuning when validation errors do not decline for 20 consecutive epochs. The learning rate and L2 coefficient are chosen from $10^{-1}, 10^{-2}, \ldots, 10^{-8}$. The hidden units for MLPs are chosen from {32, 64}. The batch size is 32. We train our model for 500 epochs and utilise Adam [KB15] for optimisation. We use the first three months' cascades in the Sina Weibo and Twitter dataset to pre-train users' initial profiles: their topic preference, susceptibilities and influences. For the AMINER dataset, the first two years' cascades are used. As for DeepCas and TempCas, the sequence lengths and the number of sampled sequences are set to 100 and 8, respectively. The size of node embedding vectors of CasCN and DeepHawkes is fixed at 50. All other hyperparameters remain the same as they are recommended in the original papers or the published source codes.

7.5.4 Overall Prediction Performance

We compare the performance of our CasSIM model to the baselines for both the two cascade objectives: popularity prediction and final adopter prediction. As we discussed previously, only FOREST [YTS+19] and CoupledGNN [CSG+20] can achieve these two objectives simultaneously. As a result, for each objective, we conduct the comparison with the baselines that can achieve the objective. We independently train each model 5 times and report the average results on testing sets.

Popularity prediction. We outline the performance of all the benchmarks and our CasSIM model on Sina Weibo, AMINER and Twitter in Table 7.3, 7.4 and 7.5, respectively. We do not consider DyHGCN in this comparison since it can only conduct microscopic prediction, i.e., predicting the next single adopter. Our objective is to examine whether our CasSIM model can outperform the baselines in different scenarios. If not, we analyse the possible causes so as to understand the scenarios when our model can work the best. In general, we can observe that FOREST, CasFlow and TempCas are the best baselines in terms of popularity prediction. In addition, the prediction becomes more accurate when observation periods are longer. These two observations are consistent with the experimental evaluation in the literature [TLH+21, CZZ+19]. This consistency indicates the correctness of our implementation and our comparison is reliable and trustworthy. We highlight the best performance of each measurement in bold numbers and underline the second best.

We have three main observations. First, we observe that our CasSIM model outperforms almost all the baselines according to the three measurements in the three datasets. Tempcas only marginally outperforms CasSIM when the observation periods are set to 1 hour and 2 hours. This may be caused by the relatively large variances of cascade lengths in the Weibo dataset. The performance improvements show that our model can accurately predict the final size of cascades on both social media and citation networks where the cascading phenomenon exists. Second, compared to CoupleGNN, our CasSIM model can produce overwhelmingly more accurate predictions, especially when measured by WroPerc. For instance, on the Sina Weibo dataset, the increase is larger than 17%. The improvement can even reach 35% in our Twitter dataset. This means the performance of CasSIM is more stable than CoupledGNN. We can also infer that the consideration of users' dual roles in information diffusion is necessary and our CasSIM model effectively captures the interactions between users' susceptibilities and influences. Last, the improvement of our CasSIM model is more significant when observation periods are shorter. For instance, for the Sina Weibo dataset, CasSIM increases the performance measured by MLSE by 6% compared to Tempcas when observation periods are set to one hour. The increase drops to 3% for two-hour observation periods and further decreases to 2% when observation periods are three hours. We infer that this should result from our consideration of users' topic preferences and message contents in CasSIM. When shorter observation periods are set, the baselines which only rely on early adopters' co-occurrences in cascades do not have sufficient information for prediction.

Final adopter prediction. In the literature, only FOREST can predict the final adopters while predicting the popularity. It uses a microscopic prediction module to calculate the probability distribution over users to be the next activated user. FOREST iteratively samples the next adopters until a special virtual user named by 'STOP' is sampled. Compared to FOREST, CoupledGNN and our CasSIM model assign an

Table 7.3: Popularity prediction performance on Sina Weibo dataset.

Model	1 Hour			2 Hours			3 Hours		
	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc
DeepCas	3.578	0.291	32.26%	3.421	0.288	28.74%	3.139	0.270	18.58%
DeepHawkes	2.894	0.289	26.21%	2.551	0.280	25.89%	2.240	0.268	17.57%
CasCN	2.749	0.285	27.36%	2.442	0.283	25.56%	2.181	0.279	17.23%
CoupledGNN	2.289	0.242	23.60%	2.254	0.236	17.96%	2.037	0.223	14.27%
FOREST	<u>2.156</u>	0.238	20.05%	<u>2.136</u>	0.235	18.14%	1.995	0.230	13.49%
CasFlow	2.248	0.239	20.68%	2.195	<u>0.221</u>	16.79%	1.982	0.215	12.10%
TempCas	2.290	<u>0.226</u>	**18.23%**	2.208	0.229	**14.73%**	1.960	<u>0.209</u>	<u>11.26%</u>
CasSIM	**2.148**	**0.221**	<u>19.46%</u>	**2.126**	**0.217**	<u>14.94%</u>	**1.919**	**0.202**	**11.04%**

Table 7.4: Popularity prediction performance on AMINER dataset.

Model	1 Year			2 Years			3 Years		
	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc
DeepCas	2.031	0.293	28.33%	1.916	0.260	22.69%	1.908	0.227	21.39%
DeepHawkes	2.400	0.294	27.42%	1.148	0.252	22.47%	1.735	0.191	20.73%
CasCN	2.007	0.285	27.49%	1.959	0.283	20.28%	1.876	0.183	20.99%
CoupledGNN	1.970	0.288	25.90%	1.798	0.282	20.16%	1.430	0.165	19.63%
FOREST	1.359	0.293	25.11%	1.175	0.298	19.40%	1.495	0.154	18.88%
CasFlow	1.822	0.256	26.44%	1.086	<u>0.233</u>	17.01%	1.416	0.136	14.83%
TempCas	<u>1.308</u>	<u>0.242</u>	<u>24.66%</u>	<u>1.073</u>	0.236	<u>16.87%</u>	<u>1.384</u>	<u>0.130</u>	<u>14.84%</u>
CasSIM	**1.272**	**0.231**	**24.51%**	**1.063**	**0.225**	**16.26%**	**1.376**	**0.126**	**14.09%**

activation probability for each user. As both the models can predict the number of final adopters, i.e., $\tilde{n}_m^\infty$, we can use the $\tilde{n}_m^\infty$ users with the largest activation probabilities as the set of final adopters. Considering the inevitable prediction errors, we use a tolerant parameter η to add a certain percentage of extra adopters. It may be argued that microscopic models can also be applied to predict final adopters by iteratively predicting the next adopters which is similar to FOREST. However, different from FOREST, such models do not have the mechanisms to terminate the sampling. In order to ensure the comprehensiveness of our validation, we manually add an *unfair* terminating condition, that is, the true number of final adopters are sampled. We use the state-of-the-art microscopic model DyHGCN as a representative. In Table 7.6, we list the performance regarding final adopter prediction when observation periods are 3 hours for Twitter and Sina Weibo, and 3 years for AMINER. For the tolerant parameter η, we use 10%, 20%, 30%, 40% and 50% in our experiments.

We can see that CasSIM already perform better than FOREST and CoupledGNN with the original predicted popularity with η set to 0. DyHCGN only performs slightly better than CasSIM when applied on the Sina Weibo dataset. Although the improvement is a bit marginal compared to FOREST, CasSIM has a much better performance than CoupledGNN. With the relatively high-quality cascades in Sina Weibo, CasSIM increases the three measurements by about 18%. The improvement can reach more than 30% on the other two datasets. With positive η values set, we can observe an obvious performance increase for both CoupledGNN and CasSIM. It can be expected that a too large η will eventually compromise the performance. In our experiments, on all the three datasets, we can achieve best performance when η equals 40% and the performance started to fall when η is 50%.

Table 7.5: Popularity prediction performance on Twitter dataset.

Model	1 Hour			2 Hours			3 Hours		
	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc
DeepCas	12.897	0.614	39.73%	11.145	0.579	36.02%	11.677	0.547	30.13%
DeepHawkes	10.705	0.623	36.25%	10.499	0.617	35.83%	9.188	0.553	25.28%
CasCN	10.640	0.592	35.81%	9.207	0.552	34.63%	9.048	0.550	25.62%
CoupledGNN	9.400	0.497	34.49%	9.122	0.477	32.86%	9.045	0.452	22.55%
FOREST	8.799	0.489	33.01%	8.469	0.463	30.25%	8.147	0.454	21.46%
CasFlow	8.916	0.478	31.59%	8.114	0.458	28.94%	8.081	0.446	16.33%
TempCas	8.756	0.461	28.25%	8.251	0.442	26.67%	8.070	0.426	15.38%
CasSIM	**8.569**	**0.443**	**27.53%**	**8.046**	**0.437**	**25.43%**	**8.032**	**0.422**	**14.76%**

Table 7.6: Final adopter prediction performance.

Model	Sina Weibo			AMINER			Twitter		
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1
FOREST	0.436	0.379	0.383	0.398	0.344	0.392	0.404	0.416	0.413
DyHGCN	0.449	0.393	0.394	0.413	0.388	0.410	0.441	0.448	0.442
CoupledGNN	0.371	0.323	0.332	0.318	0.297	0.262	0.315	0.300	0.302
CoupledGNN+10%	0.392	0.348	0.352	0.325	0.307	0.319	0.335	0.324	0.329
CoupledGNN+20%	0.418	0.361	0.365	0.334	0.311	0.327	0.356	0.335	0.342
CoupledGNN+30%	0.431	0.372	0.378	0.370	0.359	0.364	0.371	0.387	0.385
CoupledGNN+40%	0.437	0.379	0.381	0.376	0.364	0.370	0.380	0.391	0.387
CoupledGNN+50%	0.422	0.364	0.370	0.366	0.341	0.351	0.350	0.332	0.341
CasSIM	0.443	0.390	0.394	0.409	0.397	0.412	0.417	0.436	0.425
CasSIM+10%	0.447	0.404	0.405	0.412	0.408	0.406	0.421	0.437	0.423
CasSIM+20%	0.452	0.411	0.414	0.422	0.416	0.420	0.432	0.439	0.433
CasSIM+30%	0.474	0.423	0.428	0.437	0.433	0.435	0.440	0.442	0.441
CasSIM+40%	0.477	0.441	0.451	0.439	0.435	0.436	0.443	0.446	0.444
CasSIM+50%	0.449	0.431	0.440	0.423	0.420	0.421	0.429	0.432	0.431

Discussion. From the above analysis, we can see our CasSIM model produce promising performance for both popularity prediction and final adopter prediction. Moreover, it effectively models the two roles of users in information diffusion. The integration of message contents into our model also helps improve the prediction of popularity when observation periods are short.

7.5.5 Ablation Study

We examine the contributions of the components which are implemented in our CasSIM model and missing in previous works. As we emphasised previously, the novelty of CasSIM is the diffusion process modelling which considers users' profiles as two roles, message contents and topic-specific susceptibilities and influences. We design three variants of CasSIM to study the components related to these factors:

- **CasSIM-h/r.** We do not distinguish users' dual roles in diffusion and use the same vectors for users' susceptibilities and influences.
- **CasSIM-up.** We remove the pre-training process for the initial user profiles and use random assignments.
- **CasSIM-x.** We remove users' topic preference vectors, e.g., $\mathbf{p}_v$ and do not consider the content of messages under diffusion, e.g., $\mathbf{x}_m$.

Table 7.7: Ablation study of popularity prediction performance on all datasets

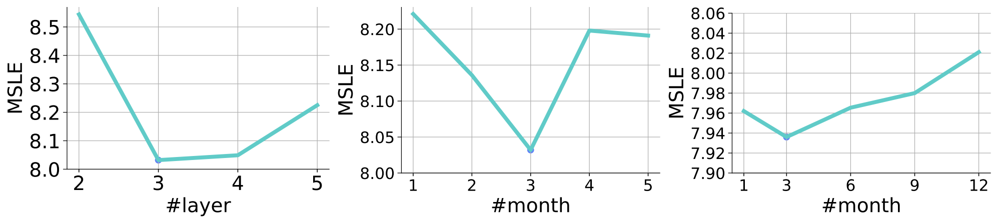
Dataset	Model	1 Hour			2 Hours			3 Hours		
		MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc	MSLE	MAPE	WroPerc
Sina Weibo	CasSIM	**2.148**	**0.221**	**19.46%**	**2.126**	**0.217**	**14.94%**	**1.919**	**0.202**	**11.04%**
	CasSIM-h/r	2.323	0.241	23.09%	2.243	0.228	16.73%	1.992	0.216	13.73%
	CasSIM-up	2.253	0.234	21.81%	2.223	0.224	15.04%	1.939	0.210	12.97%
	CasSIM-x	2.214	0.224	21.74%	2.230	0.229	16.43%	1.973	0.213	13.36%
AMINER	CasSIM	**1.272**	**0.231**	**24.51%**	**1.063**	**0.225**	**26.26%**	**1.376**	**0.126**	**14.09%**
	CasSIM-h/r	1.736	0.284	24.61%	1.403	0.278	29.63%	1.466	0.139	15.82%
	CasSIM-up	1.337	0.247	22.46%	1.370	0.231	27.17%	1.409	0.139	16.18%
	CasSIM-x	1.585	0.259	24.95%	1.355	0.246	28.12%	1.527	0.140	15.74%
Twitter	CasSIM	**8.569**	**0.443**	**27.53%**	**8.046**	**0.437**	**25.43%**	**8.032**	**0.422**	**14.76%**
	CasSIM-h/r	9.283	0.488	33.64%	9.040	0.473	30.80%	8.577	0.448	21.07%
	CasSIM-up	8.891	0.473	28.80%	8.490	0.451	27.85%	8.267	0.436	18.54%
	CasSIM-x	8.907	0.478	29.46%	9.160	0.464	28.85%	8.352	0.452	19.79%

Table 7.7 outlines the performance comparison between CasSIM and its variants in terms of popularity prediction. We have three major observations: i) CasSIM performs considerably better than its variants; ii) ignoring users' two roles in information diffusion consistently leads to the largest damage to the prediction performance; iii) except for Weibo, message content consistently ranks the second most influential component.

7.5.6 Hyperparameter Test

We examine the influence of three important hyperparameters of CasSIM. The first parameter is the number of GNN layers which can intuitively be interpreted as the number of diffusion generations. The other two parameters relate to the pre-trained user profiles. In CasSIM, we assume that users' profiles are stable over a sufficiently long time, especially for users' topic profiles, e.g., $\mathbf{p}_v$. In the previous experiments, we use the first three months' retweets in our Twitter dataset to pre-train users' susceptibility and influence vectors, and stick to them to conduct following cascade predictions. We would like to test whether this is reasonable in practice and when user profiles should be retrained. We take our Twitter dataset as an example in our investigation. We start with examining how many months in advance are needed in this pre-training process and then track the performance changes when predicting cascades in different periods after the user profile training.

Figure 7.6 shows the results. We vary the number GNN layers z from {2,3,4,5}. We can see that MSLE curve drops to the bottom when $z = 3$, then slowly climbs up when larger numbers of layers are implemented. We vary the number of months whose retweets are used for user profiles from $\{1, 2, 3, 4, 5\}$ and the result shows that the periods for user profiling can be neither too short nor too long. On our Twitter dataset, three months work the best for popularity prediction. To test the effectiveness of pre-trained user profiles, we train and test our CasSIM model on tweets in the 1, 3, 6, 9 and 12 months after the tweets used for user profile training. We can see the popularity prediction performance decreases when the trained user profiles are used to predict cascades later than 3 months. However, a closer look will reveal that the range of the change is rather small. This is consistent with our expectation that users' preferences and interests are relatively stable on social media in spite of the vast changes in social news trending.



(a) Number of GNN layers. (b) User profiling period. (c) Prediction period.

Figure 7.6: The influence of hyperparameters.

7.6 Conclusion

In this chapter, we propose a new deep learning model CasSIM which can simultaneously achieve the two most demanded cascade prediction objectives: popularity prediction and final adopter prediction. Compared to previous models, CasSIM explores the dual roles of users in diffusion processes as both receivers and distributors, and models the three basic factors in users' decision to become active: susceptibilities, influences and message contents. With effective user profiling, CasSIM successfully models the topic-specific property of susceptibilities and influences. In addition, the introduction of GNN allows CasSIM to capture the dynamics of susceptibilities and influences during information diffusion. With extensive experiments on three real-life datasets, we validate the effectiveness of CasSIM in predicting popularity and final adopters. The results show that CasSIM outperforms the state-of-the-art methods, especially when shorter cascades are observed, in both social media and other scenarios where cascades also present.

We identify a few limits of our CasSIM model which can be addressed in the future. First, we focused on messages in the form of texts. Second, CasSIM does not consider the temporal ranks between the early adopters. It is interesting to extend and test CasSIM in cascade prediction by combining other types of information in messages such as images and quotations, and further improve the performance by integrating the time stamps of early adopters.

Part IV

Public Opinion Extraction

Chapter 8

Learning Attitudes towards COVID-19 Vaccination

In the previous chapter, we focus on controlling the spread of misinformation and fake news by introducing CasSIM, a deep-learning cascade prediction model that successfully captures the dynamics of susceptibility and influence during information diffusion. As we progress to this chapter, our attention shifts to another approach to dealing with the infodemic, which is to understand public opinion and reactions to policies, particularly in relation to vaccination.

In this chapter, we present a deep learning framework that utilises text posts on Twitter to extract and constantly track users' vaccination stances. To address the impact of linguistic features such as sarcasm and irony, which are frequently used in vaccine-related discourse, we integrate recent posts from users' social network neighbours into the framework to help detect users' genuine attitudes. Based on our annotated dataset, the model instantiated from our framework achieves a 23% performance improvement in attitude extraction compared to state-of-the-art text-only models. By using this framework, we successfully validate the feasibility of employing online social networks to monitor the evolution of real-life attitudes in response to public events, such as vaccination campaigns and health policies.

8.1 Introduction

Vaccination is now unanimously accepted as an effective approach to combat the ongoing global COVID-19 pandemic, caused by the contagious coronavirus SARS-CoV-2 [MKC+21]. Despite the decreased efficacy against the infection of the recent variants, a high-level uptake of the currently available vaccines is still believed as key to restrain the numbers of severe diseases, deaths, and particularly hospitalisation, which is crucial for medical systems to remain operating as normal [ASK+22]. Regrettably, similar to the vaccines of other infectious diseases, not everyone is willing to be vaccinated [DLG+13]. The impact of *vaccine hesitancy* has been widely recognised and extensively studied in a number of countries [CPAA+21] for various groups of people, e.g., healthcare workers [SSU+21] and immigrants [AMS21].

Many related factors and their influences are evaluated and compared, e.g., education, income and gender [CPAA+21]. These scientific findings have provided policymakers with valuable references to design strategies to reduce vaccine hesitancy and fix the stagnant uptake ratios.

The success of these studies relies on the collection and accurate understanding of the public's attitudes towards vaccination. Social surveys with well-defined questions, due to their reliability and trustworthiness, have been adopted as the dominant source of public opinions in the literature. However, as conducting surveys is expensive and time-consuming, they tend to fall behind the fast development of the COVID-19 pandemic [BMM+20], and thus fail in capturing the evolution of vaccine hesitancy. Continuous tracking of public vaccination attitudes can help healthcare bodies to identify the significant fluctuations to make a timely intervention or fast capture the public's response to implemented policies. Moreover, it allows for analysing dynamic factors such as occasional social protests and misinformation propagation, in addition to the static ones like demographic profiles which rarely change in short periods.

In recent years, social media has attracted the attention of data analysts as an auxiliary data source to complement public health surveillance (PHS) [PSDN20, RRR20] and understand social events such as natural disasters [RGGG18, MZBC20, AKZ20], in spite of its inherent bias, e.g., regarding population sampling [BMM+20, ARZ20]. Due to social distancing and fear of the unknown, people spend more time than ever on social media. As social media posts have proved to encode posters' subjective opinions [JVR+20, GCH19b, RGGG18], in this chapter, we aim to leverage the enormous daily posts during the COVID-19 pandemic to extract users' vaccination attitudes and track their temporal changes.

It is widely accepted to exploit social media posts to examine public opinions including vaccine hesitancy before the pandemic with promising performance [PSDN20, RRR20]. However, we still face challenges in *analysing* and *continuously tracking* vaccination attitudes during the COVID-19 pandemic. First, the globalism of the pandemic requires a method that can deal with multilingual posts. This fails most of feature-based methods which classify posts with pre-defined textual features, e.g., through keywords, and usually focus on one single language. Second, the linguistic features such as sarcasm and irony, which are quite common in vaccination-related discourse during the pandemic, significantly impair the performance of existing models. Consider the following example: "*I wouldn't do it for their vaccine, I'm waiting for the 6G*". The user expresses his/her support for vaccination by making fun of the conspiracy theory that chips are implanted with vaccine injection. After experimenting with the state-of-the-art text feature-based classification methods, we only get an accuracy of 0.65, which is apparently not reliable enough for trustworthy analysis. Last but not least, the aim of continuous tracking prevents from exploiting community-based methods [CMK+20, JVR+20] due to the relatively stability of community affiliations.

To address the above challenges, we take advantage of the recent advances of deep learning in natural language processing (NLP), and propose a framework that can accurately classify a textual post according to the vaccination stance expressed by its originator. Recent studies revealed that a user's vaccination attitudes correlate with those of their neighbours in social networks, e.g., friends and friends of friends. For example, online social network users with negative attitudes often have social relations with users of positive attitudes [JVR+20, MMS21]. Inspired by such studies, we integrate the recent posts of a user's social network neighbours in our framework to help detect the user's genuine attitude and reduce the impact of sarcasm. To train and test models instantiated from our framework, we use the EU-vax dataset we collected. In addition to the experimental evaluation, we draw the temporal evolution of vaccination attitudes extracted from our collected tweets. We cross-validate with published social studies and manually analyse the popular social events occurring around significant changes of vaccine hesitancy levels. Our model has also been leveraged to measure individual users' vaccine hesitancy in a recent study [CCP+22b] which successfully cross-validates the consistency between Twitter and social surveys. All the validation results successfully illustrate the effectiveness of our framework, as well as the power of social media as a data source to grasp public vaccine hesitancy in practice in near real time.

Newsagents, governments, healthcare professionals and even anti-vaccine activists use social media to spread news, knowledge and suggestions to persuade or dissuade people from getting vaccinated [VBBB+20]. To showcase the practical use of our post-based attitude learning framework, to the best of our knowledge, we are the first to demonstrate that the information users perceive from social media can be used as predictors of their vaccine hesitancy changes.

Our contributions are as follows:

- We propose a framework to extract vaccination stances from textual social media posts. Our framework integrates recent posts of a user's social network neighbours to help reduce the interference of linguistic features, e.g., sarcasm and irony.
- We design models instantiating our framework. Based on our annotated dataset from Twitter, the best model can increase the performance of attitude extraction by up to 23% compared to state-of-the-art text-only models.
- Using the model with the best performance, we track the evolution of vaccination attitudes. The utility of the extracted vaccination attitudes is further validated by the consistency with published statistics and explainable significant fluctuations of vaccine hesitancy in terms of social events such as wide propagation of misinformation and negative news.
- We show a practical use of our framework by validating the possibility to predict a user's vaccination hesitancy changes with the information he/she perceives from social media.

Through this chapter, we (re-)establish the power of social media as a complementary data source in public health surveillance in spite of its inherent biases. Specifically, when exploited properly, it can provide healthcare bodies with useful information to guide or support their decision-making processes. Note that in spite of our target at COVID-19 vaccination attitudes, our framework is actually general to be used or extended to extract other pubic opinions from social media posts.

8.2 Related Work

Vaccine hesitancy is believed to be a major cause of stagnant vaccine coverage and contributor to vaccine program failure [DLG+13]. In spite of the lack of a unified definition, one widely accepted representation of vaccine hesitancy is a continuum, ranging from complete rejection of vaccines to varying degrees of scepticism [WAD14]. In this section, we concentrate on the vaccine hesitancy studies after the onset of the COVID-19 pandemic. A considerable amount of literature has been published investigating the state of vaccine hesitancy and the influential factors in different regions [KSP+21, SSS+21] for specific groups of people such as healthcare employees [BMKP21, GGA20], immigrants [ARAK+21] and college students [BND+20]. Although surveys are still the most adopted method to collect sampled populations' attitudes or stances toward vaccination [AAME+21], some recent works leverage social media as a new dimension [JVR+20, GCH19b]. Compared to self-reporting questionnaires, social media data are cost-effective to access, and more importantly, allow analysis over large populations which was not previously feasible [BMM+20, ARZ20, JVR+20].

The methods extracting vaccination attitudes from social media fall into two categories: community-based and post-based. Cossard et al. [CMK+20] found pro- and anti-vaccine users naturally cluster into communities and calculated community partitions of various communication graphs to infer users' vaccination stances. Johnson et al. [JVR+20] made use of the topics of fan pages (similar to discussion groups) on Facebook to approximate users' attitudes, and analysed the communities formed by 100 million users across the world in terms of their vaccination attitudes. Post-based methods benefit from the various types of information encoded in social media posts such as texts, labels and pictures. Gunaratne et al. [GCH19b] relied on the hashtags in tweets to approximate the vaccination attitudes in tweets. Sentiment analysis [WNY20], as part of natural language processing (NLP), aims to derive multiple types of subjective opinions expressed in texts such as political stances [EDL16, CCET16]. The introduction of deep learning leads to more powerful models that can process posts at the sentence or paragraph levels such as word2vec [MCCD13] and BERT [DCLT19, OSB22]. The sentiments extracted from texts have been used as references to study vaccine hesitancy [GADN21]. For instance, Gbashi et al. [GADN21] detected the opinions of media towards vaccines in Africa through Twitter and Google news.

Discussion. The community-based methods cannot capture the fast development of public vaccination attitudes due to the relatively stable connections between users. Moreover, community memberships are effective for analysis on the level of populations but fail to accurately derive individual users' attitudes. The post-based methods in previous studies are not specifically designed and trained for COVID-19 vaccines. As a result, they cannot capture the special linguistic characteristics of the online discourses during the COVID-19 pandemic. This is partially because of the lack of social media posts which are related to COVID-19 vaccines and annotated with vaccination stances. In this chapter, we propose a framework that can not only benefit from state-of-the-art post-based methods, but also deal with the interference of linguistic features such as sarcasm and irony in discourses related to COVID-19 vaccination. We also create the first annotated dataset of tweets which can facilitate developing future models on subjective opinion extraction. Intuitively, our framework fuses the recent online discourses of the originating user's friends (including his/her own past posts) as part of the input to help mitigate the impact of sarcasm and irony.

The previous works that are closest to our work includes [EDL16, CCET16, WNY20] which leverage auxiliary information on social media in addition to texts to classify social media posts, e.g., structure of social connections [CCET16, EDL16] and sentiment diffusion patterns [WNY20]. However, they still rely on manually selected features either for texts under analysis or the auxiliary information. For instance, Wang et al. [WNY20] used features of information diffusion trees and diffusion networks to predict the likelihood of sentiment changes when social media posts propagate among users. This likelihood is then used to adjust the classification results of text-only classifiers. Cotelo et al. [CCET16] represented social structural information with calculated community partitions and combined it with the structure embedding with word frequencies in targeted text to infer political stances. One problem is that the features effective for one type of opinions may not results in the same level of performance for other opinions. Moreover, the existing works still use word-level representations for texts and are only tested effective on a single language. Although they can be re-trained with new training data in other languages, it is not clear whether they can result in similar performance when dealing with multiple languages simultaneously. This prevents them from being applied in regions/countries with populations of various origins.

Different from the literature, we propose an end-to-end deep learning framework specifically for inferring vaccination attitudes which can automatically extract useful representation of the inputs without manual intervention. This allows our framework to be easily adopted to infer other subjective opinions when properly re-trained with new data. Considering the global requirement of monitoring vaccine hesitancy, we ensure that our framework can deal with multilingual texts.

8.3 Extracting Vaccination Attitudes from Social Media Posts

8.3.1 Problem Definition

Extracting the vaccination stance of a social media post can be technically formulated as a natural language processing (NLP) task [LPL$^+$22], i.e., classifying texts according to given class labels. In this chapter, we focus on the affective stance towards COVID-19 vaccination. Thus the set of labels is $\mathcal{L} = \{NE, PO, NG\}$ where *NE*, *PO* and *NG* correspond to neutral, positive and negative, respectively. The basic idea of text classification in NLP is first to calculate a representation of the given text which summarises its linguistic features and then output the most likely class label. Classification methods differ from each other in terms of the formats of text representation and the mapping from representations to class labels. Text classification confronts the challenge that the attitude or emotion expressed by the same words varies according to the context. For instance, the figurative usage of symptom words can fool the keyword-based classification methods and significantly deteriorate the precision of health mention classification [BJL+20]. Sentiments of the short texts with symptom words are thus introduced as auxiliary information to deal with this figurative interference.

As discussed previously, we notice that during the COVID-19 pandemic, Twitter users tend to use sarcastic or ironic expressions to describe their disagreements with those with different vaccination stances. Inspired by previous works such as [BJL+20], given a post, we aim to benefit from the originator's social relations as well as their past posts to help reduce or eliminate the interference of sarcasm and irony.

We use $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the social graph recording the social relations between users where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges between nodes. A node $v \in V$ corresponds to a social media user and an edge (v, v') indicates the existence of a relationship between two users v and v'. Note that we ignore the direction of relationships in this chapter to take into account all the neighbours of a user, e.g., both followers and followees on Twitter. Thus, $(v, v') \in \mathcal{E}$ implies $(v', v) \in \mathcal{E}$. We abuse the terms *user* and *node* in the following discussion when it is clear from the context. Let $\mathcal{N}_i^k$ be the set of neighbours of node v_i within k hops, i.e., $\{v \mid d_{\mathcal{G}}(v, v_i) \leq k\}$ where $d_{\mathcal{G}}(v, v_i)$ is the shortest distance between v and v_i in the graph $\mathcal{G}$. Note that node v_i is also in $\mathcal{N}_i^k$ as $d_{\mathcal{G}}(v_i, v_i) = 0$. We use x_i^t to denote the textual message posted by user v_i at time t. We use $\mathcal{M}_i^{<t}$ to denote the list of posts originated by user v_i before time t chronologically ordered by their post time, and $\mathcal{M}_{\mathcal{N}_i^k}^{<t}$ to represent the set of post lists of the neighbours of user v_i within k hops.

Our COVID-19 vaccination attitude classification problem can be defined as calculating the probability distribution of all labels in $\mathcal{L}$. The final vaccination stance of x_i^t, i.e., $f(x_i^t)$, is determined by the label with the largest probability. Formally, we have

$$f(x_i^t) = \operatorname*{argmax}_{stance \in \mathcal{L}} \Pr(stance \mid x_i^t, \mathcal{G}, \mathcal{M}_{\mathcal{N}_i^k}^{<t}).$$

8.3.2 A Vaccination Attitude Learning Framework

To solve the classification problem formulated previously, we propose a framework which takes advantage of the recent success of adopting deep learning in NLP and graph analysis such as text embedding and GNNs.

Figure 8.1 depicts an overview of our framework by labelling its three main components in different background colours: i) a text-level information embedding module, ii) a GNN-enhanced module, and iii) a classification module. The first module

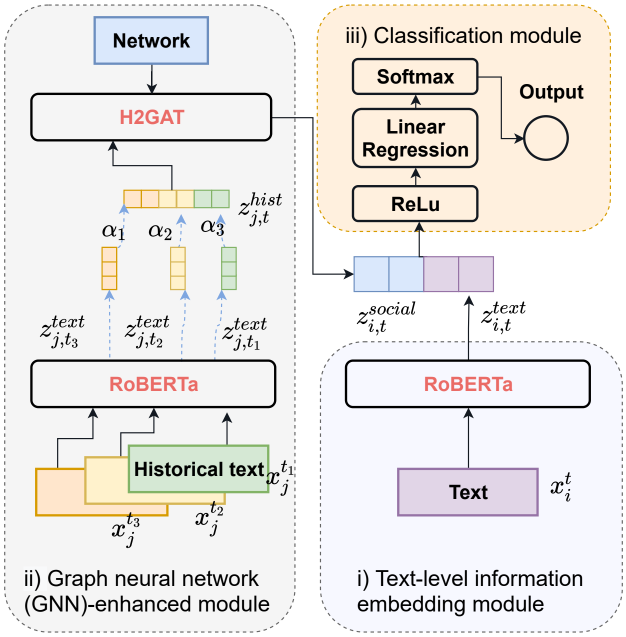


Figure 8.1: An illustration of our attitude classification framework and model.

is used to learn the linguistic representation of the targeted post while the second module summarises the linguistic features of the recent messages posted by the user's neighbours. We concatenate the outputs of these two modules as the input of the classification module. GNN [KW17] has shown its advantage in transforming graph information, including structures and attributes of nodes and edges in both academia and industry. Intuitively, it employs a message passing scheme to integrate the information of a node's neighbourhood as the representation of the node. The calculated embedding is then used for various downstream applications such as node classification and link prediction. Variants of GNN differ from each other in terms of the implementation of their message passing schemes.

Our framework can be instantiated by assembling various methods that can achieve the corresponding tasks of the modules. In the following, we present how we implement every module of the framework and justify our choices.

8.3.3 Our Model

Text-level information embedding. In order to mine meaningful information from a limited number of annotated posts, recent solutions leverage pre-trained NLP transformers to calculate the embedding for short texts [LFX+20, OSB22]. NLP transformers have been empirically evaluated in [ZLZZ21] where RoBERTa [LOG+19] is shown to outperform the competing models. Due to its high effectiveness, we adopt RoBERTa to learn text representations in our model. The model takes a textual post, e.g., x_i^t, as input, and outputs a sentiment-oriented representation vector $z_{i,t}^{text} \in \mathbb{R}^d$ where d is the pre-defined dimension of the vector. RoBERTa will be fine-tuned with the posts in the training set.

GNN-enhanced module. Given the target post x_i^t, we utilise this module to capture the linguistic features of the recent messages posted by v_i's friends within k hops before t. The definition of "recent" is flexible and depends on application scenarios. In this chapter, we select the last λ tweets before t as a user's recent tweets. Its output will be subsequently used as complementary contextual information to further ameliorate classification performance. Therefore, the input of this module consists of the social network $\mathcal{G}$ and the historical messages of user v_i's k-hop neighbours, i.e., $\mathcal{M}^{<t}_{\mathcal{N}_i^k}$. Note that the post originator's recent messages are also considered as v_i is included in $\mathcal{N}_i^k$. The output will be a text-level embedding vector that can be intuitively interpreted as a summary of useful features of friends' recent discourse.

We take two successive steps to calculate the output text-level representation vector $z_{i,t}$. We first integrate the recent posts of each user in v_i's k-hop neighbourhood as a summary of his/her vaccination discourse. In the second step, we propose a new GNN-based model named by H2GAT to aggregate the discourse of all v_i's neighbours into the social text encoding vector.

STEP 1: Text-level encoding. For each user $v_j \in \mathcal{N}_i^k$, we use his/her last λ posts in $\mathcal{M}_j^{<t}$, denoted by the list $(x_j^{t_\lambda}, x_j^{t_{\lambda-1}}, \ldots, x_j^{t_1})$ where t_m $(1 \le m \le \lambda)$ is the time stamp of v_j's last m-th post. We apply the pre-trained RoBERTa model and then obtain the corresponding list of text-level representations, i.e., $(z^{text}_{j,t_\lambda}, z^{text}_{j,t_{\lambda-1}}, \ldots, z^{text}_{j,t_1})$. There are many ways to integrate v_j's past text-level representations while distinguishing their various temporal importance, e.g., Hawkes [Haw71] and GRU [CvMBB14]. In our implementation, we adopt the dynamic-aware *position encoding* by assigning a fixed importance factor to each past post according to its position in the list. This method is simple but more effective than other competing ones in our experiments (as shown in Section 8.4). Formally, the integrated text-level representations of user v_j is calculated as follows:

$$z^{hist}_{j,t} = \sum_{m \le \lambda} \alpha_m \cdot z^{text}_{j,t_m}$$

where α_m $(1 \le m \le \lambda)$ is trainable and describes the positional relation between the past posts. Note that k and λ are predefined hyper-parameters that should be tuned manually.

STEP 2: H2GAT. It is pointed out that the *heterophily* phenomenon widely exists among online social network users [PCWF07]. This phenomenon also exists in vaccination-related discourses as the attitudes and linguistic features can be significantly different between users. Considering the *heterophily* of vaccination discussion in a user's neighbourhood, we adopt and extend a recent GNN-based method called H2GCN [ZYZ+20]. The same as other GNNs, it also has multiple layers, the ℓ-th of which can be formulated as follows:

$$H_i^\ell = \mathsf{Combine}(\{\mathsf{Aggregate}\{H_j^{\ell-1} : j \in \widehat{\mathcal{N}}_i^{k'}\} : k' \in \{1, \ldots, k\}\})$$

where $\widehat{\mathcal{N}}_i^{k'}$ represents node v_i's k'-order neighbours, i.e., nodes that have an exact distance of k' from v_i in $\mathcal{G}$. Formally, $\widehat{\mathcal{N}}_i^{k'} = \{v_j \mid d_{\mathcal{G}}(v_j, v_i) = k'\}$. Note the difference of $\widehat{\mathcal{N}}_i^{k'}$ from $\mathcal{N}_i^{k'}$. As $\mathcal{G}$ is connected, when $k' > 0$, we have $\widehat{\mathcal{N}}_i^{k'} \subset \mathcal{N}_i^{k'}$. The initial H_j^0 for each $v_j \in \mathcal{V}$ is set to $z^{hist}_{j,t}$.

Note that different from [ZYZ+20] which adopts the Aggregate function of GCN [KW17], we use GAT [VCC+18] for better performance. The output representation of v_i, denoted by $z_{i,t}^{social}$, is calculated by combining node representations of all layers:

$$z_{i,t}^{social} = \mathsf{Combine}(\{H_i^0, H_i^1, \ldots, H_i^k\}).$$

Many ways exist to implement the function Combine. We adopt the one in H2GCN [ZYZ+20] in our model which outputs the concatenation of all inputs.

By concatenating $z_{i,t}^{social}$ and $z_{i,t}^{text}$, we obtain the text-level representation vector for classification. Formally,

$$z_{i,t} = z_{i,t}^{social} \parallel z_{i,t}^{text}.$$

Attitude classification. We implement a simple two-layer structure to conduct the classification. Note the output vector has a length of $|\mathcal{L}|$. Recall $\mathcal{L}$ is the set of class labels. The first layer applies the ReLU function to each element in $z_{i,t}$ and at the second layer, we use a linear regression where $W \in \mathbb{R}^{2d \times |\mathcal{L}|}$ and $b \in \mathbb{R}^{|\mathcal{L}|}$ is the bias vector. The softmax function is applied to calculate the probability distribution over the class labels. Formally, the distribution p is calculated as follows:

$$p = \mathsf{softmax}\left(\mathsf{ReLU}(z_{i,t}) \cdot W + b\right).$$

The class label with the largest probability will be chosen as the output. We use CrossEntropyLoss as the objective function to train the entire model.

8.4 Experimental Evaluation

Evaluation setup. We set up an evaluation pipeline following the approach for traditional supervised classification [MKC+21]. In this chapter, we utilise our EU-Vax dataset. Specifically, we split labelled tweets into training (80%), validation (10%) and testing (10%) sets. The models are optimised with the training set, and the validation set is used to tune hyper-parameters. The model performance is evaluated on the testing set.

Hyperparameter settings. We pre-process the tweets by removing mentions of other users with '@', quoted hyperlinks and 'RT' which stands for "retweet". To re-confirm the overwhelming performance of transformer-based models, we also implement two traditional machine learning models that are widely used in the literature for similar tasks such as sentiment analysis: SVM and random forests. We use TfidfVectorizer [PVG+11] to convert tweets into the bag of n-gram vectors and use them as the input textual features for these two models. Grid search is utilised as the optimiser to train the parameters.

We train our model for 400 epochs and use Adam [KB15] for optimisation with the learning rate of 10^{-5} and weight decay of 5×10^{-4}. For the text encoder, i.e., RoBERTa, we adopt the implementation XLM-RoBERTa [OL20] and follow their default settings where the maximum string length, i.e., parameter d, is 128. For our GNN-enhanced module, we set the embedding dimension as 64. The neighbourhood order k which is also the number of layers and the number of historical tweets λ are important to ensure representation quality. Therefore, we conduct an empirical study to analyse the influence of these two key hyper-parameters to ensure the final performance. In Figure 8.2, we present the classification accuracy with different values of k (on the left) and λ (on the right). We observe that these two hyper-parameters indeed significantly influence classification accuracy. Our model arrives at the best performance with $k = 2$ and $\lambda = 3$.

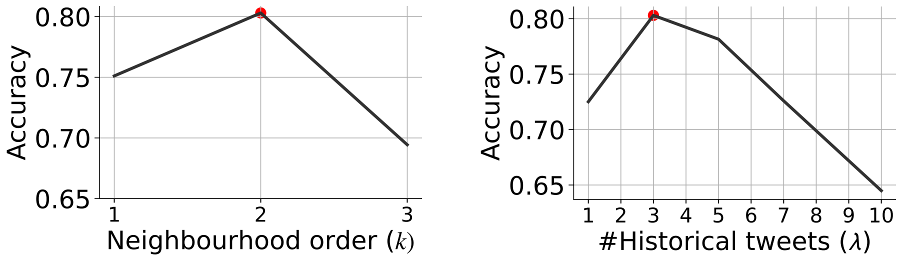


Figure 8.2: Parameter tuning for k and λ.

Experimental results. We have two objectives. One is to validate the overwhelming performance of our model over traditional machine learning models that are widely used in the literature, i.e., Random Forests and SVM. The other is to compare other possible implementations of our proposed vaccination attitude learning framework. In order to distinguish these models, we name them with two parts concatenated with '+'. The first part tells the adopted GNN model while the second part gives the method handling the temporal importance of past tweets. As all models use RoBERTa for text encoding, we do not explicitly put it in the model names. We present their performances in Table 8.1.

We reconfirmed the existing studies that transformer-based methods outperform traditional machine learning models. They can only guarantee an accuracy of around 0.4 which is at least 0.22 less than all other models. Moreover, we have four major observations that justify the effectiveness of our implementation. First, the consideration of friends' vaccination discourse increases the performance. The text-only classification model with RoBERTa only has an average accuracy of 0.65 while the other models, which are implemented with the GNN-enhanced module, achieve at least an accuracy above 0.70. Second, the vaccination discourse between friends on Twitter is actually heterophily and the choice of heterophily-aware GNN models, i.e., H2GCN and our H2GAT, can further significantly improve the performance. The next four models below RoBERTa in Table 8.1 have the same settings except for the GNN methods. Both the application of H2GCN and our H2GAT achieve an increase of about 0.04 compared to the models with GCN [KW17] and GAT [VCC$^+$18]. Third, the consideration of the temporal importance of past tweets leads to another up to 0.06 improvement. We consider four methods to combine a user's last λ tweets: MEAN, GRU, Hawkes and PE (short for positioning encoding). The method denoted by MEAN simply averages the text-level encodings. The positioning encoding method adopted in our model generates the best performance.

Table 8.1: Model performance.

Model	Precision	Recall	F1	Accuracy
Random Forests	0.4334±0.0775	0.3766±0.0716	0.4491±0.0955	0.4187±0.0606
SVM	0.4033±0.0460	0.3944±0.0196	0.4382±0.0403	0.4168±0.0110
mBERT	0.6622±0.0101	0.5769±0.0153	0.6132±0.0113	0.6466±0.0087
RoBERTa	0.6758±0.0218	0.5848±0.0232	0.6249±0.0210	0.6517±0.0348
GCN+MEAN	0.6936±0.0101	0.6932±0.0153	0.6890±0.0113	0.7033±0.0087
GAT+MEAN	0.7001±0.0091	0.7002±0.0102	0.6983±0.0078	0.7096±0.0099
H2GCN+MEAN	0.7387±0.0021	0.7144±0.0081	0.7286±0.0027	0.7412±0.0015
H2GAT+MEAN	0.7361±0.0038	0.7371±0.0007	0.7331±0.0019	0.7461±0.0008
H2GCN+GRU	0.7813±0.0025	0.7794±0.0048	0.7712±0.0027	0.7829±0.0010
H2GAT+GRU	0.7937±0.0036	0.7968±0.0021	0.7927±0.0011	0.8009±0.0008
H2GCN+Hawkes	0.7843±0.0016	0.7760±0.0038	0.7699±0.0047	0.7831±0.0025
H2GAT+Hawkes	0.7946±0.0036	0.7922±0.0010	0.7903±0.0019	0.7988±0.0023
H2GCN+PE	0.7859±0.0031	0.7813±0.0045	0.7792±0.0022	0.7889±0.0028
H2GAT+PE	**0.7948**±0.0051	**0.7954**±0.0029	**0.7936**±0.0020	**0.8030**±0.0017

Last, our extended H2GAT model outperforms the original H2GCN. Our implementation, i.e., H2GAT+PE, finally improves the text-only RoBERTa model by 23% in terms of accuracy.

Empirical complexity analysis. As the RoBERTa model is pre-trained, the models instantiated from our framework have the same complexity as the adopted GNN models. In our experiments, we conduct the training on a server with Xeon E5 CPU and Tesla V100 GPU. On average, the training time for RoBERTA is about 115 seconds for an epoch while 52.5 seconds are needed for an epoch in training the GNN-enhanced module and the classification module. What is more important is the running time of the models when processing a tweet. This will determine the practical utility of our framework in tracking public vaccination attitude in real time. We run four parallel instances of our model H2GAT+PE on the server. On average, it takes 24.68 seconds for every 1,000 tweets, which means more than 3.5 million can be processed a day. For the regions we target at, we collect in total 9 million vaccination-related tweets over two years. This implies our model is sufficiently efficient for processing posts on a daily basis.

Cross-validation. In addition to experimental evaluation, we also make use of published social studies to cross-validate our model's effectiveness. Lazarus et al. conducted a survey in June 2020, and estimated that the vaccine acceptance rates in France and Germany are 58.9% and 64.5%, respectively [LRP+21]. After applying our model to classify the tweets in the same period, we find the percentages of tweets with positive vaccination attitudes of these two countries are 42.27% and 53.12%, which are similar and retain the relative difference between the two countries. This implies that posts on Twitter can be used as a reference to fast grasp vaccine hesitancy when surveys are not available.

Vaccine hesitancy tracking and manual analysis. We draw the temporal evolution of the percentage of tweets for each selected label in Figure 8.3 on a daily basis

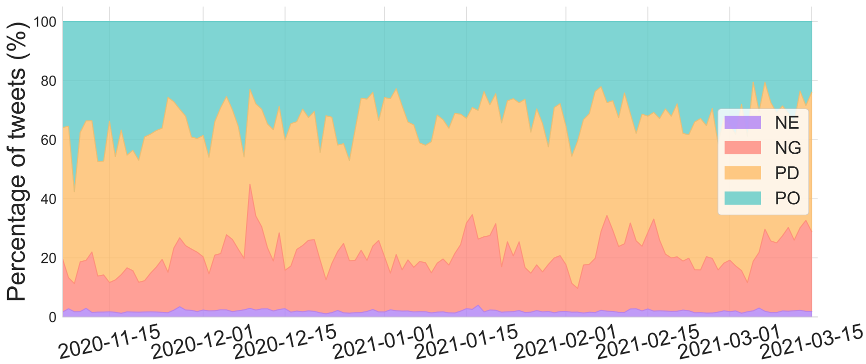
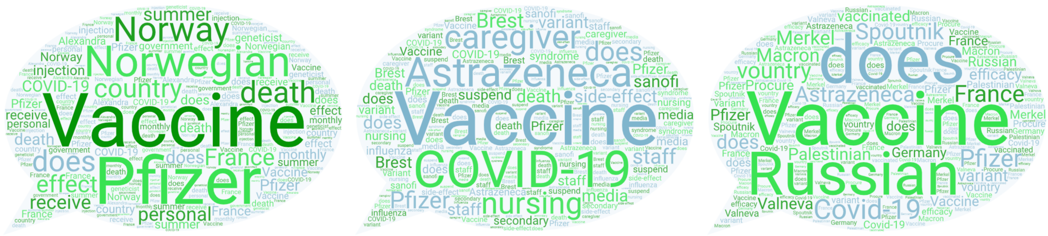


Figure 8.3: Temporal distribution of tweets with different vaccination attitude labels.

starting from November 8, 2020. Based on previous research reporting that the content of tweets is highly correlated with real-world situations [PCHG20], we make a hypothesis that real-world events may contribute to the fluctuating proportion of tweets with different vaccination stances. In vaccine hesitancy monitoring, special attention should be paid to the fluctuations of negative attitudes. We take three time points that correspond to turning points of the curve of label *NE* as examples and discuss the potential causes. Among them, two correspond to apex points where negative tweets reach local maximum percentages and one corresponds to a base point with local minimum negative tweets. We first plot word clouds in Figure 8.4 to identify the most frequently used keywords in the week around the selected points. Then we search these keywords on the Internet to identify the events that may contribute to the changes.

The first apex occurred around January 16, 2021. We notice that this surge of negative tweets attributes to the propagation of a large volume of misinformation. Take the two most dominant pieces of misinformation as examples. One said that on January 14, the Norwegian Medicines Agency reported that a total of 29 people had suffered side effects, 13 of which were fatal. The other was about the death of an Indian healthcare worker after receiving COVID-19 vaccines. The second peak happened around February 15, 2021. One piece of negative news was reported that AstraZeneca vaccines were stopped from administration after many health workers of Morlaix hospital in France suffered from side effects. This news subsequently led to anti-vaccination discussions. The base point occurred between the two peaks around February 3, 2021. From Figure 8.4(c), we find the dominant positive news that Russia started to offer other countries such as Pakistan with its vaccines.

In addition, we conducted a recent study in [CCP$^+$22b] with the proposed model to measure individual users' vaccine hesitancy levels and successfully confirmed the consistency of Twitter with social surveys in terms of vaccine hesitancy surveillance. The consistency persists across regions and along with time. From the above discussion, we can see our model can enable the use of social media data to track on a daily basis the changes of vaccination attitudes, and capture the impact of social events on public vaccine hesitancy. This may finally help the governments identify the right time to take intervention actions.



(a) Apex 1 (01/14–01/21) (b) Apex 2 (02/10–02/17) (c) The base (01/30–02/06)

Figure 8.4: Word clouds of tweets around selected points.

8.5 Use Case: Predicting Vaccination Hesitancy Changes

In this section, we illustrate a use of our vaccination attitude learning framework. Specifically, we analyse the role of the vaccination information widely spread across Twitter in affecting users' attitudes towards vaccination. Considering the comprehensiveness of vaccination discourses, we classify the most popular vaccination-related tweets into *themes* that may correlate with vaccine hesitancy. Based on users' perceived information in these themes, we forecast their vaccination attitude changes with classic machine learning models.

8.5.1 Period Selection and Theme Labelling

The participation of vaccination discourses fluctuates over time along with the occurrence of social events related to COVID-19 vaccines. We select two time periods after the start of COVID-19 vaccination campaign, in which the volume of tweets experiences significant surges compared to adjacent periods. The first period lasts for 25 days spanning from December 27, 2020 to January 20, 2021 while the second period lasts for 15 days between January 25 and February 8, 2021. These two periods involve 161,611 original tweets posted in total among which 25,449 are retweeted at least once. The total number of times of being retweeted adds up to 242,129. We encounter two challenges to continue our analysis of the impact made by diffused information: the comprehensiveness and large volume of propagated tweets. Due to the huge volume of tweets propagated over Twitter, it is not plausible to consider all of them. Previous studies show that tweets' influence follows the power-law distribution and 80% of the impacts come from 20% of the most widely spread tweets [EJR+10]. Inspired by this result, we leverage the top 25% most widely propagated tweets in every period to approximately represent the themes expressed in the diffused information. In total, we select 501 original tweets that are retweeted 78,891 times from 72.16% of the users. To deal with the comprehensiveness, with a careful examination of the selected tweets, we categorise them into themes that are considered to be responsible for the changes of vaccination attitudes. We refer to previous studies [BMZP19, DPLA+21], especially the Parent Attitudes about Childhood Vaccines (PACV) survey [OTMS+11] and the WHO Vaccine Hesitancy Matrix [Org14], and identify 11 themes that are relevant and can cover the propagated tweets (see Table 8.2 for explanation and examples).

Table 8.2: Diffused information themes and examples.

Theme	Description	Example (Translated to English)
Positive news	Positive news about vaccines & vaccination	Pfizer/BioNTech's vaccine would be effective against the new British variant of COVID19.
Negative news	Negative news about vaccines & vaccination	Portugal : She dies 2 days after the vaccine (at 41 years old). Her family asks for explanations
Distrust in government management	Doubt about the trustworthiness of governments or medical institutions, e.g., regarding the daily update of statistics	They have lied to us so much about masks, chloroquine, contagion in children, that it will be difficult to trust them the day they will tell us about a harmless vaccine.
Dissatisfaction with politics/policies	Unsatisfactory views of politics/policies, such as ineffective vaccination programs.	I am opposed to mandatory vaccination because all of the world's health organizations say that it is not the right way for a vaccine to spread.
Perception of the pharmaceutical industry	Perception that pharmaceutical manufacturers pursue only economic interests rather than public health interests	Pfizer's CEO sold 60 percent of his shares when the Covid vaccine was announced. When the CEO sells, it stinks
Conspiracy	Content that describes the event as the secret acts of a powerful, malevolent force.	18 months they've been on the vaccine ???? When did they know there would be a Covid 19 "pandemic" ????
Beliefs, attitudes about health and prevention	Personal views on vaccines and the immune system, e.g. homeopathy, natural immunity, alternative therapies.	There is no point in a generalized vaccine for a disease whose mortality is close to 0.05%.
Positive personal expression	Personal expression of positive attitude towards vaccines	We have a new weapon against the virus: the vaccine. Hold together, again.
Negative personal expression	Personal expression of negative attitude towards vaccines	Why could actually 1.5 billion Chinese get healthy without vaccination, and with us it only works with vaccination...?
Positive information	Positive expressions about vaccines from healthcare professionals	#COVID19 #vaccinationHow does an mRNA vaccine work?
Negative information	Negative expressions about vaccines from healthcare professionals	My daughter, a nurse at the AP-HP, on the vaccine "Ah ah ah! They don't even dream about it, they start with the old ones so that we can attribute the side effects to age".

We ask two of the 10 hired annotators to manually annotate the selected tweets with their corresponding themes. The Cohen's Kappa coefficient $k = 0.82$ implies a high rate of agreement between them.

8.5.2 Predictability of Vaccine Hesitancy Changes

Handling retweets and quotations. In addition to original posts, retweets and quotations also take up a large proportion of a user's historical posts. For quotations, a user added some comments which may express opposite opinions to that of the quoted one. Therefore, we only use the texts users upload as valid posts encoding users' vaccination stances. Although retweets cannot fully represent a user's own opinion, the behaviour of retweeting itself indicates some sort of agreement with the ideas expressed in the message retweeted [DLZ+15]. Based on this idea, we take retweets into account when calculating an individual user's vaccine hesitancy. The same approach is also adopted in the vaccination attitude tracking discussed in the previous section.

Quantifying individual vaccine hesitancy. We measure the vaccine hesitancy of an individual user according to the tweets posted or retweeted by the user in a time

Table 8.3: Model performances for attitude change prediction.

Model	Precision	Recall	F1	Accuracy
SVM	0.7374	0.7382	0.7345	0.7477
Naive Bayes	0.6468	0.6559	0.6427	0.6658
Random Forest	0.6811	0.6838	0.6795	0.6958
XGBoost	0.7232	0.7229	0.7198	0.7342
GBDT	**0.7533**	**0.7516**	**0.7498**	0.7603

interval. Formally, it is calculated as: $\frac{N_p(v)-N_n(v)}{N_p(v)+N_n(v)}$ where $N_p(v)$ denotes the number of posts with positive vaccination attitudes of user v during the selected interval, and $N_n(v)$ is the corresponding number of tweets with negative attitudes. Considering our purpose being idea validation, we do not distinguish the various significance of original posts and retweets.

For each selected period, we use the tweets posted 14 days before and after the period to evaluate individual users' hesitancy levels and see how they change. In order to ensure the reliability, we only consider the users who posted or retweeted at least 3 tweets. If a user's vaccine hesitancy experiences a change smaller than 0.05, we consider the user's attitude *unchanged*, otherwise, *increased* or *decreased* depending on the change direction.

Modelling perceived information. A Twitter user perceives information from the tweets retweeted or originally posted by his/her direct friends. As our focus is the information widely diffused on Twitter, we use a vector $I_u = (c_1, c_2, \ldots, c_m)$ to approximately represent a user u's perceived information where c_i is the number of popular tweets a user receives from followers in i-th theme. As we have 11 themes, $m = 11$ in our analysis.

Model evaluation. We make use of various standard machine learning methods to predict the change of a user u's vaccination attitudes with the input of I_u. The methods consist of SVM with rbf kernel ($C = 1$, $\gamma = 0.1$), Naive Bayes ($\alpha = 1$), Random Forest (100 trees with maximum tree depth of 5), XGBoost (100 trees with maximum tree depth of 4) and GBDT (100 trees with maximum tree depth of 5). Table 8.3 shows the performance of these methods. All numbers are averaged over 5 training sessions. We can see all the methods can achieve reasonably good prediction performance and GBDT outperforms the rest models with an accuracy of 0.76. When we consider additional factors such as users' vaccine hesitancy levels before the periods, the accuracy can be improved to 0.86.

Discussion. These results show that we can make accurate predictions with users' perceived popular information. Since we have empirically illustrated the plausibility to use social media posts to track public vaccination attitudes, the results imply that the diffused information on social media like Twitter could be used as indicators to forecast the changes of vaccine hesitancy levels. As repeated many times, although such predictions cannot achieve the same level of trust as social surveys, they provide decision-makers with a method to quickly understand and get prepared for the potential damage of misinformation or compare different vaccine hesitancy intervention strategies over social media.

8.6 Conclusion and Discussion

In this chapter, we propose a deep learning framework to learn vaccination attitudes from social media textual posts. Although vaccination attitudes extracted from social media cannot be as accurate and reliable as conventional social surveys, our framework allows for continuously tracking the fast development of public vaccination attitudes and capturing the changes that deserve specific attention in time. By leveraging friends' vaccination discourse as contextual information, our model successfully reduces the interference of linguistic features such as sarcasm and irony. Our model instantiate from the framework improves the state-of-the-art text-only method by up to 23% in terms of accuracy according to our manually annotated dataset. With cross-validation with published statistics and manually analysis, we further validate the effectiveness of the model to capture public vaccine hesitancy in real life. After identifying 11 themes from widely diffused information on Twitter, with the help of our model, we validate the predictability of users' vaccine hesitancy changes by the information they perceived from social media. This show a potential use of our model in practice. Through this chapter, we establish again the power of social media data in supplementing public health surveillance, especially in combating infectious virus like COVID-19.

Limitations and future work. We have three main limitations to address in future. First, we primarily focus on Twitter which potentially induces bias in our data and analysis. Thus, it is important to extend our work to other social media platforms such as Facebook and Instagram, and cross-validate our results. Second, we only analyse users' affective vaccination stances (i.e., positive, negative and neutral), which can only be used as an indicator of users' intention of getting vaccinated. It will be interesting to look deeper into users' tweets for a longer time and identify underlying determinants that lead to vaccination acceptance. Third, we use only the top 25% most widely spread tweets as representatives to extract the themes of diffused information partly limited by manual annotation. Some information in certain themes may be missed. As an interesting future work, we will develop effective NLP models to learn different tweet themes automatically.

Ethical considerations. This work is based completely on public data and does not contain private information of individuals. Our dataset is built in accordance with the FAIR data principles [WDA+16] and Twitter Developer Agreement and Policy and related policies. Our release of the dataset is also compliant with General Data Protection Regulation (GDPR). To conclude, we have no ethical violation in the collection and interpretation of data in our study.

Chapter 9

Consistency of Social Media with Surveys

In the previous chapter, we present a deep learning framework to track users' vaccination stances using text posts on Twitter, successfully demonstrating the potential of online social networks to monitor real-life attitudes in combating the infodemic. However, we recognise that social media data, including data on vaccine attitudes and indecision, may be subject to potential biases and errors stemming from measurement, coding, and missingness. In this chapter, our focus shifts to validating whether social media data can complement social surveys for monitoring the extent of public hesitancy about the COVID-19 vaccine and exploring methods to correct potential biases. We utilise recent advances in artificial intelligence to present a framework for estimating individuals' vaccine hesitancy based on their social media posts. Using vaccine-related tweets from our Twitter dataset, we compare vaccine hesitancy levels measured with our framework against those collected from multiple consecutive waves of surveys. We successfully validate that Twitter can be used as a data source to calculate public acceptance of COVID-19 vaccines consistent with national and regional level surveys. Furthermore, this consistency persists over time, although it varies across socio-demographic subgroups.

9.1 Introduction

The last three have seen the impacts of the unprecedented global COVID-19 pandemic and the infodemic on public health and the economy. Thanks to the successful vaccination program, our societies are gradually reopening and going back to the pre-pandemic states. As of August 2022, 68.1% of worldwide populations have been fully vaccinated. This milestone cannot be achieved without fast and accurate understanding of the opinions and responses of general populations towards COVID-19 vaccines and their changes over time. For instance, it allows for identifying the right intervention time and evaluating the effectiveness of deployed measures.

Social media has shown its strengths in complementing conventional surveys to study vaccine hesitancy [NCP+20]. Social media overcomes the decreasing response rates of surveys and provides a cost-effective way to reach a significantly larger population [CCZP22c]. In addition, it allows for capturing the evolution of public opinions over time, especially, in case of emergent incidents such as a sudden outburst of misinformation when there is no sufficient time for conducting surveys. In spite of these advantages, the results derived from social media are often questioned mainly because of three inherent sources of errors: *measurements*, *coding* and *missingness* [HM17, Bak17]. Measurement errors are incurred by the fact that social media users may not express their real attitudes in their posts while coding errors come from the deficiency of methods in capturing public opinions. Missingness is caused by non-representative social media users, namely, not all people express their opinions online. For instance, Twitter is more favourable to young users while Facebook attracts the elders [RBZ20]. **Our contributions.** We aim to address these challenges confronted in measuring the levels of public vaccine hesitancy with Twitter, one of the most widely used sources of social media data [Sca18, NCP$^+$20, CPAA$^+$22]. Unlike existing works examining correlated factors [NCP$^+$20], our purpose is to exemplify that with properly designed methods, individuals' vaccine hesitancy can be *accurately* measured from social media and the estimation is consistent with surveys *continuously over time* and *across countries and regions*. To the best of our knowledge, this is the first attempt to study the temporal consistency of social media with surveys regarding vaccine hesitancy.

We perform a cross-validation by making use of the social survey of multiple waves we conducted and the collected 745,661 tweets related to COVID-19 vaccines from three Western European countries. We take advantage of recent advances in natural language processing techniques, and quantify individuals' vaccine hesitancy based on their attitudes expressed in textual posts. In order to overcome the missingness errors caused by non-representative Twitter users, we show that with three socio-demographic attributes, i.e., gender, age and political ideology, the demographic selection bias can be effectively corrected. When designing our framework, we consider its applicability in a global pandemic like COVID-19 and ensure it can be used in multilingual environments.

With comprehensive analysis, we successfully validate that Twitter is able to give close estimation of vaccine hesitancy to surveys. This closeness persists at a similar level across geographical regions and over time. The large Pearson correlation coefficients indicate at least a strong correlation between the results from surveys and Twitter. We also show that the consistency varies among different socio-demographic groups. Our research re-established the power of Twitter to act as a complementary source to continuously monitor public vaccine hesitancy in COVID-19 and future health crises of similar types.

9.2 Related Work

Since the outbreak of the COVID-19 pandemic, great efforts have been devoted to studying the potential of social media in understanding the public's hesitancy in the fast developed vaccines [CPAA$^+$22, WW20, AAOA21, SGE$^+$22], based on the pre-pandemic success in studying public opinions [Sca18, JGS$^+$20, NCP$^+$20, PVPCBH$^+$21]. For instance, Cascini et al. [CPAA$^+$22] reviewed the literature during the COVID-19 pandemic about how diffused information on social media impacts vaccination attitudes. In general, previous works aim to study the correlation of social media users' online activities, e.g., information perception, to vaccine hesitancy. According to sources of online digital traces, the related work falls into two categories. The first category makes use of questionnaires or public polls to collect participants' usage habits of various social media platforms as well as their vaccination attitudes. For instance, with a survey of 504 participants, Alfatease et al. [AAOA21] observed the dependence between social media usage and willingness to accept vaccination in Saudi Arabia. Wilson et al. [WW20] revealed the correlation between online disinformation campaign and activity organisation on social media to vaccine hesitancy. The second category leverages tools such as stance detection to infer various features of online activities from social media data of various formats including hashtags, hyperlinks and textual posts. For instance, Shaaban et al. [SGE$^+$22] studied vaccine acceptance with positions and tones of comments on various social media platforms. Lyu et al. [LWW$^+$22] inferred user demographics as well as vaccine attitudes through a text-based machine learning approach, and analysed vaccine acceptance among people with different demographic characteristics.

Three characteristics have been well accepted as the advantages of social media over surveys, i.e., *volume*, *velocity* and *variety* [RSM22] and promoted social media data as a complementary or alternative source of public opinions. However, the inherent limits such as the bias of population coverage and accuracy of extracted opinions, inevitably cause doubts about claims drawn from social media [SPG$^+$16]. Several attempts have been conducted to study the reliability of social media data in studying public opinions by comparison to surveys [DZLL17, Sca18, NCP$^+$20, ABKK20]. Davis et al. [DZLL17] compared the sentiments of tweets to the polls about public opinions of the Obamacare act and showed the comparability of Twitter public opinions with survey results. Scarborough [Sca18] illustrated the correlation of tweet sentiments to gender attitudes. Amaya et al. [ABKK20] evaluated three types of errors that generate the difference between social media and public polls.

Identified challenges. Few existing works study how and whether individuals' vaccine hesitancy can be directly estimated with digital traces on social media, and whether the estimation is consistent with surveys, especially over time. Although a number of factors have been revealed to be correlated, they can only be interpreted as indicators but not a precise estimation. Without a proper cross-validation, it is unclear whether social media can be used for real-time vaccine hesitancy monitoring as suggested [PCHG20]. The work most related to this chapter is [GLG$^+$22], which compares existing selection bias correction methods with demographic attributes extracted with machining learning models. Different from this chapter, it aims at public health status and does not study the consistency of the predictability of online discourses over time.

Table 9.1: Statistics of survey participants and Twitter users.

	#Survey participants	#Twitter users	#Tweets
Luxembourg	474	1,764	28,148
Germany	501	13,390	270,695
France	711	26,562	446,818

9.3 Survey and Twitter Data

Survey. We conducted a panel survey with Qualtrics to cover representative samples of adults aged 18 and above from France, Germany, Italy, Spain, Luxembourg, and Sweden. In the first wave, stratification was used to ensure each national sample accurately represented the distribution of gender, region, and age, while other socio-economic factors such as race were not considered. The study received ethical approval from the University of Luxembourg's Ethics Review Panel, and all research was conducted according to the applicable guidelines. All participants provided their informed consent. The survey, which takes roughly 25 minutes to complete, is conducted online. Respondents were invited to fill in online questionnaires including questions about their living conditions, mental health and opinions about vaccination. More information can be found in this paper [BCDL22]. The purpose of the survey is to collect respondents' status in the pandemic to study the impact of the Covid-19 pandemic. We select three adjacent countries, i.e., Germany, France and Luxembourg, as our research objects because of their synchronised vaccination policies and close economic connections. Moreover, the diverse origins of the people are also representative for the worldwide populations. Our survey is conducted in multiple waves at intervals of approximately 4 months. During the waves in June and October of 2021, and March 2022, we consecutively asked about individuals' vaccination attitudes through the following question:

Have you been vaccinated against COVID-19?
1 *Yes*
2 *No, but I plan to*
3 *No and I do not plan to*

More than 8,000 individuals participated in the first wave. However, only part of them participated consistently in the following waves. As one of our purposes is to test whether Twitter can capture the changes in individuals' vaccination attitudes over time, we only keep the participants that responded in all the three waves. Table. 9.1 shows the statistics of our survey data.

Vaccine hesitancy evaluation. The vaccine hesitancy is calculated as the proportion of the participants marking the third option. Let $N_i^{\ell,t}$ ($i \in \{1,2,3\}$) be the number of respondents from a given region/country ℓ ticking the i-th option, in a given survey wave t. As the first two options indicate acceptance of COVID-19 vaccines, the vaccine hesitancy of a region ℓ in the survey wave t, denoted by VH_ℓ^t, is calculated as: $VH_\ell^t = \frac{N_3^{\ell,t}}{N_1^{\ell,t}+N_2^{\ell,t}+N_3^{\ell,t}}$.

Twitter Data. We construct a dataset of Twitter users located in the target countries—Luxembourg, Germany, and France—who are actively involved in vaccine-related discussions during the periods corresponding to the three selected survey waves based on the EU-Vax dataset. We retain only 49,791 users located in Luxembourg, Germany, and France. We further download the following relationships for each user to construct a social network represented as a directed graph. A vertex represents a Twitter user, while an edge from vertex v to vertex v' indicates that the user corresponding to v follows the user represented by v'. We only keep users who post at least 5 tweets in each target month to ensure the reliability of the vaccination attitude calculation. Note that we do not consider retweets because they are more likely to carry the intent of the originating user than quotes and original tweets. Because Twitter contains accounts of organisations such as newspapers and medical departments, we remove the accounts of these organisations to ensure that vaccination attitudes belong to the general population. We apply the pre-trained model in citeWangHAGHFJ19 to identify such accounts. In total, we remove 5,070 organisational accounts. Finally, we have 1,764 Twitter users from Luxembourg, 13,390 from Germany, and 26,562 from France, which is almost 30 times the number of survey respondents.

9.4 Measuring Vaccine Hesitancy with Twitter

We select Twitter as the source of vaccination attitudes by assuming Twitter users tend to express their real opinions about COVID-19 vaccines. In other words, we hypothesise the *measurement* error is acceptable when Twitter data is used. In this section, we describe how we handle the other two inherent errors with three sequential steps. The first step targets at reducing the coding error by proposing a measurement of vaccine hesitancy while the other two steps are to correct the missingness error. Note that our aim is not to eliminate the errors but to mitigate the impact of these errors. We adopt widely accepted methodologies to avoid statistic manipulation and thus ensure the generality of our framework.

9.4.1 Measuring Individual Vaccine Hesitancy

A significant amount of research has been devoted in understanding public opinions from social media posts, varying from word-level [GDM$^+$21, BKZY21] to data-driven approaches [NTFW22, ZXL$^+$22]. We use one recent deep learning model which is specifically designed to infer COVID-19 vaccination attitudes expressed in tweets and overwhelms existing models in classification accuracy [CCZP22b]. Another reason for our selection is its power of dealing with multilingualism which is essential for the global demands of vaccination attitude monitoring. Intuitively, the model uses RoBERTa [OL20], the most popular pre-trained embedding method, to calculate the representation of tweets, and leverages social connections to integrate the recent tweets of each user's friends with a variant of H2GCN [ZYZ$^+$20]. The model takes the text representation of a tweet under analysis and the integrated embedding of the recent discourse of the originating user's friends as input and output the possibility that the tweet is classified into attitudes corresponding to *vaccine support*, *anti-vaccine*, and *neutral*. We retrain this model with the release annotations [CCP22a], our constructed social network and collected tweets. The resulted model achieves an accuracy of 0.80 and Marco F1-score of 0.79.

Vaccine hesitancy calculation. To estimate an individual user's vaccine hesitancy, we leverage the vaccination attitudes expressed in his/her tweets. Inspired by the measurements in [BGRM11, CCZP22a] which are originally proposed to evaluate subjective well-being, we construct the measurement of vaccine hesitancy. Intuitively, users who post more tweets supporting vaccination are considered more acceptable of COVID-19 vaccines and thus more likely to get vaccinated. Formally, let $N_s(u)$, $N_a(u)$ and $N_{neu}(u)$ be the numbers of tweets of user u posted in a given period t (i.e., June and October 2020, and March 2022 in our analysis), indicating his supportive, anti-vaccine and neutral stance about COVID-19 vaccination, respectively. The vaccine hesitancy of u, denoted by $VH^t(u)$, is calculated as:

$$VH^t(u) = 1 - \frac{VA^t(u)+1}{2}, \text{ where } VA^t(u) = \frac{N_s(u)-N_a(u)}{N_s(u)+N_a(u)} \cdot \left(\frac{N_s(u)+N_a(u)}{N_s(u)+N_a(u)+N_{neu}(u)}\right)^{\frac{1}{2}}. \tag{9.1}$$

Note that $VA^t(u)$ is extended by Chen et al. [CCZP22a] from [BGRM11] with *neutral* messages considered by adding a scaling factor. It actually measures the vaccination acceptance of user u and is between -1 and 1. We first normalise it to the domain between 0 to 1 and then compute the complement as the level of vaccine hesitancy. As a result, a user's vaccine hesitancy of 1 indicates total opposition against vaccination and 0 means complete belief in COVID-19 vaccines.

9.4.2 Inferring Socio-demographic Profiles

The missingness error is related to the socio-demographic selection bias which is a well-recognised inherent limit of social media [RBZ20]. One way to correct such bias is to adjust each individual's vaccine hesitancy level by multiplying a factor that is calculated according to the difference between the distributions of social media users and the general population. Despite the large number of out-of-box methods inferring various demographic attributes such as education [GLG$^+$22] and income [MMSS19], few can be used in our analysis due to their limitation in dealing with multilingual texts. Existing methods, especially the ones based on machine/deep learning, can be extended to multilingual data with well-annotated multilingual data for training and testing. However, due to the privacy protection regulations such as GDPR, it is challenging to collect people's social media accounts and their corresponding socio-demographic information.

In order to ensure our framework applicable globally, we need to select the demographic attributes that can be inferred with multilingual data and effectively mitigate socio-demographic selection bias. Considering these two requirements, we select three socio-demographic attributes, i.e., age, gender and political ideology. We detail the models adopted or extended to infer these three attributes.

Age and gender. We use the multi-modal deep neural model M3 [WHaPAG+19] to infer the age and gender of Twitter users. These two attributes are simultaneously inferred by M3 with users' account name, screen name, self-descriptive description and profile image. A user's age falls into one of the three ranges: $19-29$, $30-39$ and ≥ 40. Multilingual textual inputs are first translated into English word by word, and the $3,000$ most frequent characters are selected to calculate users' embedding. Although the performance of the M3 model has been confirmed by previous studies [ZXL+22, DLP+20], we still construct a sample dataset to test its performance on our collected Twitter data. This sample dataset consists of 100 randomly selected users, whose ages and genders are manually annotated by two annotators. The annotated labels are highly agreed between the two annotators with large Cohen's Kappa coefficients ($k = 0.95$ for gender and $k = 0.81$ for age). When tested on our sample dataset, the M3 model achieves a Macro F1 score of 0.92 and an accuracy of 0.91 for age classification. For gender classification the Macro F1 score is 0.78 and the accuracy is 0.75.

Political ideology. We infer Twitter users' political ideology by the Multi-task Multi-relational Embedding model (TIMME) [XSX+20]. TIMME utilises the heterogeneous types of relationships between Twitter users including '*following*', '*retweet*', '*mention*' and '*like*' in conjunction with tweets to infer users' bipolar political ideologies, i.e., left and right. As TIMME is designed for English-only data, we have to re-train it on a multilingual dataset. One distinguishing feature of TIMME is that it can be trained only with a sparsely annotated training set. This allows us to prepare a new training set of a relatively small size from our collected Twitter data with the public Twitter parliamentarians dataset [vVTU20]. The dataset [vVTU20] contains manually verified parliamentarians from 26 countries, including France, Germany, and Luxembourg, with their names, political ideology, and Twitter IDs. The political ideology is evaluated in a scale from 0 to 10. We manually update the list of parliamentarians of the three countries by i) adding new politicians who joined after the data release, and ii) updating the obsolete Twitter IDs. In total, we constructed a training dataset of 1,021 parliamentarians. We encode the political ideology scores into left, centre and right. Specifically, a score smaller than 4 is encoded as left while a score larger than 6 is encoded as right. A score between 4 and 6 belongs to centre.

We conduct two extensions to TIMME. First, we extend TIMME to a triple classification model (with '*centre*' added) by replacing the binary cross-entropy loss function with a categorical cross-entropy loss function. Second, to enable TIMME to handle multilingual texts, we replace the word-level embedding with RoBERTa [OL20]. We train the extended TIMME model with the parliamentarian dataset and achieve an accuracy of 0.77 and Marco F1-score of 0.78.

Socio-demographic selection bias in our Twitter data. Figure 9.1 presents the socio-demographic distributions of the survey participants and our collected Twitter users in the three targeted countries. A significant difference between the two distributions in every country is observed. Moreover, the difference varies from one country to another. When measured by KL-Divergence, we have the distances

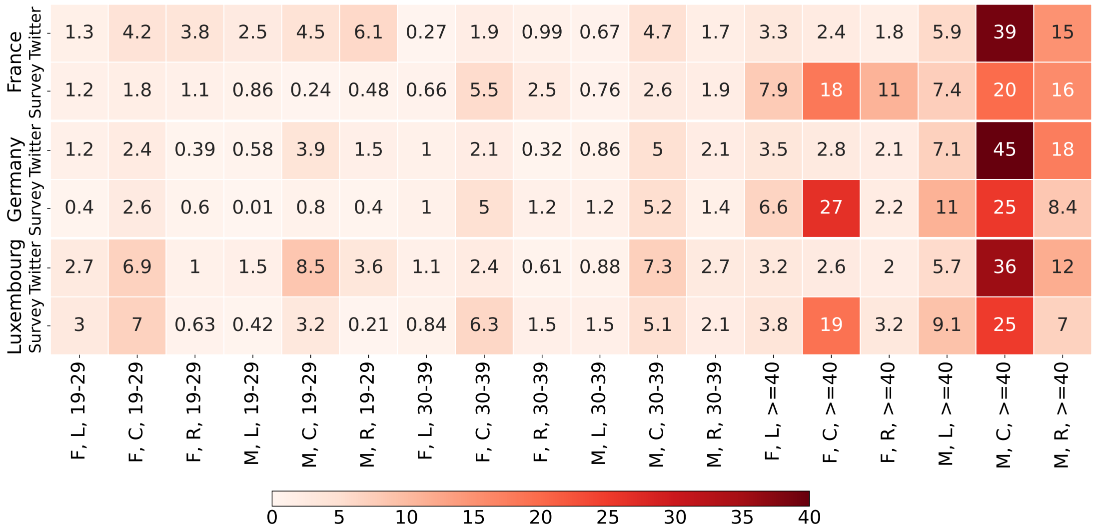


Figure 9.1: Population distribution according to age, gender (F: female, M: male) and political ideology (L: left, C: centre, R: right).

of 0.52, 0.29 and 0.38 in France, Germany and Luxembourg. This shows the non-representation of Twitter users and the necessity of correction.

9.4.3 Correcting Socio-demographic Selection Bias

The general idea of socio-demographic correction is to re-weigh non-representative samples' vaccine hesitancy with scalars calculated according to their percentage differences from the representative population. Let ϕ_u be the socio-demographic attributes of user u in the form of age, gender and political ideology. Suppose $\Pr_\ell^S(\phi_u)$ ($\Pr_\ell^T(\phi_u)$) be the percentage of survey participants (Twitter users) with the same demographic attribute as u in region ℓ. We use $\mathcal{U}^\ell$ to denote the set of users located in the given region ℓ. Thus, the corrected average vaccine hesitancy of region ℓ in time period t is

$$\widehat{VH}_\ell^t = \frac{1}{|\mathcal{U}^\ell|} \sum_{u \in \mathcal{U}^\ell} VH^t(u) \cdot \frac{\Pr_\ell^S(\phi_u)}{\Pr_\ell^T(\phi_u)}. \tag{9.2}$$

According to the availability of the joint distributions (i.e., $\Pr_\ell^S$ and $\Pr_\ell^T$), we can use different implementations. When the joint distributions are available, the correction is called *post-stratification*. When the two joint distributions are not both available, *naive post-stratification* [LW17] and *Raking* [DSS93] are applicable. The former assumes independent socio-demographic variables while Raking adopts an iterative approach to adjust each sample's marginal to match the representative population distribution.

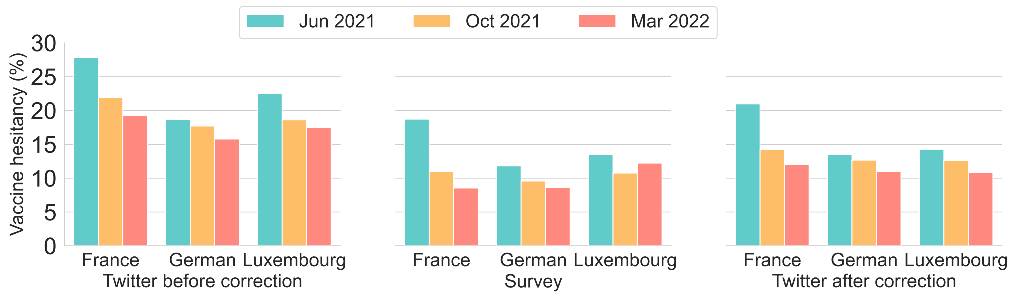


Figure 9.2: Vaccine hesitancy across countries.

9.5 Cross-validation

Our objective of the cross-validation is to test whether the vaccine hesitancy inferred from Twitter with our framework is similar to that collected from the survey and whether the similarity, if validated, persists over time and across regions/countries. As vaccine hesitancy varies among countries and regions [HTD+21], we study both the country- and region-level vaccine hesitancy. Note that we use post-stratification as the correction method because of the availability of the joint distributions of the three selected socio-demographic attributes.

9.5.1 Vaccine Hesitancy Across Countries

In Figure 9.2, we show the average vaccine hesitancy in Germany, France and Luxembourg in the three survey waves calculated with Twitter data and surveys. In general, Twitter users are more negative about vaccination. In addition, we have three other observations. First, we observe similar changes of vaccine hesitancy over time. This complies with the latest updates derived from surveys/polls around the world which indicate a decreasing trend of vaccine hesitancy [BKG+22, BBCC22]. This trend presents in all the three countries even without the socio-demographic selection bias correction. Special attention should be paid to the survey of Luxembourg in the last wave. The vaccine hesitancy increased by about 0.015 compared to the second wave. With a manual check, we notice that about 8 participants changed their choice from '*No but I plan to*' to the third option '*No and I do not plan to*'. This increase is actually not consistent with the continuous increase of vaccinated population since October 2021 and may be caused by the relatively smaller number of respondents in Luxembourg. Second, when ordered by their vaccine hesitancy, the countries have similar rankings. Residents in France are relatively more reluctant to get vaccinated compared to the other two countries and people in Germany are more favourable to vaccination. Our third observation is that without bias correction, the vaccine hesitancy calculated with Twitter data is rather different from the survey while correcting selection bias can significantly reduce the difference and ensure a similar estimation. Without the bias correction, the average vaccine hesitancy differences of the three countries are 0.083, 0.089 and 0.077 in the three waves, respectively. The differences drop by more than 70% to 0.019, 0.027 and 0.024 after correction.

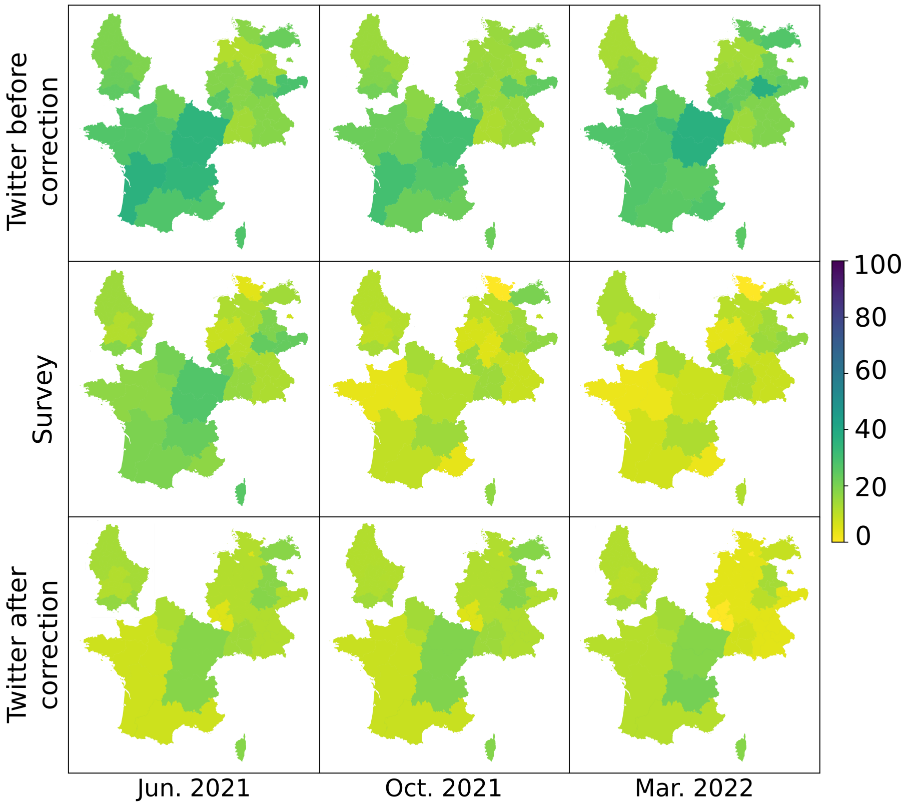
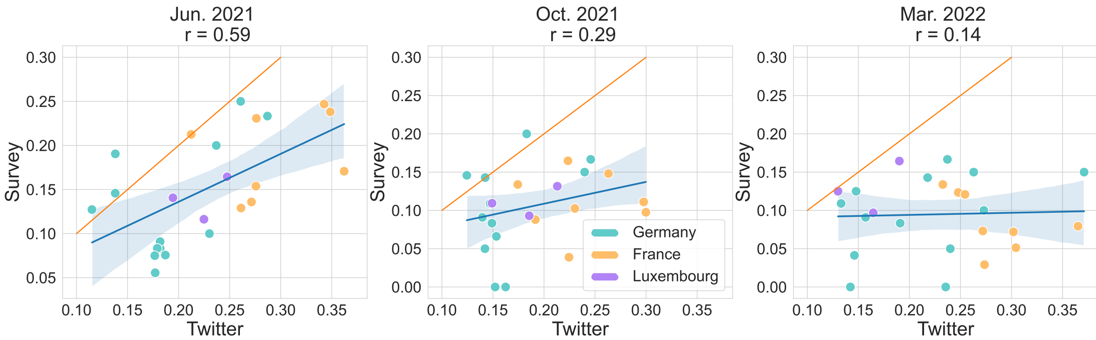


Figure 9.3: Vaccine hesitancy across regions from Twitter and survey.

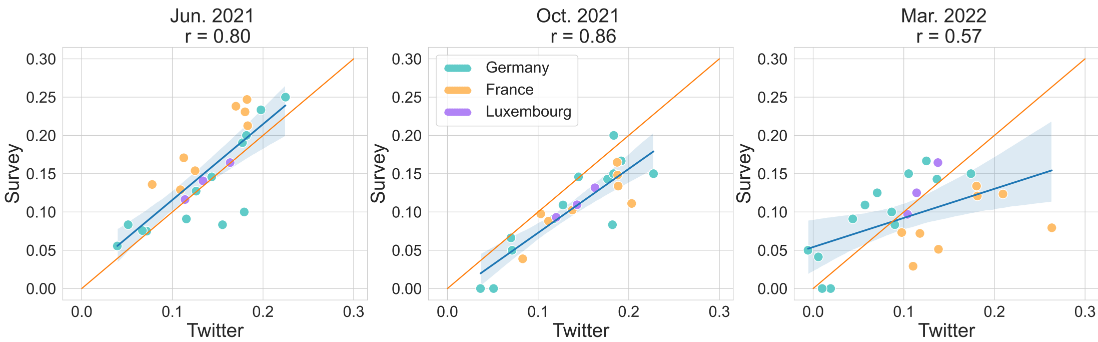
9.5.2 Vaccine Hesitancy Across Regions

We obtain the regions according to the administrative divisions of the three countries. As the distribution of our survey respondents is not uniform across regions, to ensure the reliability of the vaccination reluctance calculated from our survey, we remove those regions with fewer than 11 participants. In total, we obtain 24 regions including 8 administrative regions of France and 13 states of Germany. Due to the small size of Luxembourg communities, we divide Luxembourg into three regions: north, south and central.

Figure 9.3 illustrates the region-level vaccine hesitancy in the three survey waves. We can clearly see that after bias correction, Twitter data can reflect similar levels of vaccine hesitancy to the surveys despite the relatively big differences in certain regions. This similarity persists in all the three waves. In Figure 9.4, we further show the Pearson correlation coefficient r between the hesitancy levels calculated from Twitter and surveys before and after the socio-demographic bias correction. Each point corresponds to a region with a coordinate (x, y) where x is the vaccine hesitancy derived from Twitter and y is that from the surveys. The orange line is composed by the points where $x = y$. After correction, the Pearson correlation coefficients reach over 0.80 in the first two waves, which indicates a *very strong* correlation according to the well-accepted standard [Ako18]. In the third wave, the correlation strength decreases to 0.57 which is still interpreted as *strong*. After a closer look, we observe that the points that are relatively far from the orange line



(a) Twitter vs. Survey before bias correction ($p < 0.05$).



(b) Twitter vs. Survey after bias correction ($p < 0.05$).

Figure 9.4: Region-level correlations of vaccine hesitancy between Twitter and Survey.

mainly belong to France and Twitter users acted more negatively about COVID-19 vaccines. We check the news in March 2022 and find that this is the period when the Omicron variants were transmitted fast in eastern France regions. This also implies that Twitter can capture the changes of vaccination attitudes faster than survey in emergent incidents.

We further check whether the above identified correlation persists in various socio-demographic sub-populations. We divide the survey respondents and Twitter users into 18 groups according to their age, gender and political ideology. Then we calculate the region-level vaccination reluctance rates for each group, and compute the Pearson correlation between the reluctance rates of each Twitter group and the corresponding survey group. Figure 9.5 depicts the results. The general observation is that the correlation indeed varies among different demographic groups. The correlation increases for groups with larger ages but remains almost the same regardless of gender and political ideology. This implies that younger people may actively participate in discussion about vaccines, but they are less willing to express their real intention of vaccination on Twitter. In addition, the correlation decreases with time, which implies when a high-level vaccination rate is reached, the topics on Twitter about COVID-19 vaccines become less relevant to users' intention of vaccination.

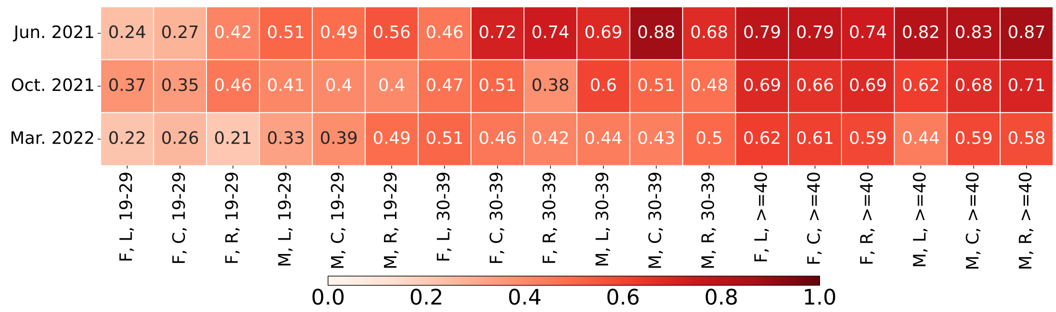


Figure 9.5: Pearson correlations (r) between Twitter and survey across 24 regions by Age, gender (F: female, M: male) and political ideology (L: left, C: centre, R: right).

9.6 Discussion and Conclusion

We have propose a framework to directly estimate public vaccine hesitancy from Twitter. Our framework addresses the widely recognised inherent errors when analysing social media data with a quantitative measurement of vaccine hesitancy and an adapted method correcting socio-demographic selection bias. With our multi-wave surveys and collection of tweets, we conduct the first attempt to validate the consistency between Twitter and surveys regarding monitoring COVID-19 vaccine hesitancy both across regions and over time. Through comprehensive cross-validation, we show that Twitter can capture the public vaccine hesitancy and generate at least strongly correlated estimation with that inferred by surveys. Moreover, this correlation is consistent over time on levels of both countries and regions although it varies among different socio-demographic sub-populations. Last but not least, we consider the global demands of vaccination attitude monitoring and empower our framework to deal with multilingual texts. With this chapter, we re-establish the power of social media in complementing social surveys to continuously capture the fast evolution of vaccine hesitancy in public health crises like COVID-19. Moreover, our work can encourage social scientists to use social media in studies, especially for the topics which are hard to formulate in questionnaires e.g., influences of social interactions on vaccine hesitancy.

We have a few limitations that will be addressed in future. First, our cross-validation is conducted in Western Europe. Similar studies in other areas can further validate the generality of our framework and our findings. Second, with vaccine hesitancy consistency validated, it will be helpful to examine whether existing social findings such as correlated factors can also be confirmed on social media data. Third, we only test three socio-demographic attributes. In spite of their effectiveness in bias correction, other socio-demographic attributes can also be tested and added to the bias correction if they can lead to better performance, especially with new progress in artificial intelligence.

Ethical consideration. This chapter is grounded in public data and does not involve any private information from individuals. The research process was established in full compliance with FAIR data principles, Twitter Developer Agreement & Policy and relevant policies. The survey is also approved by the Ethics Review Panel of our institution.

Chapter 10

Conclusion and Future Work

This thesis focuses on two key challenges in virtual ethnography research: data acquisition and the application of computational tools. We exemplify virtual ethnography research during the COVID-19 pandemic by concentrating on combating infodemics. In terms of data acquisition, two datasets are collected for different downstream research purposes, covering a two-year period from October 22, 2019, to December 31, 2021. Concerning the application of computational tools, particularly focusing on SNA, we investigate three research questions: social characteristics, information diffusion, and sentiment analysis. In social characteristics, our objective is to identify users and subgroups who genuinely facilitate information diffusion. For information diffusion, we propose prediction models based on users' susceptibility, influence, and message content to estimate the ultimate popularity of a message and pinpoint users who would eventually retweet the message. In sentiment analysis, we develop a GNN-based text classification model for the classification and monitoring of vaccine hesitancy content. Through cross-validation, we demonstrate that social media data can serve as a significant complement to surveys when tracking vaccine hesitancy.

10.1 Conclusion

We restate the five research questions proposed in Chapter 1.

> Research Question 1
>
> How can we gather an adequate amount of data to fulfil the demands of virtual ethnography studies?

To address this question, we introduced a data collection method tailored for acquiring comprehensive COVID-19-related Twitter datas, which includes tweets, users' geographic locations, social networks, and annotations on users' attitudes towards COVID-19 vaccines. The resulting datasets supports various virtual ethnographic studies related to COVID-19 and online social networks.

> Research Question 2
>
> How to design a measurement to capture the actual bridging performance of social media users in terms of spreading COVID-19-related information?

To tackle this question, we proposed two novel measurements to assess the true bridging performance of individual users and user subgroups in spreading COVID-19 information. Unlike existing metrics based on social connections, our metric is derived directly from information diffusion history. By manually analysing the collected dataset, we were able to identify previously overlooked influential health professionals and volunteers, in addition to super tweeters. Furthermore, we discovered a significant negative correlation between a user's bridging performance and their subjective well-being during the pandemic, validating the applicability of our measurements for virtual ethnographic studies.

> Research Question 3
>
> How to accurately predict the popularity of information?

To achieve this goal, we conducted a two-part study. In the first part of the study, we investigated the influence of message content on users' retweeting behaviour. We analysed collected data and confirmed the existence of the "info-exposure spillover effect" which demonstrates that exposure to different information affects social media users' behaviour in diffusing information during a pandemic. We then developed three new models using GNNs and a temporal coding method, which outperformed the baseline for all types of information.

In the second part of our study, we incorporated topic-specific and context-dependent susceptibility and influence into our model. We introduced a new deep learning model, CasSIM, that achieves both popularity prediction and final adopter prediction by exploring the dual role of users as receivers and distributors in the diffusion process. CasSIM models the topic-specific property of susceptibilities and influences. through effective user analysis and captures the dynamics of susceptibilities and influence during information dissemination using GNN. Through extensive experiments on three real-life datasets, we validated the effectiveness of CasSIM in predicting popularity and final adopters, and the results showed that CasSIM outperforms state-of-the-art methods, especially when shorter cascades are observed in social media and other scenarios where cascades also exist.

> Research Question 4
>
> How to accurately extract users' attitudes from their posts?

In response, we designed a framework for learning vaccination attitudes from social media text posts. Although vaccination attitudes extracted from social media may not be as reliable as conventional social surveys, our framework allows for faster tracking of public vaccination attitudes and capturing changes that require specific attention. We leveraged friends' vaccination discourse as contextual information to reduce interference from linguistic features such as sarcasm and irony. Our model instantiated from the framework improved the state-of-the-art text-only method by up to 23% in accuracy according to our manually annotated dataset. We validated the effectiveness of the model in capturing public vaccine hesitancy in real-life through cross-validation with published statistics and manual analysis. Using our model, we identified 11 themes from widely diffused information on Twitter and validated the predictability of users' vaccine hesitancy changes based on the information they perceived from social media. This demonstrates the potential use of our model in practice.

Research Question 5

Is it possible to accurately measure individuals' vaccine hesitancy from social media using appropriately designed methods?

To address this research question, we proposed a framework for directly estimating public vaccine hesitancy from Twitter. Our framework included a quantitative measurement of vaccine hesitancy and an adapted method to correct socio-demographic selection bias. To validate the consistency between Twitter and surveys regarding monitoring COVID-19 vaccine hesitancy across regions and over time, we conducted multi-wave surveys and collected tweets. Through comprehensive cross-validation, we demonstrated that Twitter can capture public vaccine hesitancy and generate strongly correlated estimations with those inferred by surveys. We also found that this correlation is consistent over time on both country and regional levels, although it varies among different socio-demographic subpopulations.

10.2 Limitation

While numerous SNA tools have been created to support virtual ethnographic research, various potential avenues remain to be investigated more thoroughly.

10.2.1 The Lack of Multiple Data Sources

A notable limitation of this dissertation is its exclusive focus on Twitter data as the primary source of virtual ethnographic research concerning COVID-19. While the proposed data collection method are extensive for the Twitter, and the results offer valuable insights into social characteristics, information diffusion, and sentiment analysis, these findings may not be generalise to other social media platforms or diverse types of online communities. This restricted scope risks neglecting the wider context of online communication and the potential influence of various virtual spaces on the diffusion of information during a public crisis.

First, different social media platforms exhibit unique user demographics, interaction modalities, and content-sharing mechanisms. These distinguishing features can considerably affect the applicability and effectiveness of the computational methods and models proposed in this study. For instance, social media platforms like Instagram and TikTok display distinctive user behaviours and content formats compared to Twitter. The absence of intra-platform reposts on these platforms presents a different landscape for user engagement. Additionally, their focus leans more heavily towards visual content such as images and videos, contrasting with Twitter's text-focused format. These unique attributes might lead to variations in the mechanisms of information diffusion and sentiment dynamics, which might not be fully captured by models that are trained solely on Twitter data.

Second, relying solely on Twitter data may introduce bias into the analysis, as it may not represent the full spectrum of public opinion or information diffusion processes. Moreover, other platforms, such as online forums or chat groups like Reddit or Telegram, foster different types of interactions and host distinct communities with diverse perspectives on the COVID-19 pandemic. Focusing only on Twitter users, who may exhibit specific characteristics or behaviours that do not accurately reflect the broader population, could limit the ecological validity and generalisability of the study results. Expanding the scope of analysis to include additional platforms and online communities would enhance the robustness of the findings and provide a more comprehensive understanding of virtual ethnographic research during a pandemic.

10.2.2 The Impact of Multimedia Content

A second limitation of this thesis is its heavy reliance on text-based information for developing and evaluating computational models, overlooking the potential impact of other forms of content on information dissemination. Various forms of multimedia content, such as images, videos, or links to external websites, are frequently shared by users on social media platforms and can significantly influence users' decisions to engage with and disseminate information.

Not accounting for these additional content types may cause the proposed model to overlook the full range of factors that contribute to the communication process, potentially leading to an incomplete or biased understanding of the underlying dynamics. Furthermore, multimedia content may convey subtle or complex meanings that are difficult to interpret through textual analysis alone. For instance, a tweet featuring visual content like an image or video may express sentiment or opinion that isn't directly translatable into a text-based message, yet it can significantly impact user behaviour and the spread of the message. Similarly, links to external sites may provide contextually relevant information that can influence users' decisions about sharing information. Excluding these content types from the analysis could result in oversimplifying the virtual ethnographic landscape and limit the model's ability to accurately predict and comprehend the complexity of information diffusion in the context of public crises. Incorporating multimedia content and external links into the analysis would provide a more holistic understanding of information dissemination and improve the robustness of the computational models.

10.3 Future Work

Based on the limitations identified in our thesis, there are several promising directions for future research to enhance the applicability and generalisability of the proposed computational methods and models in virtual ethnographic studies related to public crises and other contexts.

10.3.1 Generalising to Other Online Platforms

One direction for our future research is to extend the data collection methodologies and computational models to other social media platforms and online communities, such as Facebook, Instagram, TikTok, Reddit, and Telegram, among others. This would allow for a broader understanding of the social characteristics, information diffusion, and sentiment analysis related to COVID-19 and other topics of interest.

To achieve this, we could first develop platform-specific data collection methodologies that account for the unique features, user demographics, and content-sharing mechanisms of each platform. Next, we will adapt and evaluate the proposed computational methods and models in the context of these platforms, taking into consideration the distinct interaction patterns and network structures. Comparative studies could then be conducted to investigate the similarities and differences in information diffusion and sentiment dynamics across various online environments. By expanding the scope of virtual ethnographic studies to include multiple platforms and communities, we would be better equipped to understand the complexities of information diffusion and public sentiment during the COVID-19 pandemic or other public crises.

This comprehensive approach could provide valuable insights for policymakers, public health professionals, and other stakeholders, as they work to design more effective communication strategies and interventions to address public concerns and combat the spread of misinformation.

10.3.2 Incorporating Multimodal Content

Since our current model primarily focuses on text-based information, future work should investigate incorporating other content types, such as images, videos, and links to external websites, into the computational methods and models. This will enable us to gain a more comprehensive understanding of the factors influencing information dissemination and user behaviour, as users often base their decisions on a combination of text, visual, and other multimedia content.

To tackle this challenge, we plan to develop multimodal deep learning models in the future that efficiently process and analyse various forms of content while synchronising them with textual data. This may involve employing computer vision techniques for image and video analysis, natural language processing for textual information, and GNNs for social network analysis.

By integrating multimodal content, the proposed model can offer a more inclusive perspective on information diffusion and sentiment dynamics in online communities, thus enhancing its effectiveness and applicability across diverse contexts and platforms.

Overall, the future work outlined above aims to build on the contributions of this thesis by enhancing the applicability and generalisability of the proposed computational methods and models for virtual ethnographic studies. By addressing these limitations, we can continue to advance the field and provide valuable insights into the virtual world.

Bibliography

[AAAH+20] Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. Top concerns of tweeters during the covid-19 pandemic: infoveillance study. *Journal of Medical Internet Research*, 22(4):e19016, 2020.

[AAME+21] Rasmieh Al-Amer, Della Maneze, Bronwyn Everett, Jed Montayre, Amy R Villarosa, Entisar Dwekat, and Yenna Salamonson. COVID-19 vaccination intention in the first year of the pandemic: A systematic review. *Clinical Nursing*, 31(1-2):62–86, 2021.

[AAOA21] Adel Alfatease, Ali M Alqahtani, Khalid Orayj, and Sultan M Alshahrani. The impact of social media on the acceptance of the COVID-19 vaccine: a cross-sectional study from Saudi Arabia. *Patient Prefer Adherence*, 15:2673–2681, 2021.

[ABKK20] Ashley Amaya, Ruben Bach, Frauke Kreuter, and Florian Keusch. *Measuring the strength of attitudes in social media data*, chapter 5, pages 163–192. John Wiley & Sons, Ltd, 2020.

[ÄEN20] Andreas Älgå, Oskar Eriksson, and Martin Nordberg. Analysis of scientific publications during the early phase of the covid-19 pandemic: Topic modeling study. *Journal of Medical Internet Research*, 22(11):e21559, 2020.

[AGJ+20] Pushkal Agarwal, Kiran Garimella, Sagar Joglekar, Nishanth Sastry, and Gareth Tyson. Characterising user content on a multi-lingual social network. In *Proc. 14th International Conference on Web and Social Media (ICWSM)*, pages 2–11. AAAI Press, 2020.

[Ako18] Haldun Akoglu. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3):91–93, 2018.

[AKZ20] Muhammad Abulaish, Ashraf Kamal, and Mohammed J. Zaki. A survey of figurative language and its computational detection in online social networks. *ACM Transactions on the Web*, 14(1), 2020.

[AMH20] Sirry Alang, Donna D McAlpine, and Rachel Hardeman. Police brutality and mistrust in medical institutions. *Journal of Racial and Ethnic Health Disparities*, pages 1–9, 2020.

[AMS21] Shiva Raj Acharya, Deog Hwan Moon, and Yong Chul Shin. Assessing attitude toward COVID-19 vaccination in South Korea. *Frontiers in Psychology*, 12:3222, 2021.

[ARAK+21] Majid Alabdulla, Shuja Mohd Reagu, Abdullatif Al-Khal, Marwa Elzain, and Roland M Jones. COVID-19 vaccine hesitancy and attitudes in Qatar: A national cross-sectional survey of a migrant-majority population. *Influenza and Other Respiratory Viruses*, 15(3):361–370, 2021.

[ARZ20] Allison E Aiello, Audrey Renson, and Paul N Zivich. Social media- and internet-based disease surveillance for public health. *Annual Review of Public Health*, 41(1):101–118, 2020.

[ASK+22] Nick Andrews, Julia Stowe, Freja Kirsebom, Samuel Toffa, Ruchira Sachdeva, Charlotte Gower, Mary Ramsay, and Jamie Lopez Berna. Effectiveness of COVID-19 booster vaccines against COVID-19 related symptoms, hospitalisation and death in england. *Nature Medicine*, 28:831–837, 2022.

[AT08] S. Aranganayagi and Kuttiyannan Thangavel. Clustering categorical data using silhouette coefficient as a relocating measure. In *Proc. 2007 International Conference on Computational Intelligence and Multimedia Applications (ICCIMA)*, volume 2, pages 13–17. IEEE, 2008.

[AXV+11] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J Passonneau. Sentiment analysis of Twitter data. In *Proc. 2011 Workshop on Language in Social Media (LSM)*, pages 30–38, 2011.

[Bai75] Norman TJ Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Charles Griffin & Company, 1975.

[Bak17] Reg Baker. *Big Data*, chapter 3, pages 47–69. John Wiley & Sons, Ltd, 2017.

[BBCC22] Vojtěch Bartoš, Michal Bauer, Jana Cahlíková, and Julie Chytilová. Communicating doctors' consensus persistently increases COVID-19 vaccinations. *Nature*, 606(7914):542–549, 2022.

[BC12] Danah Boyd and Kate Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5):662–679, 2012.

[BCCEAN20] Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1644–1650. ACL, 2020.

[BCDL22] Liyousew G Borga, Andrew E Clark, Conchita D'Ambrosio, and Anthony Lepinteur. Characteristics associated with COVID-19 vaccine hesitancy. *Scientific Reports*, 12(1):1–9, 2022.

[BGRM11] Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. Happiness is assortative in online social networks. *Artificial Life*, 17(3):237–251, 2011.

[BH17] Georgina Born and Christopher Haworth. Mixing it: Digital ethnography and online research methods—a tale of two global digital music genres. In *The Routledge Companion to Digital Ethnography*, pages 96–112. Routledge, 2017.

[BJL+20] Rhys Biddle, Aditya Joshi, Shaowu Liu, Cécile Paris, and Guandong Xu. Leveraging sentiment distributions to distinguish figurative from literal health reports on twitter. In *Proc. 2020 International Conference on World Wide Web (WWW)*, pages 1217–1227. ACM, 2020.

[BKG+22] Ronelle Burger, Timothy Köhler, Aleksandra M Golos, Alison M Buttenheim, René English, Michele Tameris, and Brendan Maughan-Brown. Longitudinal changes in COVID-19 vaccination intent among South African adults: evidence from the NIDS-CRAM panel survey, February to May 2021. *BMC Public Health*, 22(1):1–10, 2022.

[BKZY21] Datian Bi, Jingyuan Kong, Xue Zhang, and Junli Yang. Analysis on health information acquisition of social network users by opinion mining: case analysis based on the discussion on COVID-19 vaccinations. *Journal of Healthcare Engineering*, 2021, 2021.

[BL99] Margaret M Bradley and Peter J Lang. Affective norms for English words (anew): Instruction manual and affective ratings. Technical report, the Centre for Research in Psychophysiology, University of Florida, 1999.

[BLG16] Simon Bourigault, Sylvain Lamprier, and Patrick Gallinari. Representation learning for information diffusion through social networks: an embedded cascade model. In *Proc. 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 573–582. ACM, 2016.

[BLL+14] Simon Bourigault, Cédric Lagnier, Sylvain Lamprier, Ludovic Denoyer, and Patrick Gallinari. Learning social network embeddings for predicting information diffusion. In *Proc. 7th International Conference on Web Search and Data Mining (WSDM)*, pages 393–402. ACM, 2014.

[BM19] Madeleine Burnette-McGrath. Packingham v. north carolina. *Ohio Northern University Law Review*, 44(1):6, 2019.

[BM21] Debanjan Banerjee and K. S. Meena. COVID-19 as an 'infodemic' in public health: Critical role of the social media. *Frontiers in Public Health*, 9:231–238, 2021.

[BMKP21] Nirbachita Biswas, Toheeb Mustapha, Jagdish Khubchandani, and James H Price. The nature and extent of COVID-19 vaccination hesitancy in healthcare workers. *Journal of Community Health*, 46(6):1244–1251, 2021.

[BMM+20] Jobie Budd, Benjamin S. Miller, Erin M. Manning, Vasileios Lampos, Mengdie Zhuang, Michael Edelstein, Geraint Rees, Vincent C. Emery, Molly M. Stevens, Neil Keegan, Michael J. Short, Deenan Pillay, Ed Manley, Ingemar J. Cox, David Heymann, Anne M. Johnson, and Rachel A. McKendry. Digital technologies in the public-health response to COVID-19. *Nature Medicine*, 26(8):1183–1192, 2020.

[BMZP19] Aida Bianco, Valentina Mascaro, Rossella Zucco, and Maria Pavia. Parent perspectives on childhood vaccination: How to deal with vaccine hesitancy and refusal? *Vaccine*, 37(7):984–990, 2019.

[BND+20] Serena Barello, Tiziana Nania, Federica Dellafiore, Guendalina Graffigna, and Rosario Caruso. Vaccine hesitancy among university students in italy during the COVID-19 pandemic. *European Journal of Epidemiology*, 35(8):781–783, 2020.

[BNJ01] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. In *Proc. 2001 Neural Information Processing Systems (NIPS)*, pages 601–608. MIT Press, 2001.

[BNW+20] Ana I. Bento, Thuy Nguyen, Coady Wing, Felipe Lozano-Rojas, Yong Yeol Ahn, and Kosali Simon. Evidence from Internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences of the United States of America*, 117(21):11220–11222, 2020.

[BR08] Luís M.A. Luis M A Bettencourt and Ruy M. Ribeiro. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*, 3(5):e2185, 2008.

[BRMA12] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada A. Adamic. The role of social networks in information diffusion. In *Proc. 21stInternational Conference on World Wide Web (WWW)*, pages 519–528. ACM, 2012.

[BWJ+17] William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318, 2017.

[BZSL14] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *Proc. 2nd International Conference on Learning Representations, (ICLR)*, 2014.

[CAD+14] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proc. 2014 International Conference on World Wide Web (WWW)*, pages 925–936, 2014.

[CBHK02] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[CCET16] Juan M. Cotelo, Fermin L. Cruz, Fernando Enríquez, and José A. Troyano. Tweet categorization by combining content and structural knowledge. *Information Fusion*, 31:54–64, 2016.

[CCP22a] Ninghan Chen, Xihui Chen, and Jun Pang. A multilingual dataset of COVID-19 vaccination attitudes on Twitter. *Data in Brief*, 44:108503, 2022.

[CCP+22b] Ninghan Chen, Xihui Chen, Jun Pang, Liyousew G. Borga, Conchita D'Ambrosio, and Claus Vögele. Measuring COVID-19 vaccine hesitancy: Consistency of social media with surveys. In *Proc. 13th International Conference on Social Informatics (SocInfo)*, volume 13618 of *Lecture Notes in Computer Science*, pages 196–210. Springer, 2022.

[CCZP21] Ninghan Chen, Xihui Chen, Zhiqiang Zhong, and Jun Pang. From #jobsearch to #mask: improving COVID-19 cascade prediction with spillover effects. In *Proc. 2021 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 455–462. ACM, 2021.

[CCZP22a] Ninghan Chen, Xihui Chen, Zhiqiang Zhong, and Jun Pang. The burden of being a bridge: analysing subjective well-being of Twitter users during the COVID-19 pandemic. In *Proc. 2022 European Conference on Machine Learning and 26th Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, Lecture Notes in Computer Science. Springer, 2022.

[CCZP22b] Ninghan Chen, Xihui Chen, Zhiqiang Zhong, and Jun Pang. "Double vaccinated, 5G boosted!": learning attitudes towards COVID-19 vaccination from social media, 2022.

[CCZP22c] Ninghan Chen, Xihui Chen, Zhiqiang Zhong, and Jun Pang. Exploring spillover effects for COVID-19 cascade prediction. *Entropy*, 24(2):222, 2022.

[Cen10] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.

[Cho15] Chun Tung Chou. A markovian approach to the optimal demodulation of diffusion-based molecular communication networks. *IEEE Transactions on Communications*, 63(10):3728–3743, 2015.

[CJS+20] Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John Philip McCrae. A sentiment analysis dataset for code-mixed malayalam-english. In Dorothee Beermann, Laurent Besacier, Sakriani Sakti, and Claudia Soria, editors, *Proc. 2020 Joint Workshop on Spoken Language Technologies for Under-resourced languages and Collaboration and Computing for Under-Resourced Languages (SLTU)*, pages 177–184. European Language Resources Association, 2020.

[CJY+13] Peng Cui, Shifei Jin, Linyun Yu, Fei Wang, Wenwu Zhu, and Shiqiang Yang. Cascading outbreak prediction in networks: a data-driven approach. In *Proc. 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 901–909. ACM, 2013.

[CKG+20] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 8440–8451. ACL, 2020.

[CL11] Chih-Chung Chung Chang and Chih-Jen Jen Lin. LIBSVM: A library for support vector machines. *Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011.

[CLCR+19] Alberto Chong, Gianmarco León-Ciliotta, Vivian Roza, Martín Valdivia, and Gabriela Vega. Urbanization patterns, information diffusion, and female voting in rural paraguay. *American Journal of Political Science*, 63(2):323–341, 2019.

[CLF20] Emily Chen, Kristina Lerman, and Emilio Ferrara. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273, 2020.

[CMK+20] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. Falling into the echo chamber: the Italian vaccination debate on twitter. In *Proc. 2020 International AAAI Conference on Web and Social Media (ICWSM)*, volume 14, pages 130–140. The AAAI Press, 2020.

[CPAA+21] Fidelia Cascini, Ana Pantovic, Yazan Al-Ajlouni, Giovanna Failla, and Walter Ricciardi. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review. *EClinicalMedicine*, 40:101113, 2021.

[CPAA+22] Fidelia Cascini, Ana Pantovic, Yazan A. Al-Ajlouni, Giovanna Failla, Valeria Puleo, Andriy Melnyk, Alberto Lontano, and Walter Ricciardi. Social media and attitudes towards COVID-19 vaccination: a systematic review of the literature. *eClinicalMedicine*, 48:101454, 2022.

[CQG+20] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The COVID-19 social media infodemic. *Scientific Reports*, 10(1):16598, 2020.

[CR10] Claudio Cioffi-Revilla. A methodology for complex social simulations. *Journal of Artificial Societies and Social Simulation*, 13(1):7, 2010.

[CSC+17] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. DeepHawkes: Bridging the gap between prediction and understanding of information cascades. In *Proc. ACM on Conference on Information and Knowledge Management (CIKM)*, pages 1149–1158. ACM, 2017.

[CSG+20] Qi Cao, Huawei Shen, Jinhua Gao, Bingzheng Wei, and Xueqi Cheng. Popularity prediction on social platforms with coupled graph neural networks. In *Proc. 13th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 70–78. ACM, 2020.

[CvMBB14] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In *Proc. 8th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST)*, pages 103–111. ACL, 2014.

[CZP21] Ninghan Chen, Zhiqiang Zhong, and Jun Pang. An exploratory study of COVID-19 information on twitter in the Greater Region. *Big Data and Cognitive Computing*, 5(1):5, 2021.

[CZZ+19] Xueqin Chen, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Ting Zhong, and Fengli Zhang. Information diffusion prediction via recurrent cascades convolution. In *Proc. 35th IEEE International Conference on Data Engineering (ICDE)*, pages 770–781. IEEE, 2019.

[DBG+20] Souvik Dubey, Payel Biswas, Ritwik Ghosh, Subhankar Chatterjee, Mahua Jana Dubey, Subham Chatterjee, Durjoy Lahiri, and Carl J Lavie. Psychosocial impact of COVID-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5):779–788, 2020.

[DCLT19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186. ACL, 2019.

[DELG85] ED Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. The satisfaction with life scale. *Journal of Personality Assessment*, 49(1):71–75, 1985.

[DLG+13] Eve Dubé, Caroline Laberge, Maryse Guay, Paul Bramadat, Réal Roy, and Julie A Bettinger. Vaccine hesitancy: An overview. *Human Vaccines & Immunotherapeutics*, 9(8):1763–1773, 2013.

[DLP+20] Viet Duong, Jiebo Luo, Phu Pham, Tongyu Yang, and Yu Wang. The ivory tower lost: how college students respond differently than the general public to the COVID-19 pandemic. In *Proc. 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 126–130. IEEE, 2020.

[DLZ+15] Adam G Dunn, Julie Leask, Xujuan Zhou, Kenneth D Mandl, and Enrico Coiera. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: An observational study. *Journal of Medical Internet Research*, 17(6):e144, 2015.

[DOL03] Ed Diener, Shigehiro Oishi, and Richard E Lucas. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, 54(1):403–425, 2003.

[DPLA+21] Giorgia Della Polla, Francesca Licata, Silvia Angelillo, Concetta Paola Pelullo, and Aida Bianco. Characteristics of healthcare workers vaccinated against influenza in the era of COVID-19. *Vaccines*, 9(7):695, 2021.

[DPT+21] Matthew R. DeVerna, Francesco Pierri, Bao Tran Truong, John Bollenbacher, David Axelrod, Niklas Loynes, Christopher Torres-Lugo, Kai-Cheng Yang, Filippo Menczer, and John Bryden. Covaxxy: A collection of english-language twitter posts about COVID-19 vaccines. In Ceren Budak, Meeyoung Cha, Daniele Quercia, and Lexing Xie, editors, *Proc. 2021 International AAAI Conference on Web and Social Media, (ICWSM)*, pages 992–999. AAAI Press, 2021.

[DSS93] Jean-Claude Deville, Carl-Erik Särndal, and Olivier Sautory. Generalized raking procedures in survey sampling. *Journal of the American statistical Association*, 88(423):1013–1020, 1993.

[DZLL17] Matthew A Davis, Kai Zheng, Yang Liu, and Helen Levy. Public response to obamacare on Twitter. *Journal of Medical Internet Research*, 19(5):e167, 2017.

[EAC+20] Adam B. Evans, Sine Agergaard, Paul Ian Campbell, Kevin Hylton, and Verena Lenneis. 'Black Lives Matter:' sport, race and ethnicity in challenging times. *European Journal for Sport and Society*, 17(4):289–300, 2020.

[EDL16] Javid Ebrahimi, Dejing Dou, and Daniel Lowd. Weakly supervised tweet stance classification by relational bootstrapping. In *Proc. 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1012–1017. ACL, 2016.

[EG12] Irene Eleta and Jennifer Golbeck. Bridging languages in social networks: How multilingual users of Twitter connect language communities? In *Proc. Association for Information Science and Technology (ASIS&T)*, volume 49, pages 1–4. Wiley, 2012.

[EJB20] Tarik Endale, Nicole St Jean, and Dina Birman. COVID-19 and refugee and immigrant youth: A community-based mental health perspective. *Psychol Trauma*, 12(S1):S225–S227, 2020.

[EJR$^+$10] David Ediger, Karl Jiang, E. Jason Riedy, David A. Bader, Courtney D. Corley, Robert M. Farber, and William N. Reynolds. Massive social network analysis: Mining twitter for social good. In *Proc. 2021 International Conference on Parallel Processing (ICPP)*, pages 583–593. IEEE, 2010.

[ENW$^+$21] Pascale M.J. Engel de Abreu, Sascha Neumann, Cyril Wealer, Neander Abreu, Elizeu Coutinho Macedo, and Claudine Kirsch. Subjective well-being of adolescents in Luxembourg, Germany, and Brazil during the COVID-19 pandemic. *Journal of Adolescent Health*, 69(2):211–218, 2021.

[FFG22] Roberto Stefan Foa, Mark Fabian, and Sam Gilbert. Subjective well-being during the 2020–21 global coronavirus pandemic: Evidence from high frequency time series data. *PLOS ONE*, 17(2):1–22, 02 2022.

[FG22] Alexander J. Freund and Philippe J. Giabbanelli. An experimental study on the scalability of recent node centrality metrics in sparse complex networks. *Frontiers in Big Data*, 5, 2022.

[FLP13] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

[FLŞ$^+$20] Senaka Fernando, Julio Amador Díaz López, Ovidiu Şerban, Juan Gómez-Romero, Miguel Molina-Solana, and Yike Guo. Towards a large-scale Twitter observatory for political events. *Future Generation Computer Systems*, 110:976–983, 2020.

[Fre78] Linton C Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[FSMS21] Pierri Francesco, Pavanetto Silvio, Brambilla Marco, and Ceri Stefano. Vaccinitaly: monitoring italian conversations around vaccines on twitter. *CoRR*, abs/2101.03757, 2021.

[GADN21] Sefater Gbashi, Oluwafemi Ayodeji Adebo, Wesley Doorsamy, and Patrick Berka Njobeh. Systematic delineation of media polarity on COVID-19 vaccines in Africa: Computational linguistic modeling study. *JMIR Medical Informatics*, 9(3):e22916, 2021.

[GC06] Derek Greene and Padraig Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proc. 23rd International Conference on Machine Learning (ICML)*, volume 148, pages 377–384, 2006.

[GCCH19] Zakariya Ghalmane, Chantal Cherifi, Hocine Cherifi, and Mohammed El Hassouni. Centrality in Complex Networks with Overlapping Community Structure. *Scientific Reports*, 9(1):10133, 2019.

[GCH19a] Keith Gunaratne, Eric A Coomes, and Hourmazd Haghbayan. Temporal trends in anti-vaccine discourse on Twitter. *Vaccine*, 37(35):4867–4871, 2019.

[GCH19b] Keith Gunaratne, Eric A Coomes, and Hourmazd Haghbayan. Temporal trends in anti-vaccine discourse on twitter. *Vaccine*, 37(35):4867–4871, 2019.

[GDM+21] Davide Gori, Francesco Durazzi, Marco Montalti, Zeno Di Valerio, Chiara Reno, Maria Pia Fantini, and Daniel Remondini. Mis-tweeting communication: a vaccine hesitancy analysis among Twitter users in Italy. *Acta Bio Medica: Atenei Parmensis*, 92(S6), 2021.

[GGA20] Victor Grech, Charmaine Gauci, and Steve Agius. Vaccine hesitancy among maltese healthcare workers toward influenza and novel COVID-19 vaccination. *Early Human Development*, 1:105213, 2020.

[GGLNT04] Daniel Gruhl, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proc. 2004 International Conference on World Wide Web (WWW)*, pages 491–501. ACM, 2004.

[GGLT04] Daniel Gruhl, Ramanathan V. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In Stuart I. Feldman, Mike Uretsky, Marc Najork, and Craig E. Wills, editors, *Proc. 13th International Conference on World Wide Web (WWW)*, pages 491–501. ACM, 2004.

[GHW17] Lin Gong, Benjamin Haines, and Hongning Wang. Clustered model adaption for personalized sentiment analysis. In *Proc. 26th International Conference on World Wide Web, (WWW)*, pages 937–946. ACM, 2017.

[GLG+22] Salvatore Giorgi, Veronica E. Lynn, Keshav Gupta, Farhan Ahmed, Sandra Matz, Lyle H. Ungar, and H. Andrew Schwartz. Correcting sociodemographic selection biases for population prediction from social media. In Ceren Budak, Meeyoung Cha, and Daniele Quercia, editors, *Proc. 2022 International AAAI Conference on Web and Social Media (ICWSM)*, pages 228–240. AAAI Press, 2022.

[GLK12] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data*, 5(4):21:1–21:37, 2012.

[GPG+17] Sheng Gao, Huacan Pang, Patrick Gallinari, Jun Guo, and Nei Kato. A novel embedding method for information diffusion prediction in social network big data. *IEEE Transactions on Industrial Informatics*, 13(4):2097–2105, 2017.

[GPGC21] Stefano Guarino, Francesco Pierri, Marco Di Giovanni, and Alessandro Celestini. Information disorders during the COVID-19 infodemic: The case of Italian facebook. *Online Social Networks Media*, 22:100124, 2021.

[Gra78] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[GSBC09] Angela Cora Garcia, Alecea I Standlee, Jennifer Bechkoff, and Yan Cui. Ethnographic approaches to the internet and computer-mediated communication. *Journal of Contemporary Ethnography*, 38(1):52–84, 2009.

[GSL+16] E Dario Gutiérrez, Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi. Detecting cross-cultural differences using a multilingual topic model. *Transactions of the Association for Computational Linguistics*, 4:47–60, 2016.

[GW19] Matteo M. Galizzi and Lorraine Whitmarsh. How to measure behavioral spillovers: A methodological review and checklist. *Frontiers in Psychology*, 10:342, 2019.

[HAHO+21] Ma'mon M Hatmal, Mohammad AI Al-Hatamleh, Amin N Olaimat, Malik Hatmal, Dina M Alhaj-Qasem, Tamadur M Olaimat, and Rohimah Mohamud. Side effects and perceptions following COVID-19 vaccination in jordan: A randomized, cross-sectional study implementing machine learning for predicting severity of side effects. *Vaccines*, 9(6):556, 2021.

[Hal14] Scott A. Hale. Global connectivity and multilinguals in the Twitter network. In *Proc. 32th Conference on Human Factors in Computing Systems (CHI)*, pages 833–842. ACM, 2014.

[Haw71] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[HD96] J A P Heesterbeek and Klaus Dietz. The concept of Ro in epidemic theory. *Statistica Neerlandica*, 50(1):89–110, 1996.

[HDD11] Liangjie Hong, Ovidiu Dan, and Brian D. Davison. Predicting popular messages in twitter. In *Proc. 20th International Conference on World Wide Web (WWW)*, pages 57–58. ACM, 2011.

[HGS22] Golo Henseke, Francis Green, and Ingrid Schoon. Living with COVID-19: Subjective well-being in the second phase of the pandemic. *Journal of Youth and Adolescence*, 51(9):1679–1692, 2022.

[HHSC11] Brent J. Hecht, Lichan Hong, Bongwon Suh, and Ed H. Chi. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proc. 2011 International Conference on Human Factors in Computing Systems (CHI)*, pages 237–246. ACM, 2011.

[HHW+21] Raquel G Hernandez, Loni Hagen, Kimberly Walker, Heather O'Leary, and Cecile Lengacher. The COVID-19 vaccine social media infodemic: Healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, 17(9):2962–2964, 2021.

[HL04] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proc. 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 168–177. ACM, 2004.

[HLKX20] Zhao Hu, Xuhui Lin, Atipastsa Chiwanda Kaminga, and Huilan Xu. Impact of the COVID-19 epidemic on lifestyle behaviors and their association with subjective well-being among the general population in mainland China: Cross-sectional study. *Journal of Medical Internet Research*, 22(8):e21176, 2020.

[HLZ+14] Shaobin Huang, Tianyang Lv, Xizhe Zhang, Yange Yang, Weimin Zheng, and Chao Wen. Identifying node role in social network based on multiple indicators. *PLoS One*, 9(8):e103733, 2014.

[HM17] Yuli Patrick Hsieh and Joe Murphy. *Total Twitter error*, chapter 2, pages 23–46. John Wiley & Sons, Ltd, 2017.

[HMT+17] Vuokko Heikinheimo, Enrico Di Minin, Henrikki Tenkanen, Anna Hausmann, Joel Erkkonen, and Tuuli Toivonen. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *International Journal of Geo-Information*, 6(3):85, 2017.

[HRS+20] Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation. In *Proc. 2020 International Conference on Machine , (ICML)*, volume 119, pages 4411–4421. PMLR, 2020.

[HTD+21] Zhiyuan Hou, Yixin Tong, Fanxing Du, Linyao Lu, Sihong Zhao, Kexin Yu, Simon J Piatek, Heidi J Larson, Leesa Lin, et al. Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: a global social listening study. *Journal of Medical Internet Research*, 23(6):e27632, 2021.

[HTH+21] Amir Hussain, Ahsen Tahir, Zain Hussain, Zakariya Sheikh, Mandar Gogate, Kia Dashtipour, Azhar Ali, and Aziz Sheikh. Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. *Journal of Medical Internet research*, 23(4):e26627, 2021.

[HZ11] Yulan He and Deyu Zhou. Self-training from labeled features for sentiment analysis. *Information Processing & Management*, 47(4):606–616, 2011.

[JG10] Niklas Jakob and Iryna Gurevych. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proc. 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1035–1045. ACL, 2010.

[JGS+20] Kokil Jaidka, Salvatore Giorgi, H. Andrew Schwartz, Margaret L. Kern, Lyle H. Ungar, and Johannes C. Eichstaedt. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 117(19):10165–10171, 2020.

[JR20] Kia Jahanbin and Vahid Rahmanian. Using Twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine*, 13:26–28, 2020.

[JVR+20] Neil F Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. The online competition between pro- and anti-vaccination views. *Nature*, 582(7811):230–233, 2020.

[KB06] Arnd Christian König and Eric Brill. Reducing the human overhead in text categorization. In *Proc. 12th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 598–603. ACM, 2006.

[KB15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. 2015 International Conference on Learning Representations (ICLR)*. OpenReview.net, 2015.

[KCT+20] William DS Killgore, Sara A Cloonan, Emily C Taylor, Michael A Miller, and Natalie S Dailey. Three months of loneliness during the covid-19 lockdown. *Psychiatry Research*, 293:113392, 2020.

[KE05] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

[KJ17] Farzan Kolini and Lech Janczewski. Clustering and topic modelling: A new approach for analysis of national cyber security strategies. In *Pacific Asia Conference on Information Systems*, page 126. Association For Information Systems, 2017.

[KK15] Naoko Kaida and Kosuke Kaida. Spillover effect of congestion charging on pro-environmental behavior. *Environment, Development and Sustainability*, 17(3):409–421, 2015.

[KKT03] David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Lise Getoor, Ted E. Senator, Pedro M. Domingos, and Christos Faloutsos, editors, *Proc. 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 137–146. ACM, 2003.

[KL20] Jing Xuan Koh and Tau Ming Liew. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds. *Journal of Psychiatric Research*, 2020.

[KOU+12] Andrey Kupavskii, Liudmila Ostroumova, Alexey Umnov, Svyatoslav Usachev, Pavel Serdyukov, Gleb Gusev, and Andrey Kustarev. Prediction of retweet cascade size over time. In *Proc. 21st International Conference on Information and Knowledge Management (CIKM)*, pages 2335–2338. ACM, 2012.

[Kri70] Klaus Krippendorff. Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, 30(1):61–70, 1970.

[KSP+21] Jagdish Khubchandani, Sushil Sharma, James H Price, Michael J Wiblishauser, Manoj Sharma, and Fern J Webb. Covid-19 vaccination hesitancy in the united states: A rapid national assessment. *Journal of Community Health*, 46(2):270–277, 2021.

[KW17] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proc. 2017 International Conference on Learning Representations (ICLR)*. OpenReview.net, 2017.

[KWWO14] Suin Kim, Ingmar Weber, Li Wei, and Alice H. Oh. Sociolinguistic analysis of Twitter in multilingual societies. In *Proc. 25th Conference on Hypertext and Social Media (HT)*, pages 243–248. ACM, 2014.

[Lee18] Francis LF Lee. The spillover effects of political scandals: The moderating role of cynicism and social media communications. *Journalism & Mass Communication Quarterly*, 95(3):714–733, 2018.

[LFX+20] Xinlong Li, Xingyu Fu, Guangluan Xu, Yang Yang, Jiuniu Wang, Li Jin, Qing Liu, and Tianyuan Xiang. Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access*, 8:46868–46876, 2020.

[LL21a] Siru Liu and Jialin Liu. Public attitudes toward COVID-19 vaccines on English-language twitter: A sentiment analysis. *Vaccine*, 39(39):5499–5505, 2021.

[LL21b] Siru Liu and Jialin Liu. Public attitudes toward COVID-19 vaccines on English-language twitter: A sentiment analysis. *Vaccine*, 39(39):5499–5505, 2021.

[LLC11] Yung-Ming Li, Cheng-Yang Lai, and Ching-Wen Chen. Discovering influencers for marketing in the blogosphere. *Information Sciences*, 181(23):5143–5157, 2011.

[LLLY04] Bing Liu, Xiaoli Li, Wee Sun Lee, and Philip S. Yu. Text classification by labeling words. In Deborah L. McGuinness and George Ferguson, editors, *Proc. 16th Conference on Innovative Applications of Artificial Intelligence*, pages 425–430. AAAI Press, 2004.

[LMGM17] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. DeepCas: An end-to-end predictor of information cascades. In *Proc. 26th International Conference on World Wide Web (WWW)*, page 577–586, 2017.

[LOG+19] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.

[LPA+09] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Computational social science. *Science*, 323(5915):721–723, 2009.

[LPL+22] Qian Li, Hao Peng, Jianxin Li, Congying Xia, Renyu Yang, Lichao Sun, Philip S. Yu, and Lifang He. A survey on text classification: From traditional to deep learning. *ACM Transactions on Intelligent Systems and Technologies*, 13(2):41, 2022.

[LPN+11] Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary. Twitter trending topic classification. In *2011 IEEE International Conference on Data Mining Workshops (ICDM)*, pages 251–258. IEEE, 2011.

[LRP+21] Jeffrey V Lazarus, Scott C Ratzan, Adam Palayew, Lawrence O Gostin, Heidi J Larson, Kenneth Rabin, Spencer Kimball, and Ayman El-Mohandes. A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, 27(2):225–228, 2021.

[LW17] Lucas Leemann and Fabio Wasserfallen. Extending the use and prediction precision of subnational public opinion estimation. *American Journal of Political Science*, 61(4):1003–1022, 2017.

[LW18] Yang Liu and Yi-fang Brook Wu. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proc. 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 354–361. AAAI Press, 2018.

[LWW+22] Hanjia Lyu, Junda Wang, Wei Wu, Viet Duong, Xiyang Zhang, Timothy D Dye, and Jiebo Luo. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. *Intelligent Medicine*, 2(1):1–12, 2022.

[LYL19] Peiqin Lin, Meng Yang, and Jianhuang Lai. Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification. In *Proc. 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5088–5094. IJCAI, 2019.

[MBS+20] Milad Mirbabaie, Deborah Bunker, Stefan Stieglitz, Julian Marx, and Christian Ehnis. Social media in times of crisis: Learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *Journal of Information Technology*, 35(3):195–213, 2020.

[MCCD13] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *Proc. 2013 International Conference on Learning Representations (ICLR)*. OpenReview.net, 2013.

[MH18] Leland McInnes and John Healy. UMAP: Uniform manifold approximation and projection for dimension reduction. *CoRR*, abs/1802.03426, 2018.

[MHM18] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[MKC+21] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. Deep learning–based text classification: A comprehensive review. *ACM Computing Surveys*, 54(3):62:1–62:40, 2021.

[ML10] Seth A. Myers and Jure Leskovec. On the convexity of latent social network inference. In *Proc. 2010 International Conference on Neural Information Processing Systems (NIPS)*, volume 2, page 1741–1749. Curran Associates Inc., 2010.

[MMS21] Gianmarco De Francisci Morales, Corrado Monti, and Michele Starnini. No echo in the chambers of political interactions on reddit. *Scientific Reports*, 11(1):1–12, 2021.

[MMSS19] Sandra C Matz, Jochen I Menges, David J Stillwell, and H Andrew Schwartz. Predicting individual-level income from Facebook profiles. *PLOS ONE*, 14(3):e0214369, 2019.

[MOASN21] Chad A Melton, Olufunto A Olusanya, Nariman Ammar, and Arash Shaban-Nejad. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, 14(10):1505–1512, 2021.

[Mol95] Denis Mollison. *Epidemic Models: Their Structure and Relation to Data*, volume 5. Cambridge University Press, 1995.

[Moo97] Christopher Z Mooney. *Monte carlo simulation*, volume 116. Sage Publications, 1997.

[MS00] Chris Mann and Fiona Stewart. *Internet Communication and Qualitative Research: A Handbook for Researching Online*. Sage Publications, 2000.

[MSB+18] Feng Mai, Zhe Shan, Qing Bai, Xin (Shane) Wang, and Roger H.L. Chiang. How does social media impact bitcoin value? a test of the silent majority hypothesis. *Journal of Management Information Systems*, 35(1):19–52, 2018.

[MU20] Mohammed A Mamun and Irfan Ullah. COVID-19 suicides in Pakistan, dying off not COVID-19 fear but poverty?–the forthcoming economic challenges for a developing country. *Brain, Behavior, and Immunity*, 2020.

[MXKP21] Chen Mingxuan, Chu Xinqiao, and Subbalakshmi K. P. Mmcovar: Multimodal COVID-19 vaccine focused data repository for fake news detection and a baseline architecture for classification. *CoRR*, abs/2109.06416, 2021.

[MZBC20] Alexandros Mittos, Savvas Zannettou, Jeremy Blackburn, and Emiliano De Cristofaro. Analyzing genetic testing discourse on the web through the lens of Twitter, Reddit, and 4chan. *ACM Transactions on the Web*, 14(4), 2020.

[NCP+20] Sarah A. Nowak, Christine Chen, Andrew M. Parker, Courtney A. Gidengil, and Luke J. Matthews. Comparing covariation among vaccine hesitancy and broader beliefs within Twitter and survey data. *PLOS ONE*, 15(10):1–16, 10 2020.

[NLGB10] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. Automatic evaluation of topic coherence. In *Proc. 2010 Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 100–108. ACL, 2010.

[NLL14] Guolin Niu, Yi Long, and Victor O. K. Li. Temporal behavior of social network users in information diffusion. In *Proc. 2014 International Joint Conferences on Web Intelligence and Intelligent Agent Technologies (WI-IAT)*, pages 150–157. IEEE, 2014.

[NTFW22] Serge Nyawa, Dieudonné Tchuente, and Samuel Fosso-Wamba. COVID-19 vaccine hesitancy: a social media analysis using deep learning. *Annals of Operations Research*, 16:1–39, 2022.

[OCM+21] Katriona O'Sullivan, Serena Clark, Amy McGrane, Nicole Rock, Lydia Burke, Neasa Boyle, Natasha Joksimovic, and Kevin Marshall. A qualitative study of child and adolescent mental health during the COVID-19 pandemic in Ireland. *International Journal of Environmental Research and Public Health*, 18(3):1062, 2021.

[OGC+15] Derek O'callaghan, Derek Greene, Joe Carthy, Pádraig Cunningham, Derek O'Callaghan, Derek Greene, Joe Carthy, and Pádraig Cunningham. An analysis of the coherence of descriptors in topic modelling. *Expert Systems with Applications*, 42(13):5645–5657, 2015.

[OL20] Xiaozhi Ou and Hongling Li. Ynu_oxz @ haspeede 2 and AMI : Xlm-roberta with ordered neurons LSTM for classification task at EVALITA 2020. In *Proc. 2020 Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA)*, volume 2765, pages 683—689. CEUR-WS.org, 2020.

[OR02] Evelien Otte and Ronald Rousseau. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6):441–453, 2002.

[Org14] World Health Organization. Report of the sage working group on vaccine hesitancy., 2014. Available on`http://www.who.int/immunization/sage/meetings/2014/october/SAGE`.

[OSB22] C. Oswald, Sona Elza Simon, and Arnab Bhattacharya. SpotSpam: Intention analysis–driven SMS spam detection using BERT embeddings. *ACM Transactions on the Web*, 16(3), 2022.

[OTMS+11] Douglas J Opel, James A Taylor, Rita Mangione-Smith, Cam Solomon, Chuan Zhao, Sheryl Catz, and Diane Martin. Validity and reliability of a survey to identify vaccine-hesitant parents. *Vaccine*, 29(38):6598–6605, 2011.

[PA08] Helle Pullmann and Jüri Allik. Relations of academic and general self-esteem to school achievement. *Personality and Individual Differences*, 45(6):559–564, 2008.

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

[PBR19] Hao Peng, Ceren Budak, and Daniel M. Romero. Event-driven analysis of crowd dynamics in the black lives matter online social movement. In *Proc. 2019 International Conference on World Wide Web (WWW)*, pages 3137–3143. ACM, 2019.

[PCHG20] Neha Puri, Eric A Coomes, Hourmazd Haghbayan, and Keith Gunaratne. Social media and vaccine hesitancy: New updates for the era of covid-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics*, 16(11):2586–2593, 2020.

[PCWF07] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos. Netprobe: A fast and scalable system for fraud detection in online auction networks. In *Proc. 2007 International Conference on World Wide Web (WWW)*, pages 201–210. ACM, 2007.

[Pet92] Steven A Peterson. Church participation and political participation: The spillover effect. *American Politics Quarterly*, 20(1):123–139, 1992.

[PHP+15] Sarah Pink, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi. *Digital Ethnography: Principles and Practice*. Sage Publications, 2015.

[PL04] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. 42nd Annual Meeting of the Association for Computational Linguistics (ACL))*, pages 271–278. ACL, 2004.

[PMC+20] Patrick Cheong-Iao Pang, Dana McKay, Shanton Chang, Qingyu Chen, Xiuzhen Zhang, and Lishan Cui. Privacy concerns of the australian my health record: Implications for other large-scale opt-out personal health records. *Information Processing & Management*, 57(6), 2020.

[PMFT22] Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. Challenges and opportunities in information manipulation detection: An examination of wartime Russian media. In *Findings of the 2022 Association for Computational Linguistics (EMNLP)*, pages 5209–5235. ACL, 2022.

[PMV20] George Panagopoulos, Fragkiskos D. Malliaros, and Michalis Vazirgiannis. Influence maximization using influence and susceptibility embeddings. In *Proc. 14th International AAAI Conference on Web and Social Media (ICWSM)*, pages 511–521. AAAI Press, 2020.

[PPC20] Han Woo Park, Sejung Park, and Miyoung Chong. Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. *Journal of Medical Internet Research*, 22(5):e18897, 2020.

[PSDN20] Annamaria Porreca, Francesca Scozzari, and Marta Di Nicola. Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination. *BMC Public Health*, 20(1):259, 2020.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[PVPCBH+21] Hilary Piedrahita-Valdés, Diego Piedrahita-Castillo, Javier Bermejo-Higuera, Patricia Guillem-Saiz, Juan Ramón Bermejo-Higuera, Javier Guillem-Saiz, Juan Antonio Sicilia-Montalvo, and Francisco Machío-Regidor. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. *Vaccines*, 9(1), 2021.

[QSZ+20] Jianyin Qiu, Bin Shen, Min Zhao, Zhen Wang, Bin Xie, and Yifeng Xu. A nationwide survey of psychological distress among chinese people in the COVID-19 epidemic: implications and policy recommendations. *General Psychiatry*, 33(2), 2020.

[QTM+18] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. Deepinf: Social influence prediction with deep learning. In *Proc. 2018 International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2110–2119, 2018.

[RA20] Julien Riou and Christian L Althaus. Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020. *Eurosurveillance*, 25(4):2000058, 2020.

[Ram03] Juan Ramos. Using tf-idf to determine word relevance in document queries. In *Proc./ 2003 International Conference on Machine Learning (ICML)*, volume 242, pages 133–142, 2003.

[RBZ20] Filipe N. Ribeiro, Fabrício Benevenuto, and Emilio Zagheni. How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. In *Proc. 2020 ACM Conference on Web Science (WebSci)*, pages 325–334. ACM, 2020.

[RFN17] Sara Rosenthal, Noura Farra, and Preslav Nakov. Semeval-2017 task 4: Sentiment analysis in Twitter. In *Proc. 11th International Workshop on Semantic Evaluation (SemEval)*, pages 502–518, 2017.

[RGGG18] Koustav Rudra, Niloy Ganguly, Pawan Goyal, and Saptarshi Ghosh. Extracting and summarizing situational information from the Twitter social media during disasters. *ACM Transactions on the Web*, 12(3), 2018.

[RRR20] Viju Raghupathi, Jie Ren, and Wullianallur Raghupathi. Studying public perception about vaccination: A sentiment analysis of tweets. *International Journal of Environmental Research and Public Health*, 17(10):3464, 2020.

[RRV13] Antonio Reyes, Paolo Rosso, and Tony Veale. A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation*, 47(1):239–268, 2013.

[RSM22] Maud Reveilhac, Stephanie Steinmetz, and Davide Morselli. A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. *Multimedia Tools and Applications*, 81:10107–10142, 2022.

[SA13] Christian Sturm and Hossam Amer. The effects of (social) media on revolutions - perspectives from Egypt and the Arab Spring. In *Proc. 15th International Conference of Human-Computer Interaction. (HCI)*, volume 8006 of *Lecture Notes in Computer Science*, pages 352–358. Springer, 2013.

[SA21] Naw Safrin Sattar and Shaikh Arifuzzaman. COVID-19 vaccination awareness and aftermath: Public sentiment analysis on Twitter data and vaccinated population prediction in the USA. *Applied Sciences*, 11(13):6128, 2021.

[SB07] Norbert Schwarz and Herbert Bless. Mental construal processes: The inclusion/exclusion model. *Assimilation and Contrast in Social Psychology*, pages 119–141, 2007.

[SBB$^+$20] Lisa Singh, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, Kornraphop Kawintiranon, Colton Padden, Rebecca Vanarsdall, Emily Vraga, and Yanchen Wang. A first look at COVID-19 information and misinformation sharing on Twitter. *arXiv preprint arXiv:2003.13907*, 2020.

[SBPD16] Michael C Smith, David A Broniatowski, Michael J Paul, and Mark Dredze. Towards real-time measurement of public epidemic awareness: Monitoring influenza awareness through Twitter. In *Spring Symposium on Observational Studies Through Social Media and Other Human-generated Content*, page e198, 2016.

[SC03] Mark P Sinka and David W Corne. Towards modernised and web-specific stoplists for web document analysis. In *Proc. 2003 IEEE/WIC International Conference on Web Intelligence (WI)*, pages 396–402. IEEE, 2003.

[Sca18] William J. Scarborough. Feminist Twitter and gender attitudes: opportunities and limitations to using Twitter in the study of public opinion. *Socius*, 4:2378023118780760, 2018.

[Sch11] Norbert Schwarz. Feelings-as-information theory. *Handbook of Theories of Social Psychology*, 1:289–308, 2011.

[SDCW19] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108, 2019.

[SF16] Kim Schouten and Flavius Frasincar. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830, 2016.

[SFG+15] Kaisong Song, Shi Feng, Wei Gao, Daling Wang, Ge Yu, and Kam-Fai Wong. Personalized sentiment classification based on latent individuality of microblog users. In *Proc. 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2277–2283. AAAI Press, 2015.

[SGE+22] Ramy Shaaban, Ramy Mohamed Ghazy, Fawzia Elsherif, Nancy Ali, Youssef Yakoub, Maged Osama Aly, Rony ElMakhzangy, Marwa Shawky Abdou, Bonny McKinna, Amira Mohamed Elzorkany, Fatimah Abdullah, Amr Alnagar, Nashwa ElTaweel, Majed Alharthi, Ali Mohsin, Ana Ordóñez-Cruickshank, Bianca Toniolo, Tâmela Grafolin, Thit Thit Aye, Yong Zhin Goh, Ehsan Akram Deghidy, Siti Bahri, Jarntrah Sappayabanphot, Yasir Ahmed Mohammed Elhadi, Salma Mohammed, Ahmed Nour El-Deen, Ismail Ismail, Samar Abd ElHafeez, Iffat Elbarazi, Basema Saddik, Ziad El-Khatib, Hiba Mohsin, and Ahmed Kamal. COVID-19 vaccine acceptance among social media users: a content analysis, multi-continent study. *International Journal of Environmental Research and Public Health*, 19(9), 2022.

[SGT+08] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

[SGT+09] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[SHL17] Chonggang Song, Wynne Hsu, and Mong-Li Lee. Temporal influence blocking: Minimizing the effect of misinformation in social networks. In *Proc. 33rd IEEE International Conference on Data Engineering (ICDE)*, pages 847–858. IEEE Computer Society, 2017.

[SHW+20] Shadi Shahsavari, Pavan Holur, Tianyi Wang, Timothy R Tangherlini, and Vwani Roychowdhury. 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, 2020.

[SPG+16] Michael F. Schober, Josh Pasek, Lauren Guggenheim, Cliff Lampe, and Federick G. Conrad. Social media analyses for social measurement. *Public Opinion Quarterly*, 80(1):180–211, 2016.

[SPXZ20] Mingwang Shen, Zhihang Peng, Yanni Xiao, and Lei Zhang. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in china. *BioRxiv*, 2020.

[SRI19] J Fernando Sánchez-Rada and Carlos A Iglesias. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, 52:344–356, 2019.

[SRZ+22] Ling Sun, Yuan Rao, Xiangbo Zhang, Yuqian Lan, and Shuanghe Yu. MS-HGAT: Memory-Enhanced sequential hypergraph attention network for information diffusion prediction. In *Proc. 36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4156–4164. AAAI Press, 2022.

[SSR20] Swaroop Gowdra Shanthakumar, Anand Seetharam, and Arti Ramesh. Understanding the socio-economic disruption in the United States during covid-19's early days. *CoRR*, abs/2004.05451, 2020.

[SSS+21] Susan M Sherman, Louise E Smith, Julius Sim, Richard Amlôt, Megan Cutts, Hannah Dasch, G James Rubin, and Nick Sevdalis. COVID-19 vaccination intention in the uk: Results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. *Human Vaccines & Immunotherapeutics*, 17(6):1612–1621, 2021.

[SSU+21] Rahul Shekhar, Abu Baker Sheikh, Shubhra Upadhyay, Mriganka Singh, Saket Kottewar, Hamza Mir, Eileen Barrett, and Suman Pal. COVID-19 vaccine acceptance among health care workers in the United States. *Vaccines*, 9(2):119, 2021.

[Str20] Ilse Struweg. A Twitter social network analysis: The South African health insurance bill case. *Responsible Design, Implementation and Use of Information and Communication Technology*, 12067:120, 2020.

[SVY+20] Mohsen Sayyadiharikandeh, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. Detection of novel social bots by ensembles of specialized classifiers. In *Proc. 2020 ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2725–2732. ACM, 2020.

[SWSB14] Hua-Wei Shen, Dashun Wang, Chaoming Song, and Albert-László Barabási. Modeling and predicting popularity dynamics via reinforced Poisson processes. In *Proc. 28th AAAI Conference on Artificial Intelligence (AAAI)*, pages 291–297. AAAI Press, 2014.

[SZ12] Connie St Louis and Gozde Zorlu. Can Twitter predict disease outbreaks? *British Medical Journal*, 344(7861), 2012.

[SZS+18] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi. Polarization of the vaccination debate on facebook. *Vaccine*, 36(25):3606–3612, 2018.

[TFU07] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. *Using Multivariate Statistics*. Pearson Education, 2007.

[TLH+21] Xiangyun Tang, Dongliang Liao, Weijie Huang, Jin Xu, Liehuang Zhu, and Meng Shen. Fully exploiting cascade graphs for real-time forwarding prediction. In *Proc. 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 582–590, 2021.

[TO21] Takanao Tanaka and Shohei Okamoto. Increase in suicide following an initial decline during the COVID-19 pandemic in Japan. *Nature Human Behaviour*, 5(2):229–238, 2021.

[TQL16] Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. In *Proc. 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 214–224. ACL, 2016.

[Tra95] Thanh V Tran. Bilingualism and subjective well-being in a sample of elderly hispanics. *Journal of Social Service Research*, 20(1-2):1–19, 1995.

[TRPP21] Chahat Tandon, Sanjana Revankar, Hemant Palivela, and Sidharth Singh Parihar. How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics. *International Journal of Information Management Data Insights*, 1(2):100035, 2021.

[Tur02] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424. ACL, 2002.

[TZY+08] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: Extraction and mining of academic social networks. In *Proc. 14th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 990–998. ACM, 2008.

[VBBB+20] Jay J Van Bavel, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, et al. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5):460–471, 2020.

[VCC+18] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *Proc. 6th International Conference on Learning Representations (ICLR)*. OpenReview, 2018.

[vSH20] Christian von Sikorski and Christina Herbst. Not practicing what they preached! exploring negative spillover effects of news about ex-politicians' hypocrisy on party attitudes, voting intentions, and political trust. *Media Psychology*, 23(3):436–460, 2020.

[vVTU20] Livia van Vliet, Petter Törnberg, and Justus Uitermark. The Twitter parliamentarian database: analyzing Twitter politics across 26 countries. *PLOS ONE*, 15(9):e0237073, 2020.

[WAD14] Kumanan Wilson, Katherine Atkinson, and Shelley Deeks. Opportunities for utilizing new technologies to increase vaccine confidence. *Expert Review of Vaccines*, 13(8):969–977, 2014.

[WCL18] Zhitao Wang, Chengyao Chen, and Wenjie Li. A sequential neural information diffusion model with structure attention. In *Proc. 27th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1795–1798, 2018.

[WCR+01] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, and Others. Constrained k-means clustering with background knowledge. In *Proc. 2001 International Conference on Machine Learning (ICML)*, pages 577–584. Citeseer, 2001.

[WDA+16] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9, 2016.

[WH08] Barry Wellman and Caroline Haythornthwaite. *The Internet in everyday life*. John Wiley & Sons, 2008.

[WHaPAG+19] Zijian Wang, Scott A. Hale, David Ifeoluwa Adelani afnd Przemyslaw A. Grabowicz, Timo Hartmann, Fabian Flöck, and David Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *Proc. 2019 The World Wide Web Conference (WWW)*, pages 2056–2067. ACM, 2019.

[WLJH10] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitters. In *Proc. 3rd ACM International Conference on Web Search and Data Mining (WSDM)*, pages 261–270, 2010.

[WLL20] Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.

[WNY20] Lei Wang, Jianwei Niu, and Shui Yu. SentiDiff: Combining textual information and sentiment diffusion patterns for Twitter sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 32(10):2026–2039, 2020.

[WPC+21] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021.

[WSL+17] Yongqing Wang, Huawei Shen, Shenghua Liu, Jinhua Gao, and Xueqi Cheng. Cascade dynamics modeling with attention-based recurrent neural network. In *Proc. 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2985–2991. IJCAI, 2017.

[WSLC15] Yongqing Wang, Huawei Shen, Shenghua Liu, and Xueqi Cheng. Learning user-specific latent influence and susceptibility from information cascades. In *Proc. 19th AAAI Conference on Artificial Intelligence (AAAI)*, pages 477–484. AAAI Press, 2015.

[WW20] Steven Lloyd Wilson and Charles Wiysonge. Social media and vaccine hesitancy. *BMJ Global Health*, 5(10), 2020.

[WWW+20] P Walker, C Whittaker, O Watson, M Baguelin, K Ainslie, S Bhatia, et al. Who collaborating centre for infectious disease modelling, mrc centre for global infectious disease analysis, abdul latif jameel institute for disease and emergency analytics. *The global impact of COVID-19 and strategies for mitigation and suppression. Imperial College London*, 2020.

[WZLC17] Jia Wang, Vincent W. Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. Topological recurrent neural network for diffusion prediction. In *Proc. 2017 IEEE International Conference on Data Mining (ICDM)*, pages 475–484. IEEE Computer Society, 2017.

[XCH+20] Jia Xue, Junxiang Chen, Ran Hu, Chen Chen, Chengda Zheng, Yue Su, and Tingshao Zhu. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of Medical Internet Research*, 22(11):e20550, 2020.

[XSX+20] Zhiping Xiao, Weiping Song, Haoyan Xu, Zhicheng Ren, and Yizhou Sun. Timme: Twitter ideology-detection via multi-task multi-relational embedding. In *Proc. 2020 ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 2258–2268. ACM, 2020.

[XZZ+23] Xovee Xu, Fan Zhou, Kunpeng Zhang, Siyuan Liu, and Goce Trajcevski. Casflow: Exploring hierarchical structures and propagation uncertainty for cascade prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3484–3499, 2023.

[YCW+15] Linyun Yu, Peng Cui, Fei Wang, Chaoming Song, and Shiqiang Yang. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. In *Proc. 2015 IEEE International Conference on Data Mining (ICDM)*, pages 559–568, 2015.

[YDM+21] Samira Yousefinaghani, Rozita Dara, Samira Mubareka, Andrew Papadopoulos, and Shayan Sharif. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 2021.

[Yeg15] Narine S Yegiyan. Explicating the emotion spillover effect. *Journal of Media Psychology*, 2015.

[YFR+20] Joseph Younis, Harvy Freitag, Jeremy S Ruthberg, Jonathan P Romanes, Craig Nielsen, and Neil Mehta. Social media as an early proxy for social distancing indicated by the covid-19 reproduction number: observational study. *JMIR Public Health and Surveillance*, 6:e21340, 2020.

[YGLC13] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. A biterm topic model for short texts. In *Proc. 22nd International Conference on World Wide Web (WWW)*, pages 1445–1456, 2013.

[YLZ+20] Chunyuan Yuan, Jiacheng Li, Wei Zhou, Yijun Lu, Xiaodan Zhang, and Songlin Hu. DyHGCN: A dynamic heterogeneous graph convolutional network to learn users' dynamic preferences for information diffusion prediction. In *Proc. 2020 Machine Learning and Knowledge Discovery in Databases - European Conference (ECML PKDD)*, pages 347–363. Springer, 2020.

[YNF+20] Huso Yi, Shutian Ng, Aysha Farwin, Amanda Pei Ting Low, Chengmun Chang, and Jeremy Lim. Health equity considerations in COVID-19: Geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *Journal of Travel Medicine*, 28(2), 2020.

[YS16] Chao Yang and Padmini Srinivasan. Life satisfaction and the pursuit of happiness on Twitter. *PLoS One*, 11(3):e0150881, 2016.

[YTS+19] Cheng Yang, Jian Tang, Maosong Sun, Ganqu Cui, and Zhiyuan Liu. Multi-scale information diffusion prediction with reinforced recurrent networks. In *Proc. 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4033–4039, 2019.

[Zar20] John Zarocostas. How to fight an infodemic. *The Lancet*, 395(10225):676, 2020.

[ZEH+15] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proc. 2015 International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1513–1522, 2015.

[ZFCN22] Zihan Zhang, Meng Fang, Ling Chen, and Mohammad-Reza Namazi-Rad. Is neural topic modelling better than clustering? An empirical study on clustering with contextual embeddings for topics. In *Proc. 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 3886–3893. ACL, 2022.

[ZJZ20] Xinyi Zhou, Shengmin Jin, and Reza Zafarani. Sentiment paradoxes in social networks: Why your friends are more positive than you? In *Proc. 14th International Conference on Web and Social Media (ICWSM)*, pages 798–807. AAAI Press, 2020.

[ZL17] Junlong Zhang and Yu Luo. Degree centrality, betweenness centrality, and closeness centrality in social network. In *Proc. 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, pages 300–303. Atlantis Press, 2017.

[ZLT+13] Jing Zhang, Biao Liu, Jie Tang, Ting Chen, and Juanzi Li. Social influence locality for modeling retweeting behaviors. In *Proc. 23rd International Joint Conference on Artificial Intelligence (IJCAI*, pages 2761–2767, 2013.

[ZLZZ21] Lingyun Zhao, Lin Li, Xinhao Zheng, and Jianwei Zhang. A BERT based sentiment analysis and key entity detection approach for online financial texts. In *Proc. 2021 IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1233–1238. IEEE, 2021.

[ZXL+22] Chunyan Zhang, Songhua Xu, Zongfang Li, Ge Liu, Duwei Dai, Caixia Dong, et al. The evolution and disparities of online attitudes toward COVID-19 vaccines: year-long longitudinal and cross-sectional study. *Journal of Medical Internet Research*, 24(1):e32394, 2022.

[ZYXT21] Fan Zhou, Liu Yu, Xovee Xu, and Goce Trajcevski. Decoupling representation and regressor for long-tailed information cascade prediction. In *Proc. 44th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 1875–1879. ACM, 2021.

[ZYZ+20] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *Proc. 2020 Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 33, pages 7793–7804. NeurIPS, 2020.

Curriculum Vitae

2019 – 2023	Ph.D. student, University of Luxembourg, Luxembourg.
2017 – 2018	MSc in Data Science, University of Sheffield, The United Kingdom.
2011 – 2015	Bachelor of Public Administration, Central University of Finance and Economics, China.

Born on April 8, 1994, Guiyang, China.