Oxford Internet Institute, University of Oxford

Assignment Cover Sheet

Candidate Number	1051660
Assignment	SDS MSc Thesis
Term	Trinity 2021
Title / Question	Going the Distance: Information Diffusion and Influence Dynamics of Climate Science on Twitter
Word Count	13,498

By placing a tick in this box ☒ I hereby certify as follows:

a. This thesis or coursework is entirely my own work, except where acknowledgments of other sources are given. I also confirm that this coursework has not been submitted, wholly or substantially, to another examination at this or any other University or educational institution;

b. I have read and understood the Education Committee's information and guidance on academic good practice and plagiarism at https://www.ox.ac.uk/students/academic/guidance/skills?wssl=1.

c. I agree that my work may be checked for plagiarism using Turnitin software and have read the Notice to Candidates which can be seen at: http://www.admin.ox.ac.uk/proctors/turnitin2w.shtml}, and that I agree to my work being screened and used as explained in that Notice;

d. I have clearly indicated (with appropriate references) the presence of all material I have paraphrased, quoted or used from other sources, including any diagrams, charts, tables or graphs.

e. I have acknowledged appropriately any assistance I have received in addition to that provided by my [tutor/supervisor/adviser].

f. I have not sought assistance from a professional agency;

g. I understand that any false claims for this work will be reported to the Proctors and may be penalized in accordance with the University regulations.

Please remember:

- To attach a second relevant cover sheet if you have a disability such as dyslexia or dyspraxia. These are available from the Higher Degrees Office, but the Disability Advisory Service will be able to guide you.



Going the Distance: Information Diffusion and Influence Dynamics of Climate Science on Twitter

1051660

13,498 words

Thesis submitted in partial fulfilment of the requirement for the degree of MSc in Social Data Science at the Oxford Internet Institute at the University of Oxford

Trinity Term

August 2021

Acknowledgements

I would like to extend my gratitude to my supervisor, Dr. Balazs Vedres, for his constant and valuable insights, guidance, and feedback, for which this thesis would not be possible without. I also wish to thank the networks reading group for the interesting discussions and conversations throughout the year, which initially sparked my interest in researching social networks. I also would like to thank my colleagues, in particular, Yennie and Mira for their remarkable input in earlier drafts, and their unwavering support throughout the research period. Lastly, I would like to thank this year's cohort and all the faculty and staff at the Oxford Internet Institute, for making this year possible despite all the uncertainties and challenges of adapting to constant changes during COVID-19.

Abstract

An increasing distrust in scientists and experts has led to the polarisation of science. While there is a strong scientific consensus on human-induced global warming, ideological differences have led to public uncertainty around climate change and the polarisation of acceptance and rejection of the scientific authority. Climate change policies and net zero-emission goals become sources of contentious debates in the public, where the scientific evidence of climate change and the reliability of the science community are questioned. This has led to climate misinformation driven by the information asymmetry between climate experts and the public, which are further exacerbated by distortions of scientific framing in the media, or by the climate change countermovement.

While previous literature has studied the public opinions of climate sceptics and misinformation in the media, this dissertation considers the diffusion of climate science in climate change opinions and how climate misinformation spreads in online social networks. Through data gathered from Twitter, this research utilises temporal network analysis to achieve two research objectives: to analyse the diffusion patterns of scientific information in pro and anti-climate groups; and to understand the structural patterns of communities engaging in climate-related posts and their effect on the creation and spread of scientific-based ideas. In doing so, this thesis presents the current state of science-based discourse in climate change and sheds a light on the main challenges of climate misinformation and science communication, namely, the inherent uncertainty of scientific discovery.

Topical keywords: climate change communication, trust in science
Methodological keywords: network analysis, information diffusion, complex networks

Table of Contents

Acknowledgements 3
Abstract 4
Table of Contents 5
List of Table 7
List of Figures 8
1. Introduction 9
1.1. Research Questions 10
2. Literature Review 12
2.1. Science in climate change 12
2.3. Information Diffusion 15
2.4. Homophily and social influence 16
2.5. Information Cascade Networks 17
2.6. Information diffusion and related research on Twitter 18
2.7. Summary 20
3. Data and Methods 21
3.1. Data Collection 21
3.1.1. Climate Feedback Data 21
3.1.2. Twitter Data 23
3.1.3. Botometer Data 25
3.2. Data pre-processing 27
3.3. Network set-up 28
3.3.1. Cascade network ($G_{[URL]}$) 28
3.3.2. Retweet network (G_{RT}) 29
3.4. Network Analysis 30
3.4.1. Temporal properties 30
3.4.2. K-shell decomposition 30
3.4.3. Edge homophily measures 32
3.5. Limitations and Future work 32
3.6. Summary 33
4. Findings and Analysis 34

4.1. Summary Statistics 34
4.1.1. Article cascade lifetime and cascade size 35
4.1.2. RT cascades 36
4.2. Information Cascades Network 37
4.2.1. Structural properties 39
4.2.2. Temporal properties 41
4.2.3. Role of Media 42
4.3. Summary of user engagement 43
4.4. Re-tweet Network Topology 44
4.5. User Influence 45
4.6. Edge homophily 51
4.7. Influence of automated accounts 52
4.8. Patterns of misleading information 55
4.9. Summary 57
5. Discussion 59
5.1. Networks of climate information diffusion 59
5.2. The diffusion of credible and misleading climate information 60
5.3. Contributions and Recommendations 61
6. Conclusion 63
7. References 64
8. Appendices 72
8.1. Appendix A – Publishing media and their number of articles in the data 72
8.2. Appendix B 74
8.3. Appendix C – Temporal distance of proportion of tweets with CCDF plots based on media 75
8.4. Appendix D – Accounts in the k-core-shell of retweet network who shares low rated climate articles 76

List of Table

Table 1 – Summary statistics of Twitter data 24
Table 2 – Difference of distribution between high and low category 42

List of Figures

Figure 1: Distribution of articles based on rating value (right) and year published (left) 23
Figure 2: Volume of tweets across time 24
Figure 3: Bot and CAP scores showing the likelihood of accounts being automated (CAP) for each bot score 27
Figure 4: Distribution of bot scores for the top five percent node centrality 27
Figure 5: Illustration of how the two network was constructed from the data collection .. 29
Figure 6: Illustration of K-core decomposition revealing central nodes (Kitsak et al., 2010) 31
Figure 7: Distribution of Cascade Sizes 34
Figure 8: Probability density function (PDF) of cascade lifetime hours 35
Figure 9: Lifetime hours and user reached on each rating level 35
Figure 10: Number of RT cascades in each article rating value 36
Figure 11: Article-level network structure 38
Figure 12: Empirical CCDF plot of tweets based on duration (in minutes) since first Tweet 38
Figure 13: Structural properties of article cascade networks 40
Figure 14: Empirical CCDF plot of Tweets on time since first post 41
Figure 15: Tweet count per user 43
Figure 16: Degree distribution of RT cascade 44
Figure 17: Pruned whole retweet network 45
Figure 18: Misinformation network in the pruned whole retweet network (k=5) 45
Figure 19: K core network (K=15) 47
Figure 20: Influential nodes in K core network who shares both misleading and 49
Figure 21: Scientifically credible influencers 50
Figure 22: Edge homophily – the probability of sharing multiple articles from 51
Figure 23: Automated accounts' influence on the diffusion of credible and misleading information 54
Figure 24: Example of misleading articles in the "low" category 56
Figure 25: Tweet snippets from three types of climate misinformation 57

1. Introduction

Social media have changed how people find, consume, and are exposed to information, with access to information becoming more accessible. Online information sharing behaviour has shifted information consumption behaviours from broadcasted media to networked information sharing. In climate change discourse, the role of information is critical in shaping the public's understanding of the current climate. Accurate information based on climate science becomes a foundation for understanding how climate change affects people's lives, how the climate changes, and the possible steps to slow irreversible climate changes. Therefore, climate science plays a significant role in informing climate decisions through objective scientific evidence that delivers the opportunities and risks associated with the consequences of climate actions and inactions.

The impact of scientific findings in public is limited (Scheufele & Krause, 2019). Through a study of word frequencies and diffusion, the public usage of climate-science keywords follows the "boom and bust" fashion cycle, which indicates the impact of climate science on society (Bentley et al., 2012). At the core of this problem, the difficulties of public communication of science persist in climate change and climate science. Understanding climate science is integral to differentiate between accurate and misleading climate change information. Moreover, the gradual process of compiling scientific findings, and how it is used to generate scientific consensus and support theories, often complicates the public's understanding of climate change.

This paper uses the term climate change for climate change discourse, information sharing, and the diffusion dynamics of climate-related contents. The climate change information discourse is very prone to misinformation, as a result of the complexity of climate science for laypeople. The rise of pseudo-science, climate myths, or cherry-picking evidence has complicated the dynamic of climate information, thereby blurring the distinction between accurate and misleading information. This is where the strategy of climate disinformation grows, through exaggerating the uncertainties of scientific findings, some of which are correlated with the fossil fuel giants. Furthermore, social media becomes an effective platform for the spread of climate misinformation and disinformation.

Researchers have extensively studied the spread of misinformation in online social networks (del Vicario et al., 2016; Vosoughi et al., 2018), particularly to understand echo chamber effects (Törnberg, 2018), predict misinformation spread through modelling (Krishna Kumar & Geethakumari, 2013), and to contain misinformation online (Pennycook et al., 2021; Shao et al., 2018). The majority of existing research is focused on public health-related (Nsoesie & Oladeji, 2020; Roozenbeek et al., 2020) and politics related misinformation (Allcott & Gentzkow, 2017; Bovet & Makse, 2019), with only a small number of studies on the diffusion of climate change misinformation. Current studies of misinformation in climate change found echo chamber effects and polarisation between climate skeptics and activist groups (Farrell, 2016a; K. M. d. I. Treen et al., 2020; Williams et al., 2015), ways to inoculate the public from climate misinformation (Cook et al., 2017; S. van der Linden et al., 2017), and the challenges and effects of science understanding in understanding climate change (Kahan et al., 2012; Palmer & Stevens, 2019).

The public understanding of science, and its dynamic with science journalism and the scientific community, showcase the importance of science communication in climate change that contributes to shaping public views and policymaking. The study of how climate misinformation spreads in online social networks sheds a light on the damaging effects of climate misinformation campaigns and the role of climate science in fighting climate misinformation.

1.1. Research Questions

This research explores and contrasts the diffusion networks of credible and misleading climate information based on climate science on Twitter, a popular micro-blogging service. Through social network analysis, two research questions are considered.

> **RQ1:** How does social network analysis reveal the underlying network structure of the diffusion mechanisms of climate misinformation online?
>
> **RQ2:** How does the spread of credible and misleading climate information based on its scientific credibility differ in social media?

The focus of this research is on the dynamics of user influence and scientific credibility on the spread of climate change information online. To understand this, the differences of

content influence and user influence in climate change were studied. The value of the content relies on its scientific credibility and media popularity, while user influence relies on the user's audience and engagement on Twitter, which collectively affects information diffusion. Through *temporal* and *structural* network analysis, this research examines the spreading mechanism of reliable, science-based climate information, and misleading climate information. Furthermore, to understand the dynamics of climate information diffusion, I use edge homophily measures to analyse how people engage with different information and the influence of automated Twitter accounts in climate information diffusion. The results show overlapping communities of people sharing articles with high and low scientific credibility. Hence, the diffusion characteristics of misleading and credible climate content are found to be similar. Extending previous findings in climate misinformation and science communication, this research highlights the audience impact of climate change research in climate change discourse online and discovers people's behaviour in engaging with climate content with different rates of scientific credibility.

This thesis is structured as follows. With the guiding research questions above, a literature review of related theories and existing work on science expert systems, information diffusion, social networks, and misinformation in climate change are explored. Further, the research design, including data collection, methodology, and analysis plan are explained. Finally, the findings and discussion of diffusion networks of credible and non-credible climate information are presented, with limitations and recommendations for future research.

2. Literature Review

This chapter contextualises the importance of science in climate change misinformation and establishes the foundational theories and methodologies essential for this research. The chapter begins by introducing the main factors and issues in climate change communication through the lens of scientific consensus and uncertainties in science, then follows with a survey of existing theories in information diffusion, misinformation, and online social networks studies.

2.1. Science in climate change

There are three main areas where science is driving the course of climate change. Understanding the climate, predicting the climate, and actions for the climate. Climate science helps us to understand the processes and mechanisms of the planet's climate system, how these change over time, and its likely effects on humans and natural influences (Moss et al., 2010). These observations enable the growing research of climate models, which inform the impacts of climate change, assess the vulnerabilities of these impacts, and create model-based predictions. Furthermore, climate change research uses climate models to inform climate mitigation and adaptation policies (e.g. by developing plausible socio-economic, emissions, and climate model scenarios). Such research integrates various disciplines. For example, Rockström et al. (2009) developed the concept of planetary boundaries, which identifies and quantifies biophysical thresholds to accommodate human development in a changing planet. This framework integrates the development of human societies and the earth system, which provides a science-based analysis of the risk that human perturbations will destabilise the earth system at the planetary scale (Steffen et al., 2015). These developments assess how humans will respond to changes in technology, policy, and lifestyle changes, in addition to climate implications for the environment.

The most comprehensive climate predictions and analyses are summarised in synthesis reports of the full scientific, technical, and socio-economic assessments of climate change,

produced by The Intergovernmental Panel on Climate Change (IPCC)[1]. The IPCC defines the state of knowledge in climate change and identifies areas where there is a strong scientific consensus and areas where further research is needed. IPCC reports become practical guidelines for governments and help to shape international climate change negotiations, such as The Paris Agreement and The UN Climate Conference. The role of experts and climate scientists is crucial in the fight to reduce carbon emissions to reach the goal of 1.5 degrees above pre-industrial level, which is predicted to be the safest temperature increase that does not disrupt earth systems by 2100 (IPCC, 2014).

A study of global peer-reviewed papers found that the scientific consensus of human-induced anthropogenic global warming has reached 97 percent from multiple studies and scientists who publish climate research (Cook et al., 2013, 2016; Oreskes, 2004; Reusswig, 2013). However, despite the strong consensus, the perceived scientific agreement is undermined in the public (Hamilton, 2016; McCright, Dunlap, et al., 2013). This discrepancy led to discourse in the perceived consensus, which subsequently lowers the support for climate policies (Cook et al., 2017; S. L. van der Linden et al., 2015). Therefore, to enable strong support for widespread climate policies, the discrepancies of perceived scientific consensus in climate change research need to be addressed.

Although the scientific consensus is clear, the public has various risk perceptions of climate change, due to socio-cultural influences, socio-demographic characteristics, and trust in science (Kahan et al., 2012). The public opinion towards climate change is polarised due to the differences in individual values, distrusts in scientists and experts, and climate misinformation dissemination for political and economic objectives (Farrell, 2016). Moreover, the crisis of public trust in science and experts has become one of the key barriers for effective climate change communication as the core of climate change discourse (Eyal,

[1] The IPCC was created in 1988 by the World Meteorological Organization and the UN Environment Programme (UNEP) to assess the impacts and future risks of climate change, and to provide a scientific basis for governments to create climate related policies (IPCC, 2021)

2019). Therefore, conflicting interests in climate change issues, such as moral, religious, economic, and political interest, plays a role in discrediting experts or undermining the risk perceptions of climate change. This is supported by Kahan et al.'s (2012) study where people with high scientific literacy and numeracy have a lower perception of climate change risk than predicted.

The link between climate science and climate misinformation stems from strategies by the fossil fuel industry to induce scientifically inaccurate climate change information in public (Oreskes & Conway, 2010). This is known to be one characteristic of misinformation, which is to cancel out the positive effects of accurate information (Cook, 2019). Many anti-climate groups discredit science by cherry-picking evidence without contextualising the evidence with other mechanisms in climate science (Cook et al., 2017). Similar to the tobacco strategy, the act of casting doubts on a large consensus of scientific evidence distorts the scientific framing of climate change and is used to attack climate science (Oreskes & Conway, 2010).

2.2. Science as a social process

Trust in science should not be on scientists as individuals but as a social process that rigorously vets claims (Oreskes, 2021). Similar to the processes of peer-reviewed journals, scientists should be open to criticism, and the possibility of falsification. In the philosophy of science, the standard of evaluation of scientific theories is the criterion of falsifiability, where a theory is genuinely scientific only if it is possible in principle to establish that it is false. In this case, it should be possible to establish that the claim of naturally occurring global warming that is not man-made, is false in principle.

As proposed by Popper (1959), a way of demarcating science from non-science is that the theory needs to be able to be tested and conceivably proven false. Thus, science as a social process should always attempt to disprove a theory rather than continually support theoretical hypotheses. No matter how many observations are made to confirm a theory, there is a possibility that a future observation could refute it, and thus, this highlights the main attribution of science, where it is always contextualised with time. This is where science plays a significant role in distinguishing accurate and misleading climate information.

The complexity of climate science literature has contributed to the importance of science communication in public understandings of science. Daily and accessible articles such as news articles are as important as other scientific reports, or even more important to communicate with the public. Due to the importance of successful science communication in climate change, this research explores the spread of science-based and non-science-based climate information in the media. The following parts explore the literature on information diffusion.

2.3. Information Diffusion

The study of information diffusion first emerged from the study of diffusion of innovation. The first study on the spread of medical innovation and drug use (Coleman et al., 1966) becomes evidence of social contagion in the diffusion of technological innovation. Extending the results of this study, Burt (1987) found the effect of social contagion and social structure on diffusion through reanalysing the data with network theory and the social structural properties. Ever since other studies also incorporate social structure to diffusion models as the usual assumption of homogeneous mixing is sociologically uninformative (Centola, 2015; Valente, 1996a, 1996b). In diffusion studies, Strang (2016) found that embedding social structure in the diffusion analysis enriches the substantive insight into the mechanisms affecting the diffusion.

The study of information diffusion reflects the relationship between social structure and cohesion. Diffusion models are often used in the study of epidemiology to model the disease spread (e.g. the Bass model (1969) (S-shaped curve); the R number) as well as to study the reproducibility of information, as evidence of how information propagates in society. Although the informational value affects the diffusion outcome, other factors may greatly affect information diffusion irrespective of the content. For example, when information spreads in a social network, the structure of the underlying network becomes a global property that plays a factor.

In a survey of information diffusion in online social networks, the main research areas found were detecting popular content, modeling information diffusion, and identifying influentials (GuilleAdrien et al., 2013). Related works in information diffusion studies include the knowledge transfer from information networks such as academic citations, email chain

letters, blog diffusion studies (Gruhl et al., 2004), Facebook sharing networks (Bakshy et al., 2012), and Twitter (Alperin et al., 2019; Cha et al., 2012; A. Goel et al., 2015; Taxidou & Fischer, 2014). To study information diffusion in networked communication, researchers consider the measure of node centrality, which reflects the traffic flows in a network. As one of the most studied concepts in social network analysis, centrality offers a new interpretation of network flow, which affects the scale of node reachability in information diffusion (Borgatti, 2005).

2.4. Homophily and social influence

As a social contagion phenomenon, successful information diffusion in networks requires minimal structural consolidation (Centola, 2015). In social network theory, this refers to homophily, where there is a higher probability of ties between similar people compared to dissimilar people. In turn, homophily affects the social structure of information exchange, transfer, and other types of ties (McPherson et al., 2001). This has significant implications on the information people receive, as well as associated attitudes and behaviours formed as a result of that information, especially with ideologically homogenous personal networks. Homophily in social networks lead to the

A study of information diffusion further confirms the effect of social structure on information-sharing behaviour (Bakshy et al., 2012). This experiment of online information diffusion found that people are more likely to share information when they are exposed by their friends. Shalizi and Thomas (2011) argued the three factors of processes in social networks, which are social contagion (social influence), homophily (the formation of social ties due to matching individual traits), and the causal effect of an individual's covariate on their behaviour. However, in this study, they found the confounding effects of homophily and social influence in observational studies as distinguishing one from the other requires a strong assumption.

The issue of endogenous effects in observational data about human behaviour and information spread persist in many studies (Aral et al., 2009). As a result, the solution is to identify and control for the latent homophily, estimate community memberships by creating a modular network structure to reduce the confounding. Many of these social mechanisms of contagion, such as information spread, are a result of homophily and social influence.

Social media has the echo chamber effects, where repeated exposure to a particular content increases the likelihood of information sharing (Dow et al., 2013). In this case, the mixed effects of influence and homophily amplify its content influence.

An experimental study has considered the impact of informational value on sharing behaviour on Twitter. The findings highlighted that high informational value led to more RT-ing than low informational value news, indicating the effect of informational value is strong and stable (Rudat et al., 2014). This concept further validates that non-journalistic Internet users have become news multipliers, moreover, with identity and self-representation context in social media that amplifies efficient communication behaviour.

2.5. Information Cascade Networks

Generally, social networks represent the relationships between people, where nodes represent people and edges represent the ties. In this research, the edges specifically represent information exchanges and mutual information sharing. One example is a cascade network, which links people based on the direction of information flow between them. An aggregated cascade network reveals the paths that information travels, revealing a layer on top of the actual social network. A cascade is defined as a tree structure where the "seed" node triggers adoptions of behaviour (S. Goel et al., 2012). Goel (2012) found the common patterns of online diffusion cascades where large cascades are infrequent and do not resemble the multi-step patterns of epidemics arising from theoretical models of diffusion from other studies (Kempe et al., 2003; Watts, 2002).

Some of the literature has focused on cascade networks to study the influence paths of information sharing (Cheng et al., 2016; Dow et al., 2013). Information cascades are more certain to study rather than connectivity between users as they are more dynamic and can easily change with sudden bursts of information. Cascade networks are categorised as implicit networks, as it is constructed with a subset of a social network, which becomes an accurate indication of trust relationship (Bild et al., 2015). They argue that connections on the cascade network are better indications than follower/following relationships as they are created by sharing actions. This is also supported in another study that found sudden bursts of connectivity between users (follow/unfollow link) that correlates with information sharing dynamics on Twitter (Myers & Leskovec, 2014).

This research will focus on tracking and analysing information cascades on Twitter, on specific climate-related content. Using the diffusion network created from RT data, the analysis will be based on the direction of information flow between users.

2.6. Information diffusion and related research on Twitter

Twitter is one of the most popular micro-blogging platforms, which enables up to 280-character posts in a single tweet. This platform was selected as the subject of this research as it was built to disseminate information. Furthermore, Twitter's ecosystem is suitable to study the influence of content and context in information diffusion (Rudat & Buder, 2015). A Twitter-specific feature that makes news spread quickly and widely is the re-tweet mechanism, which is responsible for virality and the spread of real-time information.

Several studies have analysed polarisation and echo chamber effects between pro and anti-climate groups on Twitter (Cann et al., 2021; Farrell, 2016a, 2016b; K. M. d'I. Treen et al., 2020; Williams et al., 2015; Xia et al., 2020). Many of the findings highlight the great divide of topics such as skeptic and alarmist language in climate topics (Cann et al., 2021), corporate and institution's involvement in contrarian information (Farrell, 2016a), and the themes of 'attack' on pro-climate activist (Xia et al., 2020). Xia et al. (2020) also found that the themes of climate science (i.e. scientific evidence) are occurring less and spread less virally in groups on Twitter, whereas 'movement' themes spread more virally. From this recent study, they highlighted that echo chambers in climate discussions get further consolidated and separated, however, the lack of communication between groups of climate activists and skeptics persists.

On top of the low scientific content in climate change on Twitter, the diffusion of science in both pro and anti-climate groups are similar (Xia et al., 2020). Thus, the spread of science (or the lack thereof) amongst the public can be impacted by various factors, including the difficulty to understand science, neglection of climate topics in general, or polarisation over science (Bessi et al., 2015; McCright, Dentzman, et al., 2013; Rapport et al., 2018). The origins of the polarisation of science are found to be a mix of conflicting political ideologies, facts, and scientific claims (Rekker, 2021). Science polarisation can be explained by psychological, deriving from non-political identities such as religion, or ideological rejection, which may also be derived from non-political belief systems such as conspiracy theories, or

both. While the attitude towards science is not the focus of this research, this highlights the underlying reasons for science polarisation, which explains the basis of the diffusion of climate science in online social networks.

In a study, three main influence measures on Twitter are indegree (no. of followers), retweets (RTs), and mentions (Cha et al., 2010). While the number of followers gives a good illustration of one's popularity, Cha et al. proposed the importance of engagement with the audience through RTs and mentions to measure one's influence. In addition to user influence, an experimental study has considered the impact of informational value on sharing behaviour on Twitter (Rudat et al., 2014). The findings highlighted that high informational value led to more RT-ing than low informational value news, indicating the effect of informational value is strong and stable. This concept further validates that non-journalistic Internet users have become news multipliers, moreover, with identity and self-representation context in social media that amplifies efficient communication behaviour.

In social network analysis, the observable interactions between people, which leads to the spread of information online, heavily rely on how the platform is developed. Generally, information networks on the internet have a complex network topology, which is defined as a large-scale system that is characterised by its power-law degree distribution. In such networks, the scale-free state is affected by the growth and preferential attachment, responsible for the power-law networks scaling (Barabasi, 2000). This highlights the large chance of highly connected nodes occurring, dominating the connectivity of the network. Therefore, the dynamics of user influence and informational value in complex networks are different from traditional information spread offline.

Contrary to how information traditionally spreads, the psychological behaviour of re-tweeting information relies on user interaction with Twitter as a platform. In a study of how users interact with their timeline, a few factors that were identified to have influence user RT probability are previous responses to the same tweeter, tweeter's sending rate, the age of the tweet, and other text elements of a tweet (Comarela et al., 2012). From many studies, there are a few ways that lead to information sharing on Twitter. First is through a user's following's post that appears on one's feed, second is through hashtags, and third is through manual search of content. As a good platform to study the dissemination of information

online, many other similar studies used Twitter to analyze the spread of political misinformation in online social networks (Vosoughi et al., 2018).

2.7. Summary

The literature in climate communication, information diffusion, and social network analysis vary extensively in its methods, objectives, and context of information. In particular, I believe these are the gaps that need to be addressed. First is the gap of the study of climate misinformation based on robust scientific evidence in climate communication. The public perception of climate science in the media and social media, and its relation to the scientific community highlights the need to inform the public better in what is climate science, its importance, and the important role of science journalists that bridges between the public and the science community.

The second gap in the literature is the use of social network analysis in climate information studies. While there has been extensive literature on ways to tackle climate misinformation through inoculation, or identifying tactics of climate disinformation, the current landscape of the spread of climate information in public is limited. The application of social network analysis and information diffusion in climate information is what this paper seeks to contribute. With this method, it uncovers the mix of content shared between users and the influence of those users who effectively disseminate information to a broad audience online. The question of audience impact of science-based climate change communication is lacking in the literature. This is what this research aims to contribute, using social network analysis and science-based annotation.

3. Data and Methods

This chapter details the data collection and analysis approaches used in this research, comprising data collection, pre-processing, and network analysis. First, the data collection methods of different data sources are explained, and the summary statistics of each dataset are presented. Second, the network setup and analysis are discussed. The chapter concludes with limitations regarding data collection.

3.1. Data Collection

The analysis utilised three main datasets: climate change-related articles that are rated and reviewed by a group of scientists from Climate Feedback; Twitter data with the URLs of the selected articles; and Botometer data for the likelihood of automated Twitter accounts. These three datasets are described as follows.

3.1.1. Climate Feedback Data

Climate Feedback is a part of Science Feedback, a non-profit organization trying to improve the credibility of scientific claims online by reviewing and verifying the credibility of its scientific basis. It is certified by the International Fact-Checking Network (IFCN) by the Poynter Institute[2] who enforces broad standards of fact-checking worldwide. It upholds the Code of Principles of IFCN, which commits to fairness and non-partisanship of review outcomes, transparency of funding sources, methodology, and correction policies in fact-checking processes (IFCN, 2021). Climate Feedback was selected and is used as the basis of this research because it is one of the few fact-checking organisation that focuses on the scientific credibility of climate articles with a climate science lens. Although this results in a skewed definition of scientific credibility, Climate Feedback's review process encompasses a holistic review process from more than one scientists. Notable, the organisation has been

[2] https://ifcncodeofprinciples.poynter.org/profile/science-feedback

appointed by Facebook and Poynter to fight climate and COVID-related misinformation by building a database of websites and social media sources that actively spread misinformation (Science Feedback, 2019, 2020).

The fact-checking process proceeds as follows. First, scientists, who at least hold a Ph.D. in climate science and have published articles in peer-reviewed science journals, are invited to review climate information online through a science-based analysis perspective[3]. Each article is reviewed by multiple scientists and rated based on scientific credibility from 1, being the least credible, and 5, being the most credible. From a few ratings, an editor aggregates all rating values and provides an averaged rating. Some articles with a wide discrepancy in the rating values are then published as 'debated'.

Climate Feedback relies on a web annotation tool, *Hypothes.is*[4], to annotate climate articles directly on the web. Therefore, the data collection utilises the Hypothesis API[5], which contains all metadata reviews, including the URL of the reviewed article, the name of the reviewer, the rating value, and the author. To get more comprehensive data, this paper collected all reviews by scraping the Climate Feedback website with the *BeautifulSoup* Python library (Richardson, 2007) to acquire the references to scientific journals that the reviewers used to support their feedback, the name, title, and the institutional affiliation of the reviewers and editors, and the article metadata such as headline, date of published, and name of the publisher. There were 20 missing articles metadata, which was later collected with the *Newspaper* Python library[6] to get the date of publication, author, and name of the publisher.

There were a total of 156 article reviews gathered from the Climate Feedback's website and Hypothesis API made of articles from 47 media companies that were published from 2015 to 2021. Some of the reviews were based on non-article content, such as Facebook posts or

[3] https://climatefeedback.org/process/
[4] https://web.hypothes.is/partners/climate-feedback/
[5] https://web.hypothes.is/blog/lets-use-the-annotated-web-to-coordinate-the-struggle-against-fake-news/
[6] https://newspaper.readthedocs.io/en/latest/

YouTube videos, which were excluded from the analysis. The final dataset contains 152 articles with a balanced mixture of op-ed, news, and analysis articles from both left-leaning and right-leaning media companies, including, but not limited to, *The Guardian, New York Times, The Sun, Breitbart* (Appendix A). The distribution of article rating values and date of publication is illustrated in Figure 1.

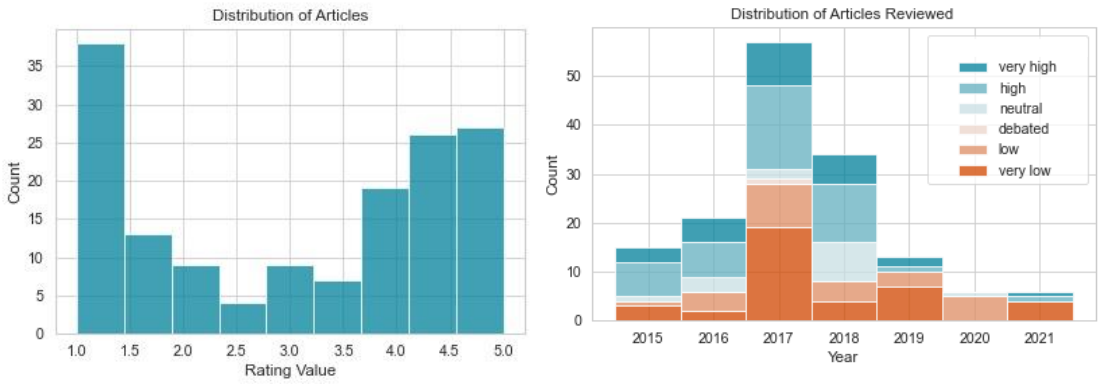


Figure 1: Distribution of articles based on rating value (right) and year published (left)

3.1.2. Twitter Data

The second dataset comprises all historical tweets gathered from Twitter's Academic Research API. All tweets and re-tweets containing the URLs from the selected articles were gathered using the filter `(url: "www.xxx.com")`, which matched tweets from 16 January 2015 to 21 June 2021. The distribution of the volume of tweets from 2015 to 2021 is illustrated in Figure 2, where there are a few spikes of tweets between 2017 and 2019. All user data were collected, which included the number of followers, followings, tweets, date joined, verified status, and bio.

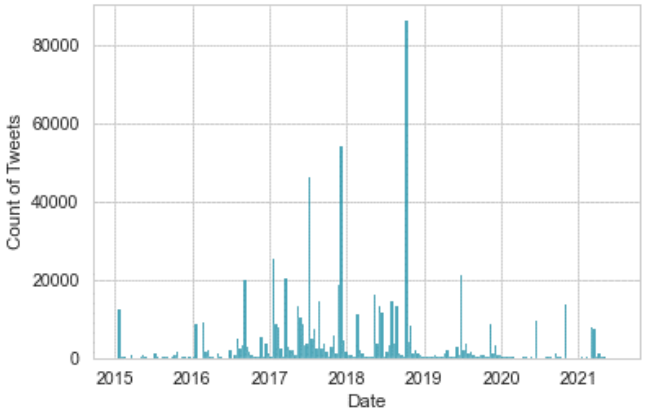


Figure 2: Volume of tweets across time

This dataset comprises a total of 740,356 English tweets, which link to the 152 articles annotated by Climate Feedback. This includes all tweet types; 76.22% of the dataset are re-tweets of the 39,219 original tweets. The rest of the dataset includes original tweets with URLs but never got re-tweeted and tweets that mentioned the URLs as a reply to another tweet in a thread. Each tweet that is re-tweeted more than once is categorised as a successful diffusion, where the user is categorised as a root user and the tweet, the root tweet. There were 19,848 successful diffusion events identified in the data, which makes up about 2.68% of the total tweets data gathered. In this data, there were 411,361 public user accounts, of which, 2.30% were verified users. The summary statistics of the tweets and user data are shown in Table 1.

Table 1 – Summary statistics of Twitter data

Item	***N***	**(%)**
Total Tweets	740,356	100.00
URL in retweets	564,309	76.22
URL in root tweets	39,219	5.30
URL in replies	15,642	2.11
Total Twitter users	411,361	100.00
Verified Users	9,465	2.30
Average user's tweets*	34,009	
Average user's followers*	5,055	
Average user's following*	1,584	

**Average values are without percentage*

There are some missing data considerations in this analysis. Due to the data filtering method, which matches any tweets containing the full URLs of the original article, tweets that refer to the same article with different URLs, such as URLs with referral or tracking codes (i.e. theguardian.com/../?smid=tw-share), or other shortened URLs (i.e. nyt.ms/34/233, bit.ly, etc.) were difficult to capture. This is due to how the URL filter for the query in the Twitter API relies on the exact match of the expanded URL and is unable to find all other URLs that direct to the same article.

Some of these URLs were captured from the re-tweet data with full URLs, which refer to original tweets with shortened articles. A further 14 shortened links of articles were captured in the data collection and filtered about 20,000 additional tweets that were not captured before. Up to the data collection period, some tweets have already been deleted and some accounts have been suspended. Thus, these missing data from protected accounts, suspended accounts, and deleted tweets, led to incomplete re-tweet data in this research. For example, if a tweet is known to be re-tweeted 300 times, the actual data collected may only show 284 tweets RT-ing it, which produces a partial network. Although there are some missing data, the data is representative of the majority of tweets and RTs linking climate articles, which can be used to illustrate the underlying network of information diffusion. The Twitter data produced an inferred diffusion network, where the reproducibility and spread of information can be analysed from the increase in information spread according to time.

3.1.3. Botometer Data

The third dataset uses the Botometer API[7] to get the scores for the bot prediction of each user, rated from 0 (humans) to 1 (bot) (Davis et al., 2016; Sayyadiharikandeh et al., 2020; Yang et al., 2019). Botometer is a supervised machine learning tool that learns the characteristics of likely automated and human accounts based on annotated datasets and

[7] https://botometer.osome.iu.edu

account features. This dataset used the V4 Botometer, which implements an ensemble of classifiers that are specifically trained for various bot classes.

Here, I used the Botometer to calculate the bot scores of the most active users in the dataset, which are those with degree centrality in the top 5 percent (95 percentile) of the degree distribution to distinguish the influence of humans and bots in information spread. From the V4 Botometer, each account is rated by its complete automation probability (CAP) scores, which is the cumulative conditional probability that accounts with a score equal or greater than the CAP score is automated. For example, if the raw score of an account is 0.78 and the CAP score is 0.70, then it means that 70% of accounts with this score is labelled as automated account. The Botometer data is utilised to understand the influence of automated accounts in amplifying and triggering the spread of credible and misleading climate information.

Many studies have found difficulties in classifying bots using Botometer. In particular, the false-positive issue of humans being classified as bots (Rauchfleisch & Kaiser, 2020). With cross-validation of other bot detection methods, the bot population on Twitter is estimated to range between 9 and 15 percent (Varol et al., 2017). To prevent misclassification error, similar to previous studies (Pozzana & Ferrara, n.d.; Varol et al., 2017), I classified accounts as bots for those with a bot score ranking in the top 5 percent of the distribution, which corresponds to a threshold value of 0.82 in this dataset. Varol et al. (2017) found no significant variation for those with scores 0.40 to 0.60, which is also supported in this dataset, where the CAP values of users with bot scores between 0.40 and 0.60 are similar, as shown in Figure 3. The frequency distribution of bot scores for accounts with the highest degrees is illustrated in Figure 3. Thus, to reduce the risk of misclassification, the maximum threshold value for humans in this dataset is set at 0.40. From this labelling, 658 (66.5%) accounts were labelled human, 56 (5.6%) bots, and 279 left unlabelled (those with a score value between 0.40 and 0.82).

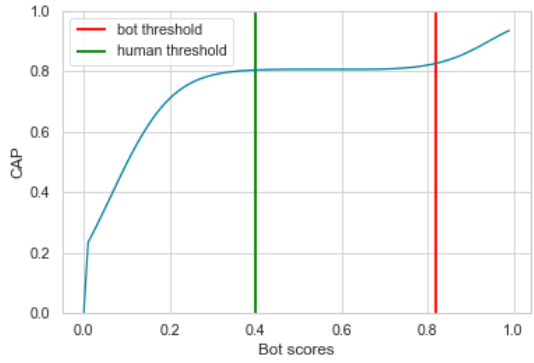


Figure 3: Bot and CAP scores showing the likelihood of accounts being automated (CAP) for each bot score

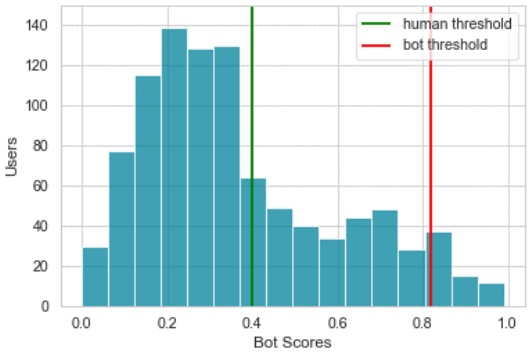


Figure 4: Distribution of bot scores for the top five percent node centrality

3.2. Data pre-processing

For data collection and pre-processing, I used Python and *BeautifulSoup* to aggregate and clean XML and JSON files. All network analysis and visualization were made with *NetworkX* and *igraph* for Python[8]. Statistical analysis was done in Python with *statsmodels*[9].

While the aggregated Twitter data show all tweets from various languages, only English tweets were included in the analysis. To construct the diffusion network, I used the Twitter dataset to gather the data for the article cascade networks. For example, when article A was

[8] https://igraph.org/python/
[9] https://www.statsmodels.org/stable/ind

published, the people who first tweeted about the article were the root users, and the users who re-tweets a tweet about the article becomes the child nodes/users. Each re-tweet data is considered as one cascade or diffusion event. From all the successful diffusion events, the networks are reconstructed based on time and tweet attributes. Each re-tweet is time-stamped; thus, a time-inferred diffusion network was built. Many of the diffusion events were only RTed once or not re-tweeted, this was categorised as unsuccessful diffusion. Thus, the analysis below was constructed based on diffusion of more than 1 RTs.

The Twitter data only shows the parent node and the user who RT-ed the original tweet. Consequently, this data does not show the entire diffusion process. Instead, it forms a time-inferred diffusion, where the timestamps indicate the hierarchical process of diffusion. Thus, with this limitation, the inferred diffusion trees show an upper bound of the actual diffusion taking place of the underlying network.

3.3. Network set-up

3.3.1. Cascade network ($G_{[URL]}$)

There are two main networks used in this analysis. First, I use article-level cascade tree networks, denoted as $G_{[URL]} = (V, E_{[RT]})$, where the nodes represent all the users who post a tweet linking an article, whether through an original tweet or RT, and the edges represent a re-tweet. Users who post an original tweet that was never RTed will be shown as isolates, and those with one RT will be dyads. This network shows the structural properties of the information diffusion and was analysed with the temporal properties to understand how different types of information (on the article level) spread in an online social network.

To analyse the diffusion from the information content, the article-level network structure is explored. Using structural properties, including the size of the largest weakly connected component, the number of connected components, the local and global clustering coefficient, and the network density. These, compared with the cumulative distribution of tweets plotted against time showcase the speed of diffusion (see section 3.4.1).

Depending on the availability of the data to the researcher, cascade networks often have a tree or forest network topology. Two main cascade features that are investigated are structural and temporal properties. The structural properties inform the relation between

users and the paths of influence between users in the cascade. The temporal properties measure the diffusion events that occur at a given time, it adds value to the structural aspect of the network analysis (Alrajebah et al., 2017).

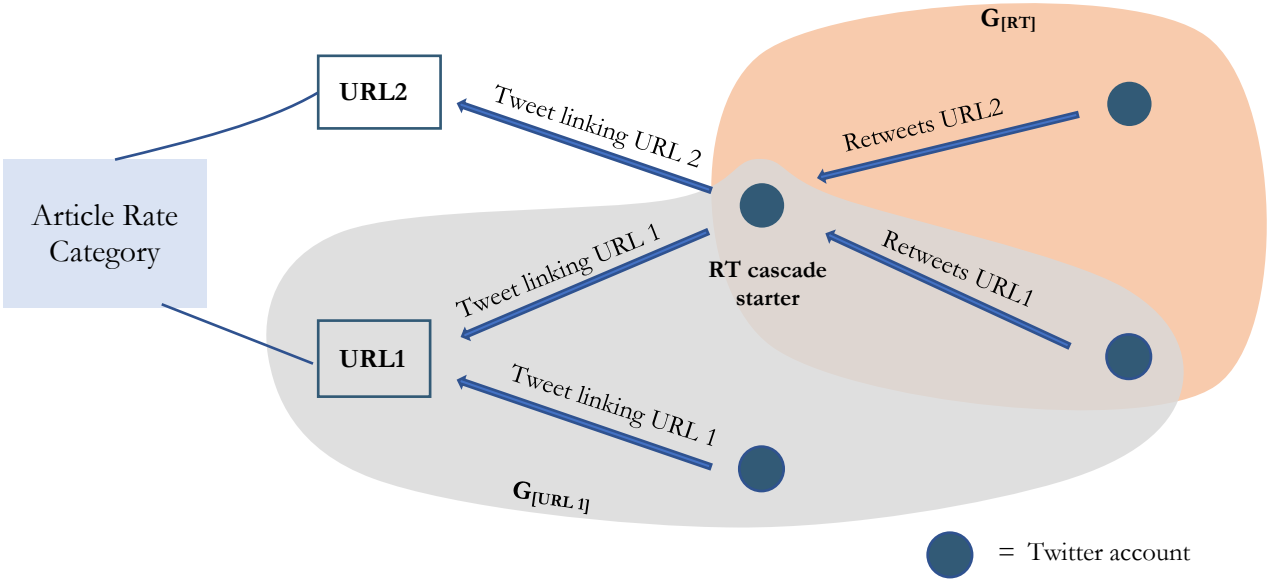


Figure 5: Illustration of how the two network was constructed from the data collection

3.3.2. Retweet network (G_{RT})

The second network is a whole RT network of all the users and articles including every cascade from every time points in the data collected, denoted as $G_{[RT]} = (V, E_{RT, T})$, where each vertex represents a user, and the edges represent RTs of URLs. The whole network contains 411,740 nodes and 503,899 edges. The data forms a temporal directed network of *T* timesteps where *T* is defined as the duration between the initial tweet and the re-tweet in minutes. In this network, isolates (those whose tweets were never RT-ed) are not included. This network presents the structural properties of user influence through re-tweet relationships, which highlights the influential nodes and the cohesiveness of the spread of credible and misleading information networks.

Using k-core decomposition (explained in section 3.3.4), I analysed the whole network of the aggregated user's engagement with climate information on Twitter to identify influential users and the type of information shared by them. Furthermore, through the whole network,

an analysis of user-level engagement was explored. User engagement and activity showed how many tweets each user has in the sample population, the percentage of cascade starters, and the number of root tweets each user has. This, with the types of information people engaged with, allows the analysis of homophily of information sharing through calculating the likelihood of engaging with information from different credibility ratings (see section 3.4.3).

3.4. Network Analysis

3.4.1. Temporal properties

There are a few temporal features that were used to understand the growth of cascades and the collective attention of information. First is the duration between RTs and original tweets, which indicates how many times a message has been passed since its initial post and shows the fade of interest in the message over time (Dow et al., 2013). Second, is the time lag between two sharing events; posting and first reshare. Third, is the speed of diffusion, which indicates when the first cascade occurs.

The tweet volume and user reach across time indicates the number of spikes and peaks in the cascade networks. These are presented through the lifetime hours of articles, which measures the temporal distance between the first and last tweet about an article. Further, with the complementary cumulative distribution function (CCDF) other properties that are examined include cascade growth and cascade re-occurrence. (Alrajebah et al., 2017; Cheng et al., 2016).

3.4.2. K-shell decomposition

K-core decomposition works by pruning nodes with degrees less than k, which uncovers the most active or influential users in climate change topics. To measure the diffusion network structure, I utilised k-shell decomposition to measure the cohesion and clustering of the core network. K-shell decomposition has been widely used to detect the core structure of a network (Seidman, 1983). More recently, k-shell decomposition has been utilised to identify the most efficient and influential spreaders in a complex network (Kitsak et al., 2010). Many other past studies have found the effectiveness of using k-shell to rank user's influence, which outperforms other ranking schemes such as PageRank and degree (Pei et al., 2014).

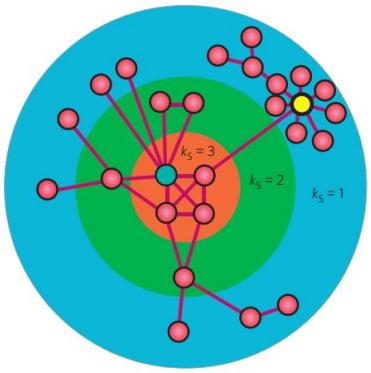


Figure 6: Illustration of K-core decomposition revealing central nodes (Kitsak et al., 2010)

This method is suitable for detecting influential spreaders and their influence in information diffusion, where individuals with high connectivity stand to contribute more to the diffusion process. In a modified k-shell decomposition, the algorithm is used to identify influential specifically on Twitter data (Brown & Feng, 2011). In k-shell decomposition, nodes with an increasing number of degrees are iteratively removed. First, all nodes with degree k=1 (in and out-degree) are removed; pruning continues until there is no node with k=1. The removed nodes are categorised to be in k shell with $k_s = 1$. In the next shell, a similar method is utilised until the shell reaches the largest k shell, which is the k-core, wherein in this network, the core is k=15. Depending on the modularity of the network, in real-world networks, the degree of the node may not always determine the location in the k-shell structure.

The core of the k-shell network highlights the most central and influential nodes. In each k-core network, its structural properties show the network cohesiveness, which measures how tightly knitted the network structure is. This is measured by the density, clustering coefficient, and diameter of the graph. The subclusters of the graph show the different communities of It is the ratio between the number of links in a graph and the maximum number of possible links. The proportion of re-tweets and users, and the number of users in k-core decomposition (k=1 to k=15) is attached in Appendix B.

3.4.3. Edge homophily measures

This section will investigate the likelihood of users sharing information with the same level of credibility. It is measured by calculating the edge homophily in cascades through conditional probability, adopted from a similar previous study (Anderson et al., 2015). For all accounts who have engaged with the sampled tweets more than once (N = 112,312), this calculates the conditional probability of switching information. In other words, this measures how likely is a user to post low-credible content given that they have posted highly credible content before and vice versa.

The conditional probability is measured as:

$$P(Cat_{II} \mid Cat_I) = \frac{P(Cat_{II} \cap Cat_I)}{P\,(Cat_I)}$$

$P(Cat_{II} \cap Cat_I)$ is measured by getting all the pairs of rate categories a user has engaged in. For example, if a user has engaged in 3 categories (high, low, neutral), they will be categorised as (high, low), (low, neutral), and (high, neutral), which produces an adjacency matrix of the number of users in each category pairings.

$P(Cat_I)$ is measured by the proportion of users (who have engaged with more than one post) that have engaged with the category.

3.5. Limitations and Future work

There are a few limitations concerning the data availability that limited the network analysis. Due to the limitations of the Twitter API, I was unable to retrieve the complete information paths of RT diffusion. The edges that form the RT network are between the original tweet author and the RTer, without data of how a user was exposed to the information. In many past studies, the real diffusion paths were reconstructed with the underlying follower network. However, this was not possible in this research due to the current Twitter API restriction to retrieve all the follower data of 411,361 users, which sums to 2,118,371,310 followers, where follower data collection is restricted to 60 user requests or at most 60,000 follower entries (whichever is smaller) per hour. Thus, this restricted the analysis of the complete cascade tree structure analysis, such as depth, breadth, and hops.

Secondly, many of the URLs used to share articles on Twitter are shortened URLs, such as bit.ly, or shortened from other platforms (i.e. lnkd.in/XX). Thus, in the temporal analysis, there are some sudden bursts of activity long after the article was published, up to six years, where is it difficult to conclude whether any other activities were mentioning the articles between the period.

3.6. Summary

This chapter discussed and explained considerations in the data collection and methodology in social network analysis that suits the research questions of this research. Steps were taken to ensure robust data collection and minimise the limitations of data collection methods to combat missing data. Moreover, data pre-processing and judgements for threshold values were explained.

4. Findings and Analysis

This chapter explains the findings and analysis of this research. Through findings from the network analysis and annotated article data, this chapter begins with summary statistics of the data gathered, then continues to the analytical discussions of the information spread from two perspectives: article (article cascade network) and user-level (RT cascade network).

4.1. Summary Statistics

Information cascade networks trace the speed and breadth of information diffusion in a social network. In each article diffusion cascade, a re-tweet network pointing to a single root tweet with an article link forms a star topology network. An article cascade network that is formed by multiple RT cascades and isolates, shows the cascade size, which is the number of people who have ever engaged with the article at least once, whether through original tweets or RTs. Figure 7 shows the distribution of article cascade sizes from all the 152 articles. The cascade size varies from 3 to 55,755 and it increases exponentially, with a large proportion of the article diffusing to more than 1000 people (median = 1051).

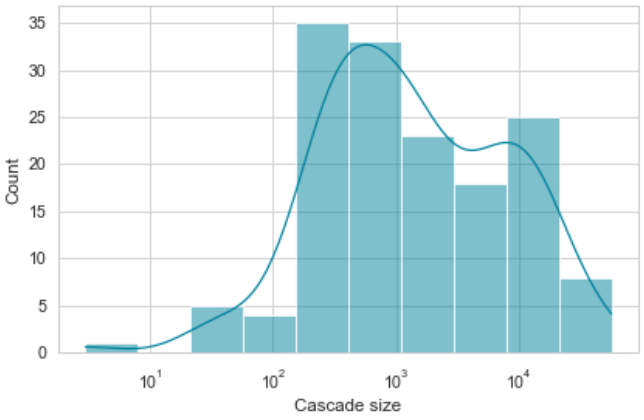


Figure 7: Distribution of Cascade Sizes

Each article has a continuous rating value, which reflects the consensus of the scientific feedback from various reviewers. The continuous rating distribution is categorised into five broad credibility categories, which are, very low ([1,2)), low ([2,3)) , neutral (3), high ([3,4)) and very high ([4,5)).

4.1.1. Article cascade lifetime and cascade size

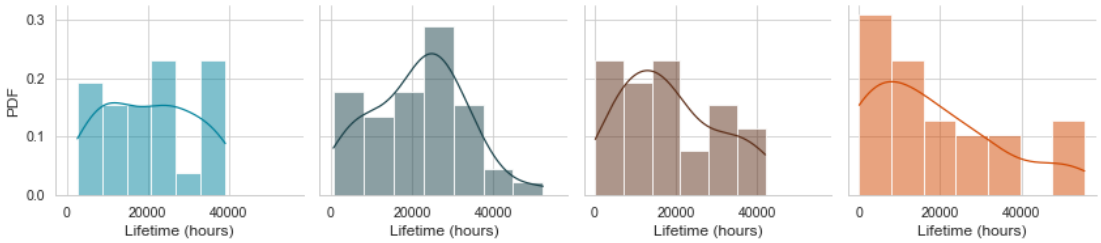


Figure 8: Probability density function (PDF) of cascade lifetime hours

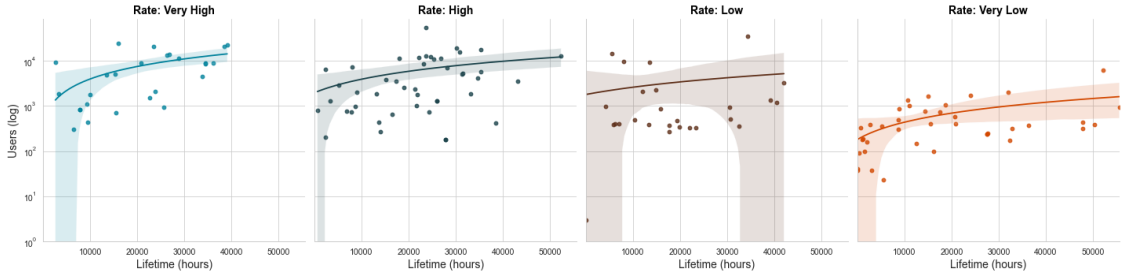


Figure 9: Lifetime hours and user reached on each rating level

The attention patterns of each article are measured by the lifetime hours, which is the temporal distance between the first and last tweet of an article (in hours). Figure 8 illustrates the probability density function of lifetime hours of articles in high category (rate > 3), neutral (rate = 3), and low (rate < 3). Figure 9 then compares the total number of users reached and the posts' lifetime hours. This highlights that higher-rated articles reach more people in a shorter period, as shown by the sharp positive correlation between lifetime hours and user reach. Less credible content has been shown to have longer lifetime hours on the platform, however, this alone does not confirm higher levels of engagement. Many of the articles stay on the platform longer, with the longest duration between the time of the first post about an article and the last being six years. This confirms that to an extent, information sharing behaviour is content-driven, and there is a positive correlation between the cascade size and lifetime hours.

The diffusion of most RTs end after 13 hours, and people stop engaging with the articles four days after the article was published. Some article cascades with longer lifetime hours are affected by accounts that RT their old tweets to amplify the engagement, which is illustrated by self-loops in the network visualisation. Collectively, there are 1003 users who self RT,

with 1683 self RTs found. This partially explains the variety of lifetime hours and the number of unique users reached. Time is a crucial factor in information diffusion as older tweets will be less noticeable in people's timelines unless people seek specific tweet content. Therefore, when there are more engagements, the article spreads faster to a larger audience.

4.1.2. RT cascades

A successful diffusion is constructed from an RT cascade from any tweet that has been RT-ed more than once. The relationship between the scientific credibility of articles and the number of RT cascades found is illustrated in Figure 10, where a positive trend between the content credibility and the number of diffusion cascades is demonstrated. Note that the y-axis is on a log scale, so the number of cascades found increases exponentially as the article rate increase. Although, there are some low-rated articles with a similar number of cascades with highly-rated articles. This implies that although information sharing is content-driven and that extremely misleading climate articles do not diffuse as widely, there are some similarities in the diffusion characteristics of low and high-rated articles, which will be discussed in the following sections.

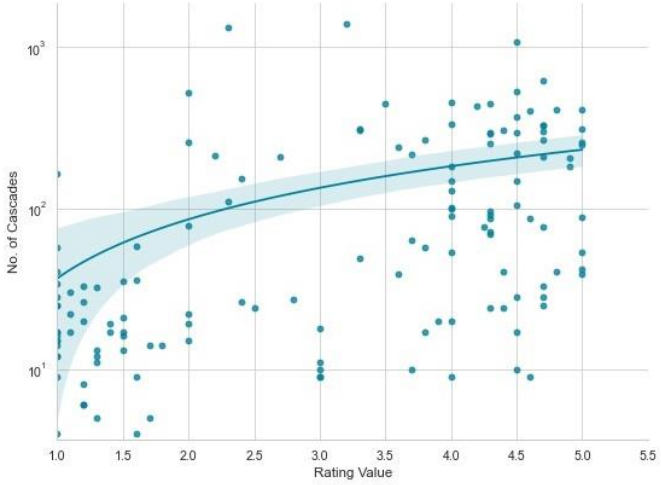


Figure 10: Number of RT cascades in each article rating value

The measure of information diffusion on Twitter can be analysed based on the reachability of nodes and the speed of information spread within a specific timeframe. As shown, some articles with low rates such as those rated 2.0 may have the same number of cascades as high rated articles (i.e. rate 4.5), indicating other factors affecting the diffusion.

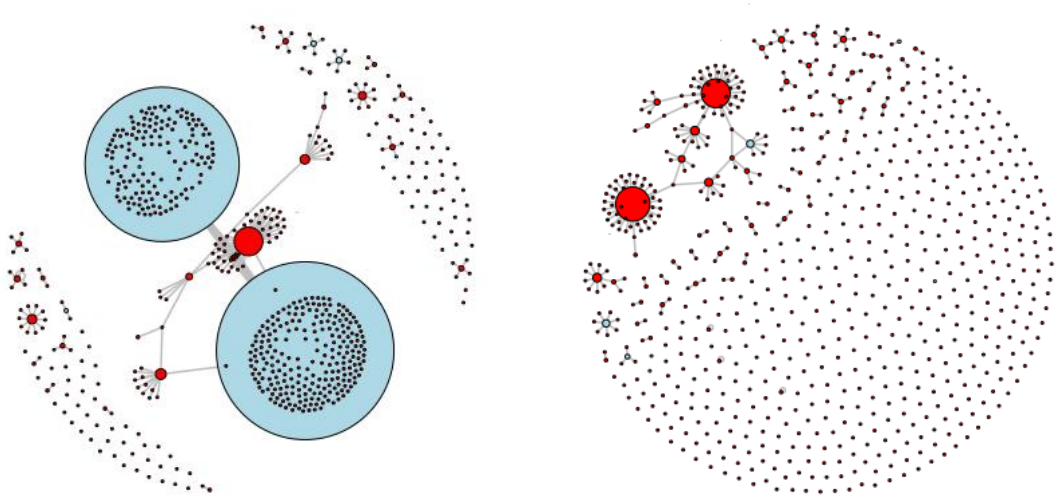
4.2. Information Cascades Network

The diffusion of articles on Twitter is explored from the article information cascade networks ($G_{[URL]}$). The following section explores the structural and temporal properties of article cascade networks concerning information credibility.

As a form of social media, to a certain extent, the outcome of information propagation is affected by user interactions and influence. The aggregated outcome of an article's reach is impacted by who initially tweets them, at what time, and whether more than one 'power user' tweets them around the same time. This led to differences in speed (time) and breadth (user reach) of diffusion, where some articles spread more evenly across time and others spread to most people in a shorter period. The interactions between users, their information-sharing behaviour, and influential users in information propagation will be further discussed in section 4.4, however, this section discusses how the article-level network structure affects the speed of diffusion cascades.

Diffusion networks that consist of multiple powerful root users create a broadcast-like diffusion structure where the large proportion of people in the network are RT-ers. On the contrary, some diffusion mimics a word-of-mouth structure, where a large proportion of people directly share links to the articles in original tweets. Articles that reach more users and spread faster tend to have multiple influential root users that tweet around the same time, which collectively results in a wider and faster spread. Figure 11 shows examples of two article cascades of comparable size from two different categories, a high-rated article from USA Today (Figure 11a) and a low-rated article cascade network from Forbes (Figure 11b).

In Figure 11a, most nodes are isolates, or dyads, which show that people are either posting original tweets, or the root tweets are only RT-ed once. It illustrates the structure of word-of-mouth-like diffusion, where people who are sharing the information are not directly re-tweeting someone, thus opinion formation is more segregated. Furthermore, this kind of diffusion structure implies that many of the tweets are referring to an article that has been posted a long time before, sometimes even articles that have been published more than a year ago. This also shows the untraceable diffusion paths on Twitter, as it is unknown from the data where the users heard the information from, especially those who share contents



(a) USA Today article, rate: high, nodes: 775

(b) Forbes article, rate: very low, nodes: 922

Figure 11: Article-level network structure
[each node represents a user (blue: verified, red: not verified),
node size: out-degree (retweet count),
and edge represent a retweet]

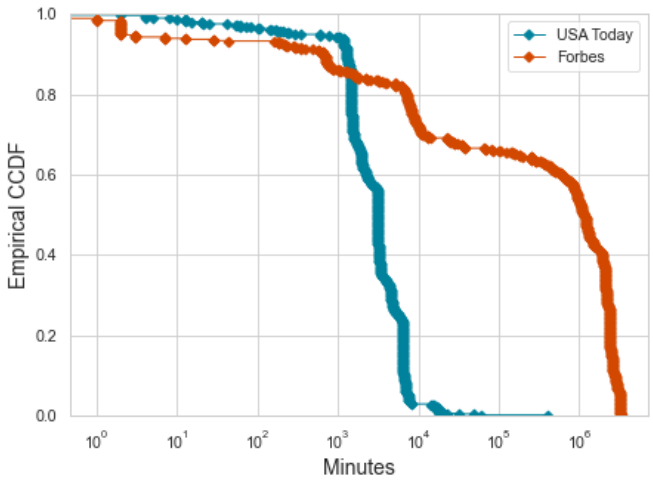


Figure 12: Empirical CCDF plot of tweets based on duration (in minutes) since first Tweet

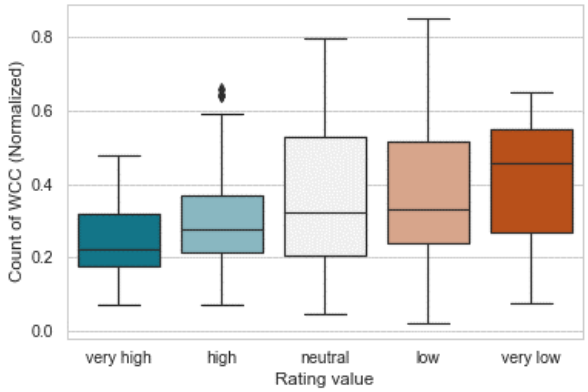
significantly later. In comparable cascade sizes, credible information spreads faster and more or less at the same time, while less credible content spreads slower.

The empirical complementary cumulative distribution function (CCDF) graph in Figure 12 illustrates how fast the articles in Figures 11a and 11b diffuses. It measures the temporal distance between each tweet to the first tweet posted about an article, measured in minutes, where the x-axis is on a log scale. In these two examples, the network structure shows the distinction between clustered and less connected components and how it affects the collective speed of information spread. Although the two cascades reach a similar number of users, the time required to reach a marginal user in the cascade differs. Furthermore, the presence of influential power root users, as shown by the large nodes, show an effective diffusion that reaches the most people in less time.

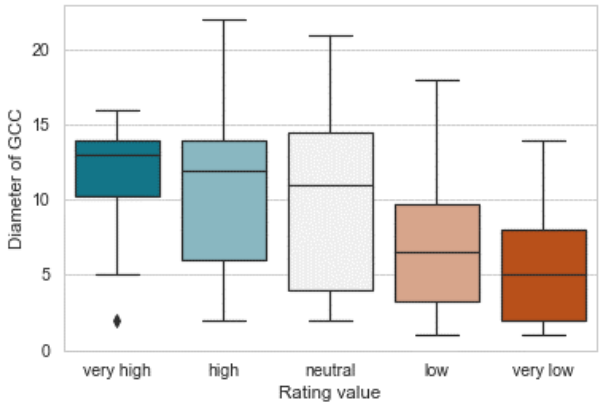
4.2.1. Structural properties

As per Figure 11, the structural properties of article cascade networks demonstrate the differences in cascading behaviour of credible and misleading information diffusion. Thus, this section explores the structural properties of all article cascades in relation to the article's credibility. Figure 13a shows a boxplot of the number of weakly connected components (WCC) normalized by the cascade, plotted against the article rate category. This highlights how the diffusion network of higher-rated articles is less clustered and more connected, while the diffusion of lower-rated articles consists of more dyads and isolates, which leads to a higher proportion of WCC. This indicates that the diffusion consists of more original tweets from individual accounts rather than RTs of a few influential accounts.

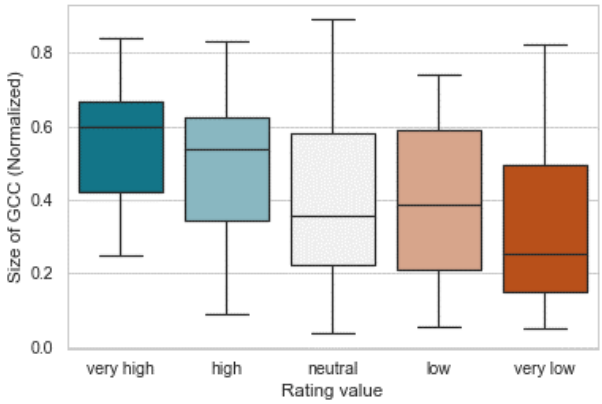
Figure 13b illustrates the diameter of the giant connected component (GCC) in the article cascade, and Figure 13c the size of the GCC, normalised by the cascade size. The decreasing trend between both the size and diameter of GCC and the article rating category demonstrates how the diffusion networks of higher-rated articles tend to be more connected and cohesive, as shown in Figure 11a. This article cascade network is made of aggregated original tweets and RTs of the same article (see Figure 5), where an edge illustrates the link between a root user and a user who RT (without true diffusion paths). Thus, the highest number of hops in the article cascade is supposedly one. However, some nodes engage with the same article more than once, connecting multiple RT cascades. Therefore, the diameter



(a) Number of weakly connected components (WCC) in each article cascade network



(b) Diameter of giant connected component (GCC) normalized by cascade size



(c) Size of giant connected component (GCC) normalized by cascade size

Figure 13: Structural properties of article cascade networks

of the GCC indicates the number of accounts that have multiple exposures and those that RT the same information from different people, thereby linking different RT cascades. The role of influential accounts that have significantly higher degrees than other nodes affects the connectedness and cohesiveness of the cascade networks. To an extent, this demonstrates the structural virality of the information cascade, where more clustered networks (such as Figure 11b) mimic word-of-mouth spread, and the more connected and cohesive networks (such as Figure 11a) mimic a broadcast-like spread, which leads to opinion leaders of a few influential accounts.

4.2.2. Temporal properties

The structural properties of the article cascade networks affect the speed of information diffusion. Thus, the temporal properties of the cascade networks complement the structural properties to understand how structural properties affect the characteristics of information diffusion. Similar to the lifetime hours, the temporal distance is measured by the duration (in minutes) between the first and last tweet about an article, which shows how fast information spreads on Twitter.

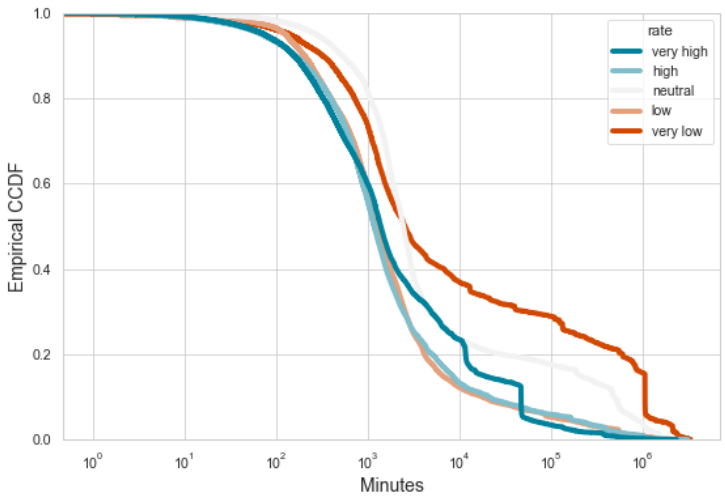


Figure 14: Empirical CCDF plot of Tweets on time since first post

Using the empirical complementary cumulative distribution function (CCDF), Figure 14 illustrates the collective attention patterns of different information rate categories. Information in the low-rate categories diffuses longer than in high-rate categories. However, the attention patterns of high and low categories are very similar. Using the Kolmogorov-

Smirnov test, the maximum difference between the high and low category empirical distribution is found to be the lowest amongst other comparisons ($D = 0.039, p \sim 0.00$), which indicates that high category information is as likely to spread as fast as low category information based on the empirical CCDF plot. Much of the low category that spreads as fast and as widely as high category articles are often articles with clickbait headlines, articles from the same media company, or controversial content.

Table 2 – Difference of distribution between high and low category

Duration (hours)	N	Mean	SD	Min	Max
High	4275.0	534.0	2440.0	0.0	52298.0
Low	1347.0	548.0	2722.0	0.0	42006.0

4.2.3. Role of Media

The similarity of diffusion characteristics of high and low rated content stems from the articles of the same media company but varies on the credibility. The role of media's reputation and influence ultimately affects the outcome of the spread of information. Naturally, traditional outlets are more well-known and thus, will have significantly more traffic, regardless of the information content. Traditional outlets such as *BBC* and *The Guardian* naturally have more traffic than curated digital media channels such *as Breitbart, Vox, or Mashable*. This is supported by empirical evidence, where some highly credible articles fail to diffuse to a wide audience and have short lifetime hours on the platform. The CCDF plots show how articles from the same media show similar diffusion characteristics even though the credibility of the content varies (Appendix C). For example, both articles (very high and low rated) from the Independent have a similar distribution of empirical CCDF of tweets. This highlights how traditional media has a higher influence on information dissemination, especially news channels, than curated media platforms.

Although media channels and article types affect the spread of information online, media and content alone do not conclude the outcome of the spread of information on Twitter. Influential users who expedite the speed of information spread and widen the reach become a significant factor.

4.3. Summary of user engagement

Due to the large volume of re-tweets in the data, understanding how opinion leadership is formed in Twitter is important to identify how the influence dynamics dictate the course of information diffusion. Among all the users, only 5.81 percent are the cascade starters (has at least one root tweet), and 76.30 percent are only re-tweeters. The proportion of users who are RT cascade starters that conclude as a successful diffusion (RT-ed more than once) is 3 percent. As shown in Figure 15, the partition of tweet count per user follows a power law, which indicates that most users only engage with the climate articles once. About 100,000 accounts have only posted climate articles once, where very few accounts have more than 100 tweets about the articles.

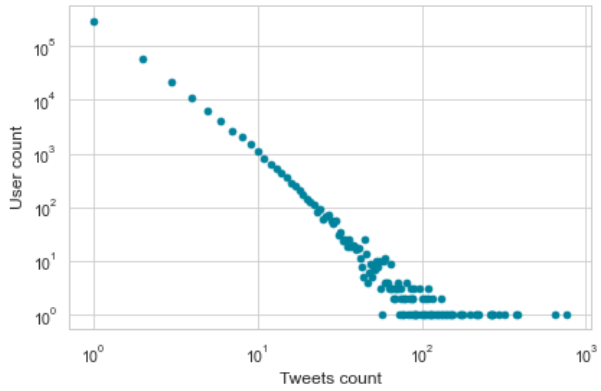


Figure 15: Tweet count per user

Similarly, most cascade starters have one root tweet, with a very small number of users who have multiple root tweets. Cascade starters were heavy-tailed with only a few people establishing more than one root tweet. For example, one user with the highest number of root tweets is *@FriendsofScience*, with 148 tweets as an RT cascade starter that links to articles selected in the sample. This is an account of a group of earth, atmospheric, solar scientists who believe that the sun is the main driver of climate change. However, the number of root tweets alone does not completely show the level of influence in information diffusion.

The level of influence is impacted by the cascade size of each root tweet. Figure 15 illustrates the degree distribution of nodes in the RT cascade, which indicates the aggregate RT cascade size from all root tweets per user. This confirms the skewness of the population of Twitter, where very few people are very active on Twitter, and among them, even fewer people have the highest degrees, who are responsible for a large proportion of information dissemination.

This highlights how the online information dynamics are skewed and how opinion leaders are formed.

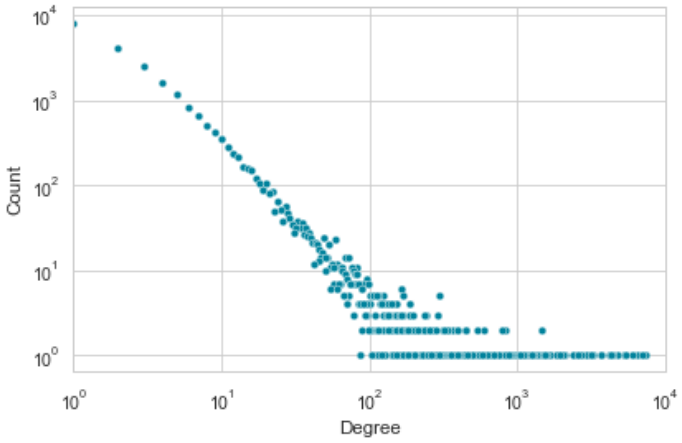


Figure 16: Degree distribution of RT cascade

4.4. Re-tweet Network Topology

The whole network shows all the users who have engaged with the set of articles, both through re-tweet and original tweet, and highlights the information flow from the network structure. In this network, users are connected through re-tweet relationships on Twitter, where the directed edge between two users is formed whenever a user's tweet is retweeted by another user (RT-ed forms out-degree). Root users and RT-ers are characterised from the node's degree, where users with zero out-degree and high in-degree signify that they are very active re-tweeters, and conversely, users with very high out-degree demonstrate that they are very influential root users.

To visualise the re-tweet network, I performed k-core decomposition on the directed network of users to discover influential users. The retweet network visualises the spread of information between groups of people and shows the underlying network structure of credible and misleading climate information. Figure 17 shows a pruned network of all nodes with degrees greater than five (k=5). The nodes represent users, and the edge colours represent the scientific credibility rating of the article link in a tweet, where blue is high and very high, orange is low, red is very low, and white is neutral, debated, or inconclusive.

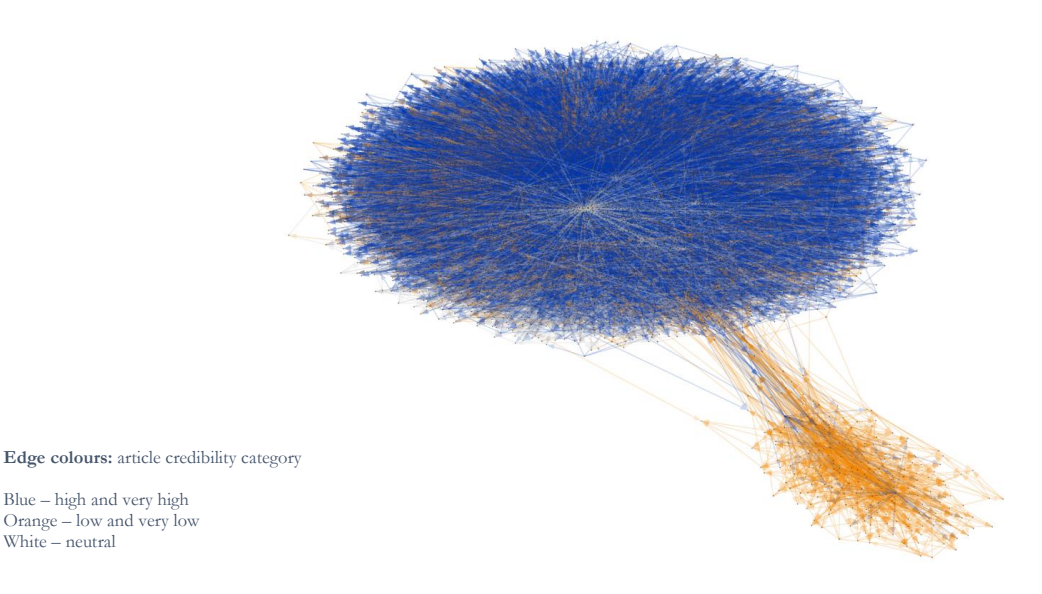


Figure 17: Pruned whole retweet network
(K=5, V=13,399, E=94,271),

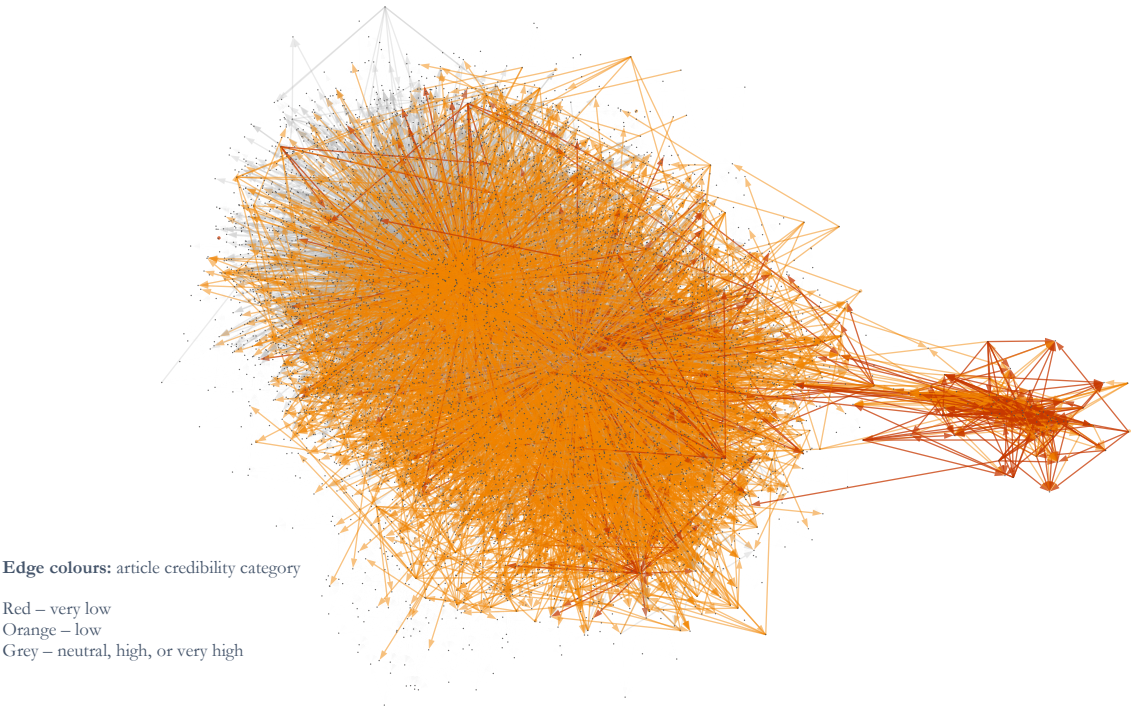


Figure 18: Misinformation network in the pruned whole retweet network (*k*=5)

Overall, it illustrates the polarisation between credible and non-credible climate information sharing on Twitter. The smaller cluster mostly contains articles of extremely low-rated articles from sites such as *Breitbart* and shows evidence of homophily and echo chamber effects. This is illustrated in Figure 18, where red edges represent very low-rated articles and orange edges represent low-rated articles. To an extent, this confirms the hypothesis of evidence of polarisation of science. However, not to a great extent, as there are some overlaps of information sharing of high and low-rated articles in the bigger cluster. This indicates that many people who engage in sharing climate information share both credible and less credible content. In particular, low-rated articles often make claims from scientific findings that are taken out of context, where implications are exaggerated or new research findings that have not reached scientific consensus, and further research is required to make claims. This results in many clickbait headlines, which show controversial contents that cite scientists out of context.

4.5. User Influence

Influential users can be discovered through centrality measures such as k-shell decomposition and degree centrality. A few users that influence a large proportion of the audience are discovered from the k-core network.

Figure 18 shows the core-shell of the network, with k=15, showing the most active and influential accounts in climate change information dissemination in the past six years. The figure shows the most active root users only by filtering nodes with the out-degree of at least one. The node size represents the out-degree, thereby showing the most influential accounts in climate topics. It includes a group of climate activists; climate authors and journalists; media accounts; scientists; and public figures. Surprisingly, many of these influencers also share less credible climate information, including the young climate activist, Greta Thunberg (Appendix D).

These central nodes exhibit the spread of both credible and non-credible information. However, when we unpack the content of the tweets with these articles, there are two types gathered. Attacks of non-credible articles; and exaggerations of scientific findings that have not reached consensus in the climate science community. This evidence supports the literature around the alarmism language that is found in many climate activists, which is not

scientifically supported. The qualitative exploration of tweets of misleading content by influential accounts continues in section 4.8.

The core-shell network forms one big cluster where there are many overlaps between misleading and credible information sharing. This implies the weak polarisation of science in climate change information dissemination, whereby a clear distinction between communities of science-based and non-science-based information dissemination networks is not present.

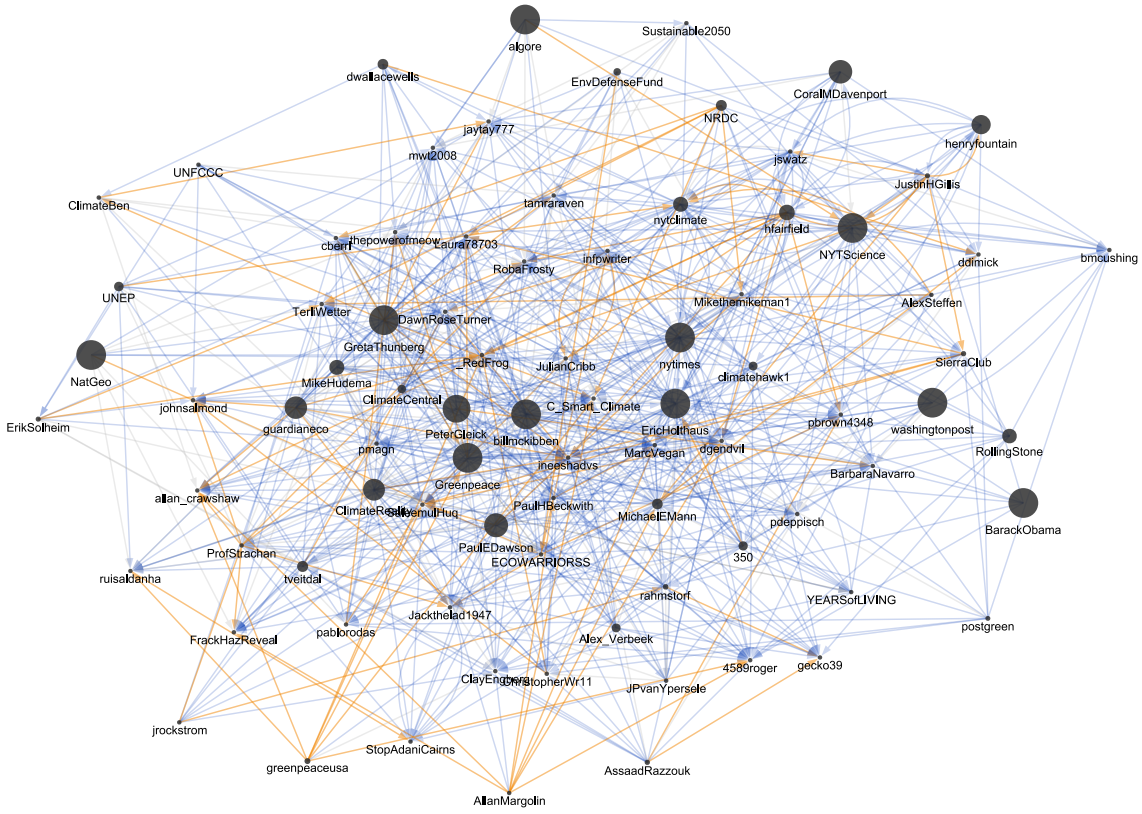


Figure 19: K core network (K=15)

When the k-core graph is further partitioned, Figure 20 shows the accounts in the core network who engage with both high and low content categories, while Figure 21 shows the scientifically credible influential accounts who have only posted highly credible content. The very few accounts that share only credible climate tend to be accounts owned by an official government-related institution such as the United Nations Framework Convention on Climate Change (UNFCCC), National Resource Defense Council (NRDC), United Nations

Environment Programme (UNEP), and Environmental Defence Fund. A few publishing companies include *National Geographic, the New York Times (@nytimes*), and *the Washington Post.* Other accounts are public figures such as Barack Obama and climate journalists, where a cluster of the network is made of writers for *the New York Times*, such as *@CoralMDavenport, @JustinHGills, @jswatz, @henryfountain.*

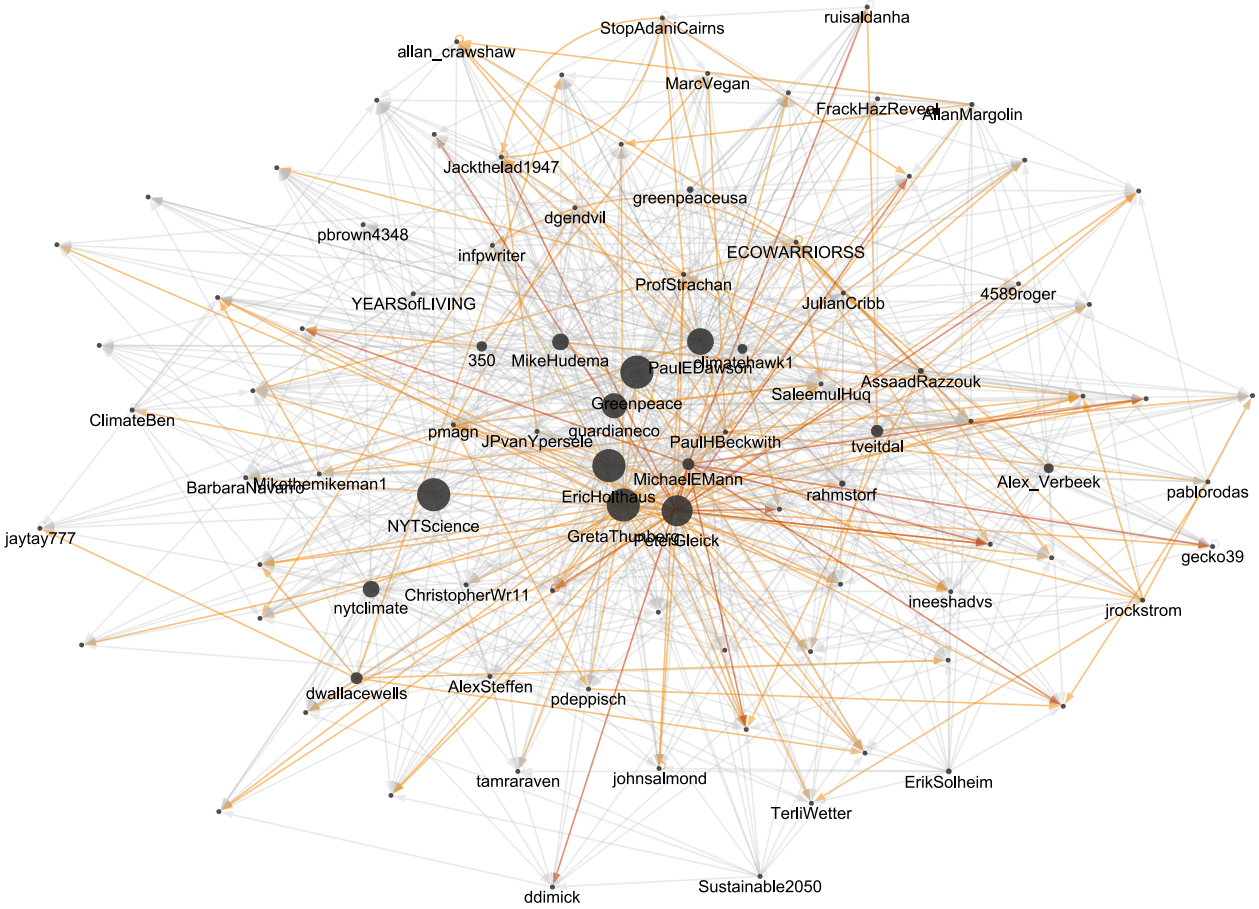


Figure 20: Influential nodes in K core network who shares both misleading and scientifically credible climate content

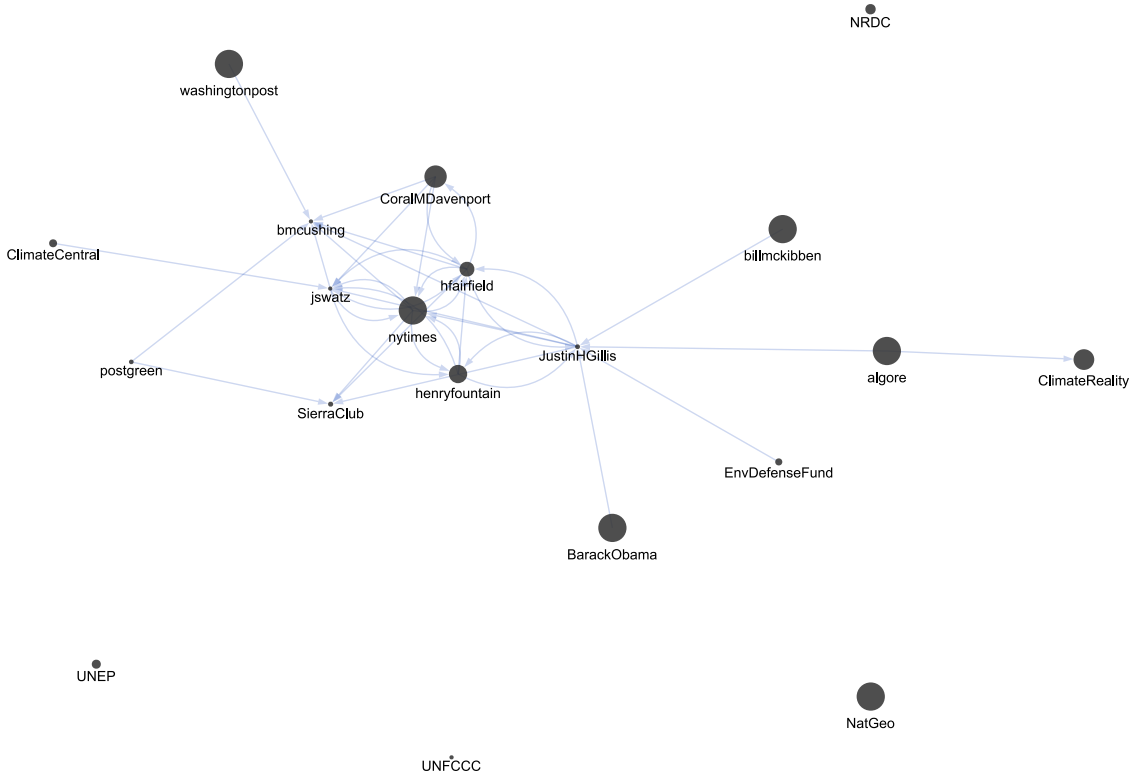


Figure 21: Scientifically credible influencers

4.6. Edge homophily

As illustrated in Figure 19, the core-shell of the retweet network highlights the mix of sharing scientifically credible and non-credible articles. Only a small proportion of users who only post scientifically credible content in the past six years from this observed dataset. Following the immense overlap between misleading and credible climate information in sections 4.4 and 4.5, I explore information mixing patterns to understand the homophily of information sharing from a scientific credibility perspective.

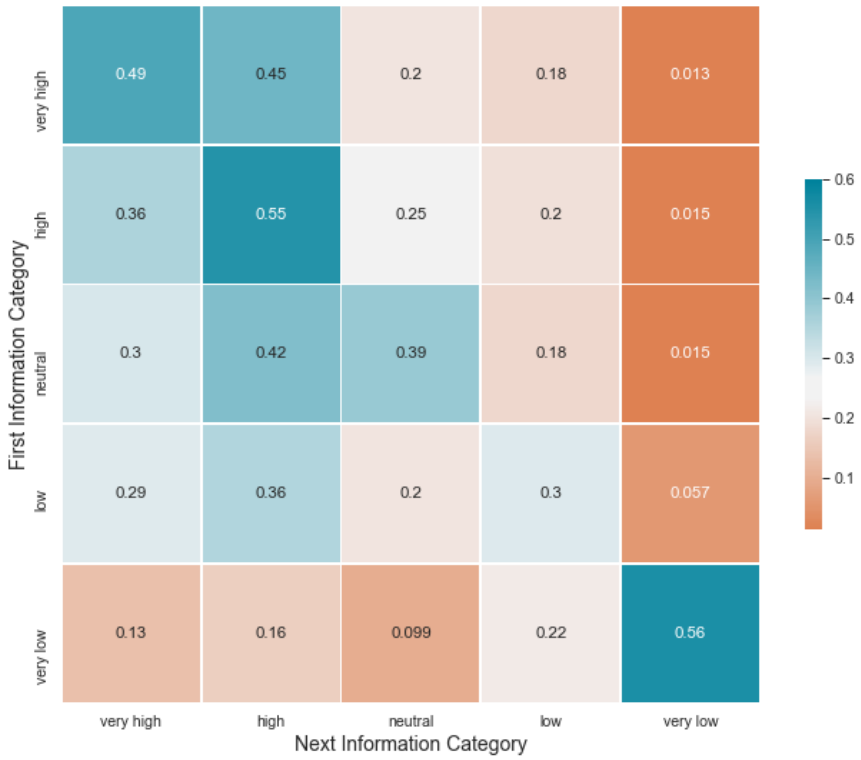


Figure 22: Edge homophily – the probability of sharing multiple articles from different information rate categories

Using the conditional probability of users who engage with climate posts multiple times (see section 3.4.3), Figure 22 highlights the likelihood of users engaging multiple climate change content with different scientific credibility. This shows whether users subsequently post the

same kinds of information with the prior tweets and the likelihood of people cross-sharing between high and low-rated articles.

Generally, the distinction between very low-rated climate articles and others is clear. This is shown by the probability for users to continue posting very low-rated information after previously posting very low-rated content (P = 0.56) and the low likelihood for them to engage with other higher-rated articles. However, low-rated information has different characteristics. The probability of engaging with information in the high or very high category after posting a low-rated article is 0.36 and 0.29 respectively, which indicates the overlap of information sharing between credible and non-credible content. This complements the findings from previous sections, where low credible content is as likely to spread as high credible content.

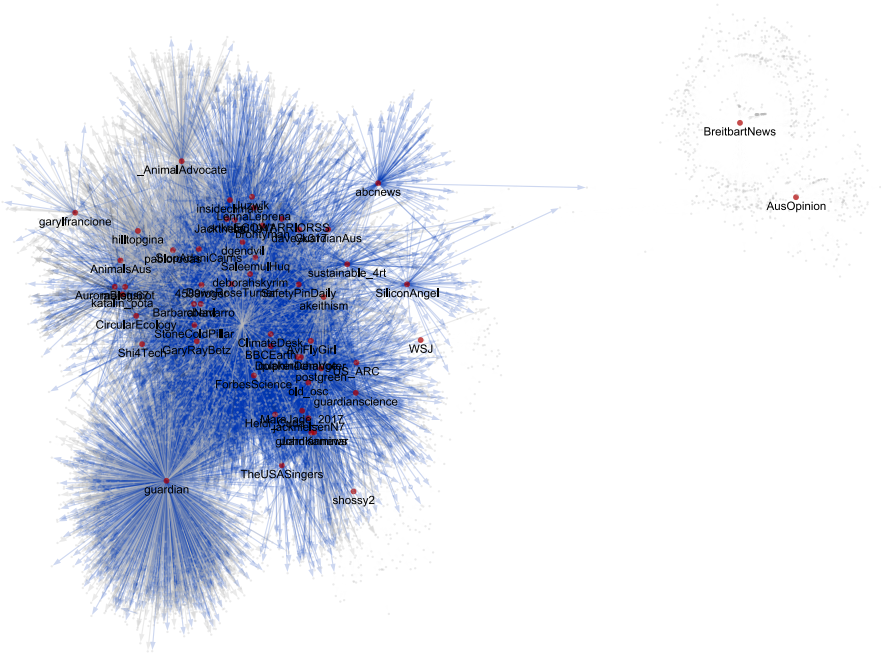
Furthermore, these findings highlight the strong homophily of users in high, and very high, where users tend to continuously engage with the same article rating category. In the bottom, right-most column, the probability of users reposting very low content given that they have posted very low content in the past is very high (P = 0.56). Similarly, there is a high probability that users continually post high-rated and very high-rated content given that they have posted in the same information category previously (P = 0.55, P = 0.49).

4.7. Influence of automated accounts

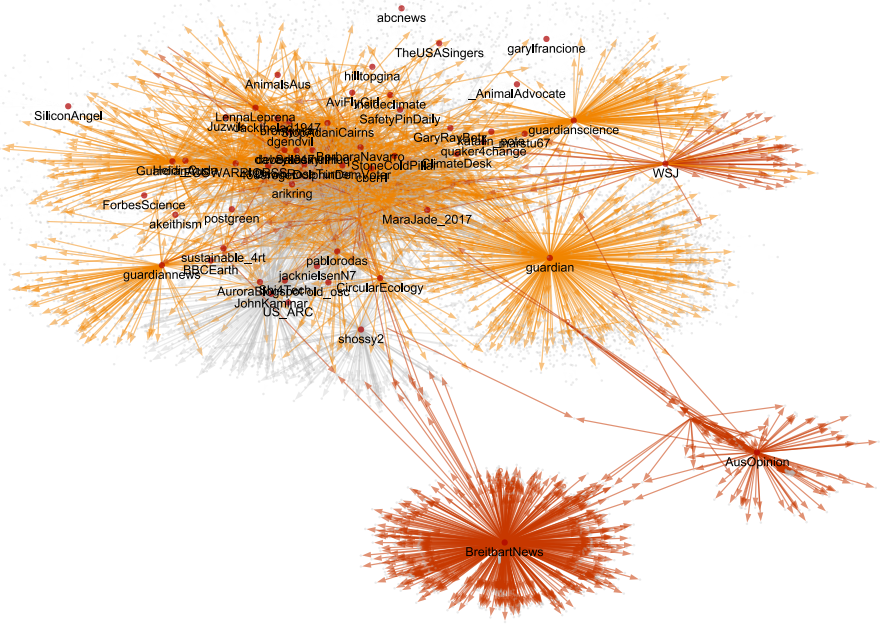
To conclude the exploration of climate misinformation networks based on climate science, this section discusses the influence of automated accounts in climate information diffusion from accounts that are categorised as software-controlled. As explained in section 3.1, the classification label of humans and automated accounts depends on how the *Botometer* was trained, which categorises media accounts such as *The Guardian* to be automated as they continuously post through another application and not the usual Twitter account like ordinary users. However, this section aims to discover the proportion of nodes with the highest centrality (degree and k-core) that are responsible for a large proportion of information cascade are automated, and whether they amplify misinformation in online social networks. Past studies have found numerous state-backed accounts or bots, that amplify misinformation campaigns in social media and, in turn, distort public information

dissemination. However, in climate change, the influence of automated accounts in misinformation campaigns is not as extensive.

Figure 23 shows the influence of automated accounts in climate information diffusion. Nodes with degree centrality in the top 5 percent of the distribution and are labelled as automated accounts are present in this figure, where the edge colors represent the scientific credibility. The red nodes are root users, and the neighborhood nodes of the root users are shown. Many of the automated accounts are media accounts that regularly publicise their newly published articles. As shown, the overlap of information sharing persists in diffusion networks from automated accounts, which signifies that climate misinformation sharing is not significantly driven by automated accounts as they also disseminate credible climate content. Except for the spread of very low credible content (red edges), from accounts like *BreitbartNews, Wall Street Journal (WSJ),* and *AusOpnion,* a few accounts do not share overlapping information between credible and non-credible. This highlights a different cause of climate misinformation in contrast to political misinformation, as misinformation strategies are not amplified by specific groups or automated accounts from specific groups.



(a) Spreading of credible climate information from automated accounts (rated high or very high)



(b) Spreading of misleading climate information from automated accounts (rated low or very low)

Figure 23: Automated accounts' influence on the diffusion of credible and misleading information

4.8. Patterns of misleading information

In misleading climate articles, there are two main kinds of climate misinformation in the media. First, there are articles with dubious claims from factually incorrect observational data, or claims based on a heavily biased interpretation, that question the integrity of science and experts. These are often labelled information with very low scientific credibility, such those articles that are published by media that often perpetuate misinformation (e.g. *Breitbart, The Daily Mail,* or *Fox New)* with headlines such as "*tumbling rocks are the causes of sea-level rise", or "Scientists Prove Man-made Global Warming is a Hoax".*

Second, articles are citing scientific evidence or scientists' claims that are published in more credible media such as *The Guardian* or *The Telegraph*. The examples of these low-rated article headlines are shown in Figure 24. These contain distorted portrayals of accurate scientific evidence, such as the information in Figure 24a, which gives impressions of uncertainties on an accurate factual statement, indicating a non-existent scientific debate. Other articles, such as in Figure 24b, were found to be exaggerating skewed scientific evidence that still requires further research. Science journalists in these media often attempt to communicate scientific findings in climate change, however, certain words, claims, or interpretations, either intentionally or unintentionally, shifts the information understanding.

From these two types of misleading climate content, three Twitter patterns of misleading information diffusion were found. This highlights the distinction of misinformation patterns between very low and low-rated ones, and how it results in differences in the network structure. The observation below is supported by tweets with misleading content from the central nodes (Figure 20).



(a) A mix of accurate scientific evidence with unsupported claims that are not supported by evidence create misleading information *(Source: The Telegraph)*

(b) Climate article citing preliminary research that lacks important context creating misleading information *(Source: The Guardian)*

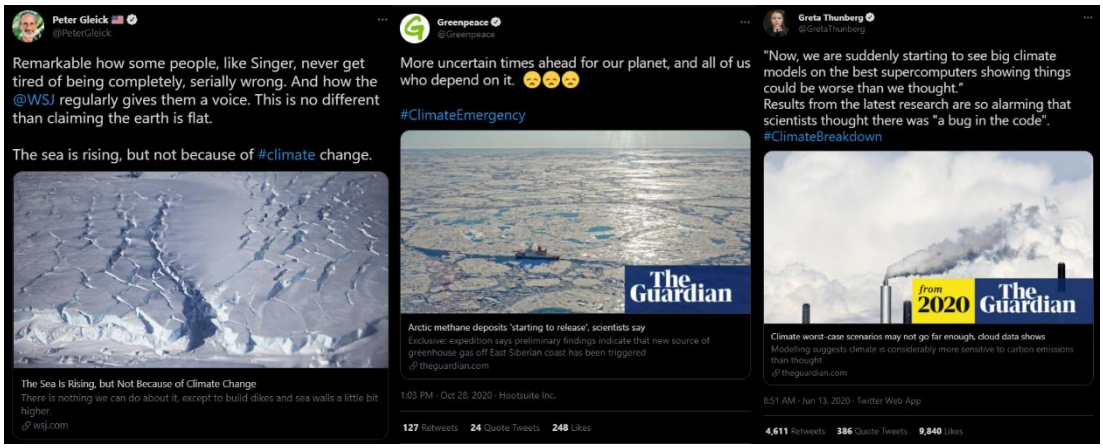
Figure 24: Example of misleading articles in the "low" category

First, some people often share very low-rated articles on Twitter but do not promote the information. They directly discredit very low-rated articles that completely misrepresent scientific findings, such as in Figure 25a.

The second type is the use of alarmism language in the Tweets to support claims made in the articles, such as "reached a point of no return", or "climate predictions are worse than current scenario" (see Figure 25b). They take advantage of the inconclusiveness of the research findings and the uncertainties inherent in climate science research to amplify a doomsday portrayal of climate change to push for an agenda. These are often portrayed in climate activists' accounts on Twitter, including *@greenpeace, @extinctionrebellion*, etc., where the goal is to warn the public about the urgency of the climate crisis regardless of the credibility of the content.

The third type is the low-rated articles that highlight new scientific research that is still debated in the scientific community. Similar to the previous type, the inherent uncertainties of scientific discoveries in climate change research become the central theme. While new findings highlight new implications, these are often areas in climate change research that need further research. The strong claims made in the articles are supported by scientific research with a weak scientific consensus, hence, making misleading claims about them. They often highlight scientific evidence with a lack of context, which results in clickbait headlines with claims like "the data shows", or "scientists say" that amplify the spread of information.

Furthermore, climate activists further amplify the spread with the attributions to "science-backed" information such as in Figure 25c.



(a) Discredit misleading article (b) Climate alarmism (c) Scientific uncertainty

Figure 25: Tweet snippets from three types of climate misinformation

These three general themes are found in the spread of low-rated climate information by influential nodes (see section 4.5). This sheds a light on the challenges of science communication, where people who are not in the scientific community are not as familiar with the process of scientific discovery (i.e. replication of results). The nature of climate misinformation heavily relies on scientific consensus for robust scientific evidence to ensure a clear distinction between upcoming and "scientifically-proven" research that supports viral claims online.

4.9. Summary

In summary, the findings do not support the hypothesis that social media constructs polarised and homogenous communities concerning science-based and non-science-based climate information. Instead, it shows a significantly smaller polarised community than hypothesised and greater cross-mixing information sharing of credible and non-credible information. The diffusion network is less polarised for low-rated articles than extremely misleading ones. This is evident in the temporal and structural analysis of information cascades, where there is a clearer distinction between extremely low-rated articles to others, but similar characteristics between low-rated articles and other categories.

The aggregated RT network shows how influential accounts also spread misleading climate information. This is confirmed in the analysis which measures the likelihood of homophily of information sharing patterns, and further explains the main patterns of misinformation characteristics, which includes discrediting misleading articles, using science uncertainties in climate alarmism, and exaggerating claims from skewed scientific evidence.

This highlights that climate misinformation perpetuates in both pro and anti-climate groups. The common themes revolve around the issues of scientific uncertainties, exaggerations, and distortion of claims from skewed scientific evidence.

5. Discussion

5.1. Networks of climate information diffusion

To answer **RQ1,** the article cascade networks' properties were examined to investigate the underlying network structure of diffusion mechanisms of climate misinformation in online social networks. The structural and temporal features highlight the cascading behaviour in online information sharing, where more credible information tends to diffuse faster and more widely. This is evident from the larger number of cascades found in higher-rated articles and from the attention patterns show in the empirical CCDF plots, where more cohesive networks improve the effectiveness of information propagation.

The structural properties of article cascades show the diffusion mechanisms for different rate categories. Collectively, the number of clusters, size, and diameter of clusters show how information spreading can have a broadcast-like or word-of-mouth-like diffusion. In a broadcast-like diffusion many users RT and thus, diffuse the information faster through the amplification of the tweet's exposure to users' followings. An interesting observation is the number of self RT found, especially in the diffusion of less credible content, that may be a pattern to improve attention to information. The network degree distribution reveals user engagement in climate content on Twitter, where a large majority of users only RT and a very small percentage of users are dominating root users. This supports previous findings of the Twitter population where a few people produce the biggest proportion of tweets and others, the "silent majority", where about 40 percent of Twitter users frequently read tweets but never tweet themselves (Kwak et al., 2010)

The article cascade networks showcase a few users who engage in the same article multiple times. Extremely low content has distinct diffusion characteristics to other rate categories. The structural and temporal properties of each article-level cascade network indicate how very low credible information diffuses slower, has lower attention, and has a more clustered network structure. Many of the unsuccessful diffusions show less cohesive networks, where there are smaller and fewer influential nodes initiating the spread.

Overall, these findings exemplify the phenomena of diffusion and contagion confounding in the spread of information in online social networks. As supported in past studies, while it

is difficult to differentiate the causation of homophily and social influence on information diffusion in observational data (Bakshy et al., 2009, 2012; Shalizi & Thomas, 2011), the temporal and structural properties of the diffusion network highlight the presence of both.

5.2. The diffusion of credible and misleading climate information

RQ2 aims to investigate the contrast between credible and misleading climate information diffusion networks. The whole RT network highlight the presence of weak polarisation over climate science between low and high-rated articles. The edge homophily and whole RT network analyses further demonstrated how and why the communities of credible and misleading climate information diffusion networks overlap. This implies the cross-mixing of information from different categories, as supported by the relatively high probability of users engaging in information from both low and high categories. Furthermore, it shows the extent of science polarisation in social media that is not as severe as the polarisation between political accounts, or between pro and anti-climate groups in the literature.

The analyses found the presence of climate misinformation in both pro and anti-climate groups. While the anti-climate groups often breed the diffusion of misleading content, pro-climate groups were found to similarly propagate misinformation. The spread of climate misinformation by the most influential accounts sheds a light on some misinformation diffusion patterns. This includes discrediting science, attacking science, highlighting scientific uncertainties for climate alarmism, or exaggerating skewed scientific evidence.

Many climate articles in popular media have low scientific credibility due to the exaggeration and cherry-picking of scientific findings or research that still undergoes research. For example, overstating the result of research while not putting it in important context, oversimplification of climate science, and overstating claims from a single finding. Furthermore, many of these articles lack important context when discussing scientific research, such as comparison of other effects or overstating causal effects, which create misleading portrayal through a few keywords such as "initiated" or "triggered". This is finding is supported in previous research, where climate misinformation strategies include cherry-picking evidence, and highlighting biased scientific research (Cook et al., 2017).

5.3. Contributions and Recommendations

Regardless of the climate ideology (pro or anti-climate), the misuse of scientific research in climate change information discourse leads to climate misinformation. For example, the scientific basis for an apocalyptic climate change is inaccurate, which is propagated through many climate activist Twitter accounts. Therefore, there is a danger of distorting scientific evidence through communicating climate information in alarmism language, which is often used by climate activists and in turn, is attacked by anti-climate groups. This further exacerbates the polarisation between the pro and anti-climate groups, while alternatively, scientific evidence should be where the two groups reach agreements. The findings highlight the presence of contestations on science in climate change rather than contestations based on science. Currently, the climate change discourse around climate science revolves around the trust or distrust in science, and the use or misuse of evidence, rather than debates around science-based actions to climate change. As a result, there needs to be a stronger scientific consensus to reduce the uncertainties in the media, and therefore, to manage climate misinformation in public.

Overall, the findings of this research contribute to the understanding of science-based information dissemination from the perspective of climate science in climate change discourse. Therein, I have highlighted how climate science diffuses amongst the public, where accurate scientific understanding relies on intermediaries such as science journalists, climate activists, and media channels. Context and consensus are important in science as the basis for accurate science-based information. The diffusion networks of climate science in climate information spreading further show the important role of scientific consensus and holistic understanding of the scientific discovery process. When communicating science-based information, it is imperative to understand scientific research in the greater context, such as where it is perceived in the scientific community, or the methods and biases inherent in the research formulation.

The findings of this research shed a light on the current state of climate misinformation in online social networks and its relation with science communication. However, to extend this thesis, there are some recommendations for future work including utilising other platforms and annotation tools other than Twitter and Climate Feedback, focusing on a specific event to monitor event-specific movement and information strategies such as the UN Climate

Change Conference, and including thread and replies interactions other than RTs to get a more comprehensive analysis of climate information discourse online. One limitation in this research is the data collection method that is anchored to the article data from Climate Feedback. Thus, future work should focus on a time-specific climate event and annotate the scientific credibility of posts regardless of the presence of articles.

6. Conclusion

This research explored the network structure of the spread of climate misinformation based on climate science. From the annotation of climate scientists from Climate Feedback, this thesis uses data from more than 150 articles to analyse the diffusion of credible and misleading climate information on Twitter. Using social network analysis, the temporal and structural features highlighted some similarities and differences between the spread of credible and misleading climate information.

The network structure of climate information diffusion found a less significant polarisation between science-based and non-science-based content than political polarisation commonly found in Twitter RT networks. The exploration of information cascades and RT networks concluded that highly credible content has similar diffusion characteristics to low-rated content. By measuring node centrality and the likelihood of users engaging with articles from different rating categories, this research further confirms that climate information with low scientific credibility is as likely to diffuse as those of high scientific credibility. Further, a qualitative exploration of how misleading climate information is portrayed from the tweets of influential accounts shows three general patterns of how climate science is inaccurately portrayed in the media. This includes the pro-climate groups, where patterns include exaggeration of scientific evidence and the use of scientific uncertainties for doomsday portrayal and climate alarmism.

This has consequences for science communication and managing climate misinformation. Clear scientific consensus in climate science helps to reduce inaccurate portrayal of science in science-based claims, opinions, or articles. Thus, this is where large, intergovernmental climate change institutions such as IPCC hold an important role in managing climate misinformation and clear communication messages, as the foundation of scientific consensus. These analyses shed a light on the importance of science journalism in reducing the spread of misleading "science-based" climate claims in online social networks.

7. References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2). https://doi.org/10.1257/jep.31.2.211

Alperin, J. P., Gomez, C. J., & Haustein, S. (2019). Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles. *Public Understanding of Science*, *28*(1). https://doi.org/10.1177/0963662518761733

Alrajebah, N., Tiropanis, T., & Carr, L. (2017). Cascades on Online Social Networks: A Chronological Account. *Internet Science*. https://doi.org/https://doi.org/10.1007/978-3-319-70284-1_31

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., & Tiwari, M. (2015). Global diffusion via cascading invitations: Structure, growth, and homophily. *WWW 2015 - Proceedings of the 24th International Conference on World Wide Web*, 66–76. https://doi.org/10.1145/2736277.2741672

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549. https://doi.org/10.1073/PNAS.0908800106

Bakshy, E., Karrer, B., & Adamic, L. A. (2009). *Social Influence and the Diffusion of User-Created Content*.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web*, 519–528. https://doi.org/10.1145/2187836.2187907

Bennett, W. L., & Manheim, J. B. (2006). The one-step flow of communication. *Annals of the American Academy of Political and Social Science*, *608*(1), 213–232. https://doi.org/10.1177/0002716206292266

Bentley, R. A., Garnett, P., O'Brien, M. J., & Brock, W. A. (2012). Word Diffusion and Climate Science. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0047966

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs Conspiracy: Collective Narratives in the Age of Misinformation. *PLOS ONE*, *10*(2), e0118093. https://doi.org/10.1371/JOURNAL.PONE.0118093

Bild, D. R., Liu, Y., Dick, R. P., Mao, Z. M., & Wallach, D. S. (2015). Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph. *ACM Transactions on Internet Technology*, *15*(1). https://doi.org/10.1145/2700060

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-07761-2

Brown, P. E., & Feng, J. (2011). *Measuring User Influence on Twitter Using Modified K-Shell Decomposition*. Association for the Advancement of Artificial Intelligence. https://ojs.aaai.org/index.php/ICWSM/article/view/14210/14059

Burt, R. S. (1987). Social Contagion and Innovation: Cohesion versus Structural Equivalence. In *Source: American Journal of Sociology* (Vol. 92, Issue 6). https://about.jstor.org/terms

Cann, T. J. B., Weaver, I. S., & Williams, H. T. P. (2021). Ideological biases in social sharing of online information about climate change. *PLoS ONE*, *16*(4 April), e0250656. https://doi.org/10.1371/journal.pone.0250656

Centola, D. (2015). The Social Origins of Networks and Diffusion [Article]. *American Journal of Sociology*, *120*(5), 1295–1338. https://doi.org/10.1086/681275

Cha, M., Benevenuto, F., Haddadi, H., & Gummadi, K. (2012). *The world of connections and information flow in Twitter*.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *Association for the Advancement of Artificial Intelligence*. www.aaai.org

Cheng, J., Adamic, L. A., Kleinberg, J. M., & Leskovec, J. (2016, April 11). Do Cascades Recur? *Proceedings of the 25th International Conference on World Wide Web*. https://doi.org/10.1145/2872427.2882993

Coleman, J. S., Katz, E., & Menzel, H. (1966). *Medical innovation: a diffusion study*. Bobbs-Merrill Co.

Comarela, G., Crovella, M., Almeida, V., & Benevenuto, F. (2012). Understanding factors that affect response rates in twitter. *HT'12 - Proceedings of 23rd ACM Conference on Hypertext and Social Media*, 123–132. https://doi.org/10.1145/2309996.2310017

Cook, J. (2019). Understanding and countering misinformation about climate change. *Handbook of Research on Deception, Fake News, and Misinformation Online*, 281–306.

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS ONE*, *12*(5). https://doi.org/10.1371/JOURNAL.PONE.0175799

Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., & Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, *8*(2), 24024. https://doi.org/10.1088/1748-9326/8/2/024024

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., & Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, *11*(4), 048002. https://doi.org/10.1088/1748-9326/11/4/048002

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). *BotOrNot: A System to Evaluate Social Bots*. 273–274. https://doi.org/10.1145/2872518.2889302

del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3). https://doi.org/10.1073/pnas.1517441113

Dow, A., Adamic, L. A., & Friggeri, A. (2013). The Anatomy of Large Facebook Cascades. *Association for the Advancement of Artificial Intelligence*. www.aaai.org

Eyal, G. (2019). *The crisis of expertise*.

Farrell, J. (2016a). *Corporate funding and ideological polarization about climate change*. *113*(1). https://doi.org/10.1073/pnas.1509433112

Farrell, J. (2016b). Network structure and influence of the climate change counter-movement. *Nature Climate Change*, *6*(4), 370–374. https://doi.org/10.1038/nclimate2875

Goel, A., Munagala, K., Sharma, A., & Zhang, H. (2015). A note on modeling retweet cascades on Twitter. *Proceedings of the 12th International Workshop on Algorithms and Models for the Web Graph*, *9479*, 119–131.

Goel, S., Watts, D. J., & Goldstein, D. G. (2012). *The Structure of Online Diffusion Networks*.

Gruhl, D., Liben-Nowell, D., Guha, R., & Tomkins, A. (2004). Information diffusion through blogspace. *Thirteenth International World Wide Web Conference Proceedings, WWW2004*, 491–501. https://doi.org/10.1145/988672.988739

GuilleAdrien, HacidHakim, FavreCecile, & A., Z. (2013). Information diffusion in online social networks. *ACM SIGMOD Record*, *42*(2), 17–28. https://doi.org/10.1145/2503792.2503797

Hamilton, L. C. (2016). Public Awareness of the Scientific Consensus on Climate. *SAGE Open*, *6*(4). https://doi.org/10.1177/2158244016676296

IFCN. (2021). *IFCN Code of Principles*. https://www.ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles

IPCC. (2014). *Climate Change 2014: Synthesis Report*.

IPCC. (2021). *What is the IPCC?* https://www.ipcc.ch/site/assets/uploads/2018/02/TOR_Bureau.pdf

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10). https://doi.org/10.1038/nclimate1547

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence : the part played by people in the flow of mass communications*. Free Press.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). *Maximizing the Spread of Influence through a Social Network*.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*. https://doi.org/10.1038/NPHYS1746

Krishna Kumar, K. P., & Geethakumari, G. (2013). Information diffusion model for spread of misinformation in online social networks. *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2013*, 1172–1177. https://doi.org/10.1109/ICACCI.2013.6637343

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is Twitter, a Social Network or a News Media?* http://bit.ly

Linden, S. L. van der, Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2015). The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence. *PLOS ONE*, *10*(2), e0118489. https://doi.org/10.1371/JOURNAL.PONE.0118489

Linden, S. van der, Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/GCH2.201600008

McCright, A. M., Dentzman, K., Charters, M., & Dietz, T. (2013). The influence of political ideology on trust in science. *Environmental Research Letters*, *8*(4). https://doi.org/10.1088/1748-9326/8/4/044029

McCright, A. M., Dunlap, R. E., & Xiao, C. (2013). Perceived scientific agreement and support for government action on climate change in the USA. *Climatic Change 2013 119:2*, *119*(2), 511–518. https://doi.org/10.1007/S10584-013-0704-9

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444. http://ezproxy-prd.bodleian.ox.ac.uk:2154/stable/2678628

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., & Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*. https://doi.org/10.1038/nature08823

Myers, S., & Leskovec, J. (2014). *The Bursty Dynamics of the Twitter Information Network*.

Nsoesie, E. O., & Oladeji, O. (2020). Identifying patterns to prevent the spread of misinformation during epidemics. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-014

Oreskes, N. (2004). Beyond the Ivory Tower: The scientific consensus on climatic change. In *Science* (Vol. 306, Issue 5702, p. 1686). American Association for the Advancement of Science. https://doi.org/10.1126/science.1103618

Oreskes, N. (2021). *Why trust science?* (First paperback e...).

Oreskes, N., & Conway, E. M. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.

Palmer, T., & Stevens, B. (2019). The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences*, *116*(49). https://doi.org/10.1073/pnas.1906691116

Pei, S., Muchnik, L., José S. Andrade, Jr., Zheng, Z., & Makse, H. A. (2014). Searching for superspreaders of information in real-world social media. *Scientific Reports 2014 4:1*, *4*(1), 1–12. https://doi.org/10.1038/srep05547

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*. https://doi.org/10.1038/s41586-021-03344-2

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Pozzana, I., & Ferrara, E. (n.d.). *Measuring bot and human behavioral dynamics*. Retrieved August 2, 2021, from https://botometer.iuni.iu.edu/

Rapport, F., Clay-Williams, R., Churruca, K., Shih, P., Hogden, A., & Braithwaite, J. (2018). The struggle of translating science into action: Foundational concepts of implementation science. *Journal of Evaluation in Clinical Practice*, *24*(1), 117–126. https://doi.org/10.1111/jep.12741

Rauchfleisch, A., & Kaiser, J. (2020). The False positive problem of automatic bot detection in social science research. *PLoS ONE*, *15*(10). https://doi.org/10.1371/JOURNAL.PONE.0241045

Rekker, R. (2021). The nature and origins of political polarization over science. *Public Understanding of Science*. https://doi.org/10.1177/0963662521989193

Reusswig, F. (2013). History and future of the scientific consensus on anthropogenic global warming. *Environmental Research Letters*, *8*(3). https://doi.org/10.1088/1748-9326/8/3/031003

Richardson, L. (2007). *Beautiful soup documentation*.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., … Foley, J. A. (2009). A safe operating space for humanity. *Nature 2009 461:7263*, *461*(7263), 472–475. https://doi.org/10.1038/461472a

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10). https://doi.org/10.1098/rsos.201199

Rudat, A., & Buder, J. (2015). Making retweeting social: The influence of content and context information on sharing news in Twitter. *Computers in Human Behavior*, *46*. https://doi.org/10.1016/j.chb.2015.01.005

Rudat, A., Buder, J., & Hesse, F. W. (2014). Audience design in Twitter: Retweeting behavior between informational value and followers' interests. *Computers in Human Behavior*, *35*. https://doi.org/10.1016/j.chb.2014.03.006

Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2020). Detection of Novel Social Bots by Ensembles of Specialized Classifiers. *International Conference on Information and Knowledge Management, Proceedings*, 2725–2732. https://doi.org/10.1145/3340531.3412698

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16). https://doi.org/10.1073/pnas.1805871115

Science Feedback. (2019). *Science Feedback partnering with Facebook in fight against misinformation*. https://sciencefeedback.co/science-feedback-partnering-with-facebook-in-fight-against-misinformation/

Science Feedback. (2020, August). *Science Feedback is awarded an IFCN Development Grant to build database on sources of misinformation*. https://sciencefeedback.co/science-feedback-is-awarded-an-ifcn-development-grant-to-build-database-on-sources-of-misinformation/

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287. https://doi.org/10.1016/0378-8733(83)90028-X

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, *40*(2), 211–239. https://doi.org/10.1177/0049124111404820

Shao, C., Hui, P.-M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PLOS ONE*, *13*(4), e0196087. https://doi.org/10.1371/JOURNAL.PONE.0196087

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W., de Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B., & Sörlin, S. (2015). *Planetary boundaries: Guiding human development on a changing planet*. https://doi.org/10.1126/science.1259855

Strang, D. (2016). Adding Social Structure to Diffusion Models: An Event History Framework. *Http://Dx.Doi.Org/10.1177/0049124191019003003*, *19*(3), 324–353. https://doi.org/10.1177/0049124191019003003

Taxidou, I., & Fischer, P. M. (2014). Online analysis of information diffusion in twitter. *WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web*, 1313–1318. https://doi.org/10.1145/2567948.2580050

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLOS ONE*, *13*(9). https://doi.org/10.1371/journal.pone.0203958

Treen, K. M. d. I., Williams, H. T. P., & O'Neill, S. J. (2020). Online misinformation about climate change. In *Wiley Interdisciplinary Reviews: Climate Change* (Vol. 11, Issue 5). Wiley-Blackwell. https://doi.org/10.1002/wcc.665

Treen, K. M. d'I., Williams, H. T. P., & O'Neill, S. J. (2020). Online misinformation about climate change. *WIREs Climate Change*, *11*(5), e665. https://doi.org/10.1002/wcc.665

Valente, T. W. (1996a). Network Models of the Diffusion of Innovations. *Computational and Mathematical Organization Theory*, *2*(2), 163–164.

Valente, T. W. (1996b). *Social network thresholds in the diffusion of innovations*.

Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 280–289. https://ojs.aaai.org/index.php/ICWSM/article/view/14871

Vosoughi, S., Roy, D., & Aral, S. (2018). *The spread of true and false news online*. http://science.sciencemag.org/

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(9), 5766–5771. https://doi.org/10.1073/pnas.082090499

Williams, H. T. P., McMurray, J. R., Kurz, T., & Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, *32*, 126–138. https://doi.org/10.1016/J.GLOENVCHA.2015.03.006

Xia, Y., Hsuan Yun Chen, T., Kivelä, M., & ii, K. (2020). *Spread of Tweets in Climate Discussions*.

Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48–61. https://doi.org/10.1002/hbe2.115

8. Appendices

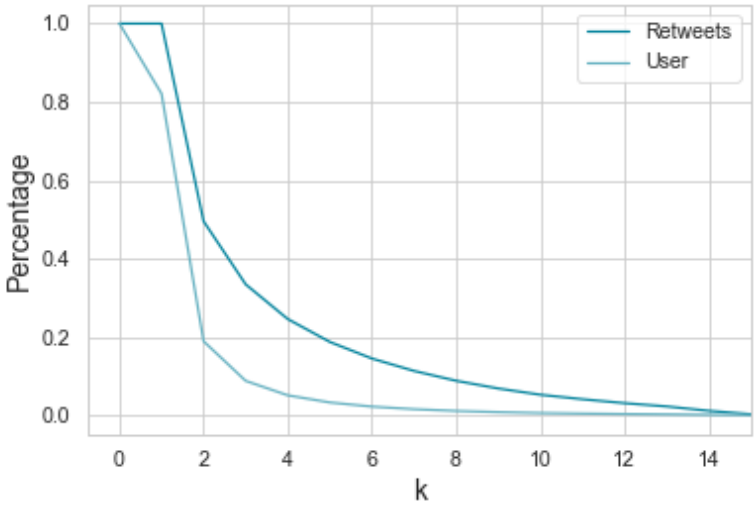
8.1. Appendix A – Publishing media and their number of articles in the data

Publisher	No of Articles
Australian Broadcasting Company	1
BBC	3
Bloomberg	1
Breitbart	6
Business Insider	2
CNN	6
Conservative Tribune	1
Daily Mail	2
Financial Post	3
Forbes	10
Fox2News, CNN	1
Futurism	1
Grist	1
IFLScience	2
Investors' Business Daily	1
Los Angeles Times	1
Mashable	5
NBC News	1
NPR	1
National Geographic	3
New York Magazine	1
New York Post	1
Rolling Stone	1
Science Magazine, E&E News	1
Sky News	1
The Atlantic	4
The Australian	5
The Boston Globe	1
The Daily Caller	2
The Daily Wire	3
The Guardian	15
The Independent	3
The New York Times	17
The Spectator	1
The Sun	1
The Telegraph	5
The Times / The Australian	1
The Wall Street Journal	9
The Washington Examiner	1
The Washington Post	13
The Washington Times	2
The Western Journal	1

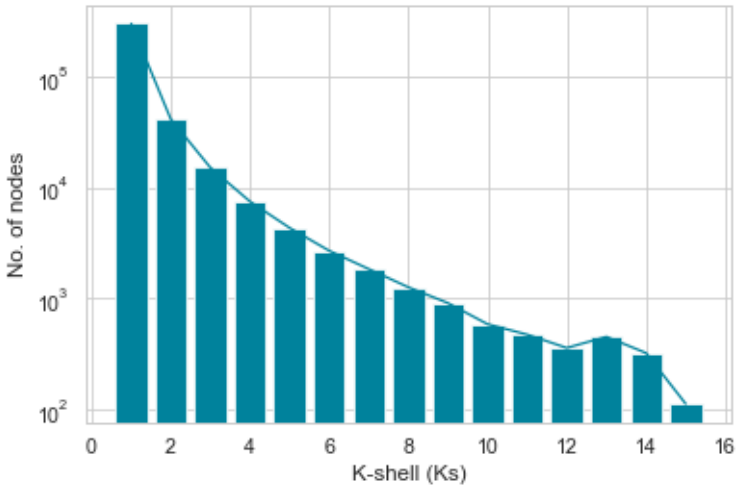
Think Progress	2
USA Today	5
Vatican.va	1
Vox	2

8.2. Appendix B

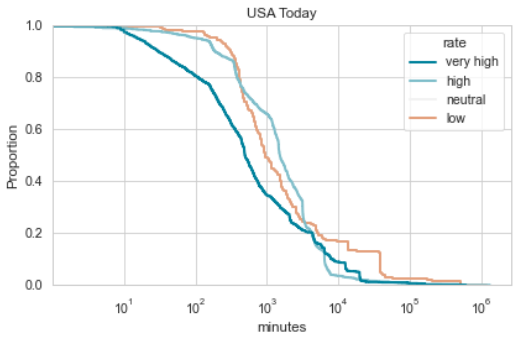
(a) Percentage of RT (edges) and users (nodes) proportion in k-core decomposition

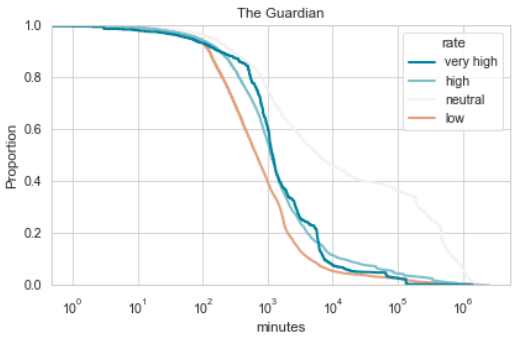


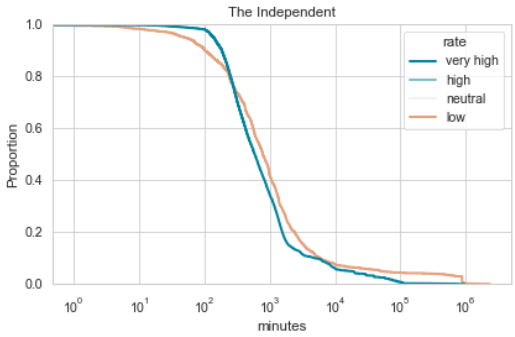
(b) Number of nodes each K-shell

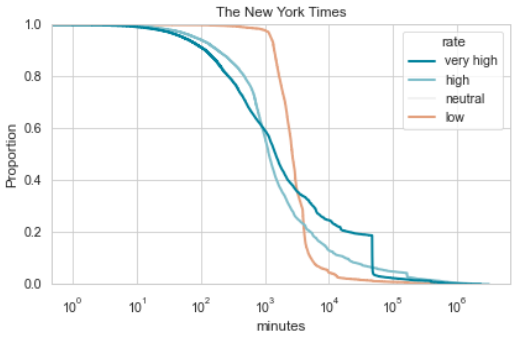


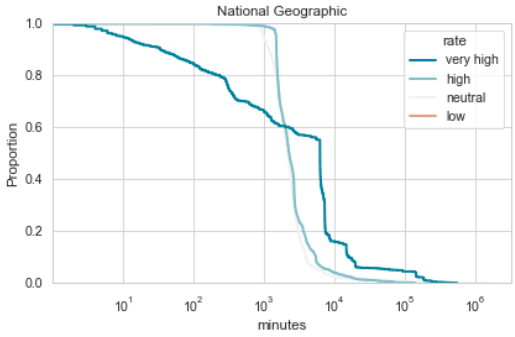
8.3. Appendix C – Temporal distance of proportion of tweets with CCDF plots based on media

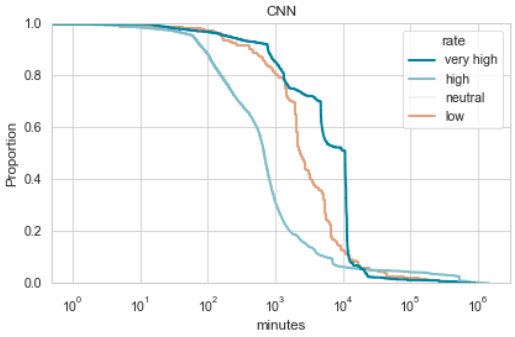












8.4. Appendix D – Accounts in the k-core-shell of retweet network who shares low rated climate articles

