



Are social media platforms accelerators of democratic fragmentation?

An inquiry into the relationship of democratic fragmentation and social media platforms in three political arenas: political campaigns, protest movements and democratic institutions

Berlin, Spring 2023

Philipp Darius

Dissertation submitted to the Hertie School
in partial fulfilment of the requirements for the degree of

Doctor rerum politicarum (Dr. rer. pol)

in the

Doctoral Programme in Governance

Advisors

First advisor

Prof. Dr Andrea Römmele
Hertie School

Second advisor

Prof. Rachel Gibson, PhD
University of Manchester

Third advisor

Prof. Dr Ulrike Klinger
European University Viadrina

Summary

Whilst social media platforms provide global communication environments, these platforms are not primarily designed for political debates. This may have adverse effects on democracy and contribute to democratic fragmentation. This dissertation project investigates the role of social media platforms as potential accelerators of fragmentation in established democracies. The overarching question of this dissertation explores to what extent social media platforms may contribute to fragmentation that may result in the erosion of democracy. This dissertation comprises three articles focusing on three political arenas 1) political campaigns, 2) protest movements, and 3) democratic institutions.

The first paper investigates the polarisation of online political behaviour on Twitter in democratic election campaigns. The analysis is based on Twitter data of German political parties and election hashtags during the final week of the 2017 and 2021 German Federal elections. The study's findings suggest that the far-right party AfD seeks to polarise online discourse as a strategy and that far-right online partisans may influence the public reception of politicians and established parties.

The second paper examines the formation of protest movements on Twitter during the Covid-19 pandemic in the United Kingdom. The study monitors two established conspiracy narratives and their communities on Twitter, the anti-vaccination and anti-5G communities, before and during the first UK lockdown. The study finds that, despite content moderation efforts by Twitter, conspiracy groups were able to proliferate their messages and influence broader public discourses on Twitter, such as #Lockdown in the United Kingdom. The findings underline social platforms' potential for protest formation that can result in disinformed social movements.

The third paper inspects social platform companies' mimicking of democratic institutions like Meta's oversight board for content decisions that may erode existing democratic institutions. The study traces the emergence of the supreme court metaphor for Meta's oversight board and its use in the US News discourse. The findings emphasise how private organisations and the use of constitutional metaphors can erode the legitimacy of supreme courts and other democratic institutions.

The dissertation's findings imply that social media platforms, besides their democratically desirable function for democratic participation and deliberative potentials, contribute to democratic fragmentation in the inspected arenas of political campaigns, disinformed protest movements and democratic institutions. Thus, extended research data access is key to better understanding the social implications of social media platforms and finding adequate regulations of recommender systems, content moderation, and advertisement-based business models.

Acknowledgments

In the past years, I often heard people speaking of the PhD journey, which sounded, on the one hand, surprising but also appealing when understanding education as a continued formation of character and knowledge. Now, after three and a half years, I am looking back, and it truly feels like a journey. From confusion in the beginning and difficult times during the Covid-19 pandemic to highlights at conferences, accepted papers and the privilege of getting to know numerous new colleagues and friends in the past three and a half years of my doctoral journey.

This dissertation would not have been possible without the support of and collaboration with many people, of which I would like to briefly mention the most important and excuse myself to those I might have forgotten. Moreover, I had the immense privilege of receiving a PhD stipend from the Hertie Foundation, without which I would not have pursued a PhD. I would like to express my gratitude to the Hertie Foundation and the Hertie School.

First, I would like to thank my supervisors Prof. Andrea Römmele, Prof. Rachel Gibson, and Prof. Ulrike Klinger, for their supervision and support throughout the past years. Working in the ERC-funded Digital Campaigning and Electoral Democracy (DiCED) project with Rachel and Andrea has taught me about political campaign research and working with an international team of scholars. Ulrike has invited me to colloquia and recommended the Milan Summer School in Political Communication, which was very helpful in developing my research ideas. I am grateful to all three of you for helping me develop a theoretical framework for my dissertation and supporting me whenever I doubted my research.

In addition, I would like to thank all Hertie staff, especially the PhD-Team, for making the Hertie School a great place to pursue a PhD in the Political and Social Sciences. Here, I also want to thank the PhD community at the Hertie School, where PhD candidates, besides an incredible diversity of topics, listen carefully to each other, provide constructive feedback and help each other out. I am proud to have been and remain a part of the Hertie School community.

Additionally, I would like to thank my collaborators, Dr Fabian Stephany and Dr Sebastian Stier. I have learned a lot about collaborative projects and conducting independent research projects, and I am looking forward to future publications and further collaboration. As much I am thankful to my co-authors and friends Michael Urquhart, Moritz Schramm, Josh Cowsls, and Dominiquo Santistevan for co-authoring studies that are part of my dissertation project. I immensely enjoyed working together and look forward to doing so in the future!

During my PhD journey, I also had the chance to visit the Data Science Institute at the London School of Economics, funded by a CIVICA mobility grant and the EUROLAB at the Leibniz Institute for Social Science (GESIS) in Cologne, funded by a GESIS Eurolab grant. I highly appreciated the time at both institutions. I would like to thank Prof. Ken Benoit and colleagues at the LSE, Dr Sebastian Stier, Prof. Ingvill Ødegaard, and Dr Marlene Mauk at GESIS for integrating me into their teams and making the visits instructive and enjoyable. Another remarkable experience was my participation in a research sprint on AI and content moderation organised by the Humboldt Institute for Internet and Society, where I learned so much about content moderation and the governance of platforms and virtually met some of my later co-authors. The sprint was, especially during the pandemic, a wonderful interdisciplinary research experience, and I would like to thank the organisers and other participants.

Without my friends and family and their love and support, I would never have had the strength to pursue this sometimes-difficult journey. I would especially like to thank my parents and my partner Penelope who makes the sun shine even on Berlin's greyest winter days. Thank you all.

Table of Contents

Summary.....	i
Acknowledgments.....	iii
List of tables.....	viii
List of figures.....	viii
List of abbreviations.....	ix
Chapter 1 - Introduction.....	1
Chapter 2 – Background.....	7
2.1. Social media platforms as fragmented communication environments.....	7
2.2. Social media platforms as environments for political election campaigns.....	9
2.3 Social media platforms as environments for civic participation and protest.....	10
2.4. Social media platforms and democratic institutions.....	11
Chapter 3 – Who polarises Twitter? Ideological polarisation, partisan groups and strategic networked campaigning on Twitter during the 2017 and 2021 German federal elections ‘Bundestagswahlen’.....	13
Abstract.....	13
3.1. Introduction.....	14
3.2. Background.....	15
3.2.2. Twitter as a political communication tool and forum for public discourse.....	16
3.2.3. Online polarisation as a strategy?.....	18
3.3. Research questions.....	21
3.4. Research design.....	22
3.4.1. Data collection.....	22
3.4.2. Network approaches for the detection of partisan communities.....	23
3.4.3. Measurement of polarisation and strategic hashtag use.....	25
3.5. Results.....	26
3.5.1. Partisan polarisation of party hashtags.....	27
3.5.2. Behaviour of partisan communities.....	28
3.5.3. Polarisation of the broader electoral debate (RQ3a & RQ3b).....	29
3.5.4. The likelihood of partisan co-occurrence in the broader electoral debate (RQ4).....	31
3.6. Discussion.....	32
3.6.1. Findings.....	33
3.6.2. Limitations of the study and methodology.....	34
3.6.3. Further research.....	35
3.6.4. Conclusions.....	36
References.....	38

Chapter 4 – Disinformed social movements: A large-scale mapping of conspiracy narratives as online harms during the COVID-19 pandemic	47
Abstract	47
4.1. Introduction	48
4.2. Background	49
4.2.1. Theoretical Framework	49
4.2.2. The Anti-5G and Anti-Vaccination Movements.....	51
4.2.3. Prior empirical research and knowledge gap	53
4.3. Methodology.....	54
4.3.1. Network representation.....	57
4.3.2. Community detection in network graphs.....	58
4.3.3. Qualitative content analysis of detected communities.....	59
4.3.4. Co-occurrence between hashtag discourses	59
4.3.5. Assessing follower growth of network hubs	60
4.4. Analysis and findings	60
4.4.1. Exploring a spectrum.....	60
4.4.2. Mapping networks before (T1) and after (T2) the introduction of the 2020 UK Lockdown	61
4.4.3. Network hubs and follower growth	63
4.4.5. Network interaction and community building	64
4.5. Discussion	69
4.5.1. Discussion of the results.....	69
4.5.2. Discussion of conspiracy theories as social movements	71
2.5.3. Limitations and further research.....	72
2.5.4. Conclusions.....	73
References	74
Chapter 5 – Constitutional metaphors: Facebook's "supreme court" and the legitimization of platform governance	79
Abstract	79
5.1. Introduction: Platform governance, the language of statehood and ‘constitutional metaphors’	80
5.2. Background	82
5.2.1. Metaphors and the language of constitutional democracy	82
5.2.2. The emergence and spread of ‘Facebook’s Supreme Court’.....	84
5.2.3. Legitimacy and legitimation in platform governance: the influence of constitutional metaphors... ..	86
5.3. Documenting the rise of the ‘supreme court’ metaphor	89
5.3.1 Methodology.....	89
5.4. Discussion of constitutional metaphors and the need for critical reflections.....	100
5.4.1. What is a supreme court anyway?	100
5.4.2. The legitimating effect of the supreme court metaphor	102
5.4.3. The construction of meaning through constitutional metaphors	103
5.5. Conclusion.....	104

References	106
Chapter – 6 Discussion and Outlook	112
<i>6.1 Discussion of findings and implications.....</i>	<i>112</i>
<i>6.2 Discussion of the limitations and ethical considerations.....</i>	<i>115</i>
<i>6.3. Discussion of regulatory initiatives and Internet governance</i>	<i>116</i>
6.4. Outlook and Conclusions	118
Bibliography	120
Appendix.....	126
List of papers.....	136

List of tables

Table 1: Summary statistics for the selected hashtag discourses 56

Table 2: Mean follower numbers and increase of conspiracy and civil society network hubs 63

Table 3: Average weighted indegree for main pro-conspiracy and civil society clusters in the RT networks..... 64

Table 4: Proportion (in %) of conspiracy and civil society cluster members that used #5G and #Vaccines between T1 (January 1, 2020 – March 22, 2020) and T2 (March 23, 2020 – June 10, 2020)..... 65

Table 5: Chronology of main events in the media discourse surrounding use of OB:SC 93

Table A1: Summary of collected retweet networks..... 126

Table A2: Top 30 most retweeted accounts in #BTW17..... 127

Table A3: Top 30 most retweeted accounts in #BTW21..... 128

Table A4: Cluster membership of official federal-level party accounts in 2017 and 2021... 129

List of figures

Figure 1: Pipeline of the analytical approach to assess the polarisation of hashtags within and between hashtag discourses 22

Figure 2: Column chart of pro-partisans for each party and contra-communities as members of the largest community consisting of other partisans 27

Figure 3: Activity of partisans of various parties contrasted between the final week of the election campaign in 2017 and 2021 28

Figure 4: Retweet network polarisation of #BTW17 (top) and #BTW21 (bottom). 30

Figure 5: Log-likelihoods of the co-occurrence of partisans in the two major communities of the broader debate on the elections on #BTW17 in 2017 and #BTW21 in 2021 32

Figure 6: Schematic diagram of the research design..... 56

Figure 7: Conversational network types on Twitter 58

Figure 8: Retweet networks of #5G in T1 and T2 with cluster names 62

Figure 9: Proportion of individual cluster nodes co-occurrence 67

Figure 10: Number of US online news articles using the OB:SC metaphor per day, October 2018 to May 2021 (Media Cloud 'United States – National' Collection) 92

Figure A 1: Illustration of 'hashjacking' in Darius and Stephany (2019)..... 130

Figure A 2: Log-likelihoods of the co-occurrence of partisans in the two major network clusters..... 130

Figure A 3: Example post 1 131

Figure A 4: Example post 2 131

Figure A 5: Example post 3 132

Figure A 6: Example post 4..... 132

Figure A 7: Example post 5 133

Figure A 8: #Lockdown RT networks with cluster names. Top: T1 (January 1, 2020 - March 22, 2020). Bottom: T2 (March 23, 2020 - June 10, 2020). 134

Figure A 9: #Vaccines RT networks with cluster names. Top: T1 (January 1, 2020 - March 22, 2020). Bottom: T2 (March 23, 2020 - June 10, 2020). 135

List of abbreviations

- DSA – Digital Services Act
- ICT – Information and Communication Technology
- OB:SC – Oversight board supreme court metaphor
- RQ – Research question

Chapter 1 - Introduction

At latest since the 1990s, scholars have discussed the social implications of Internet-based information and communication technologies (ICTs) (DiMaggio et al., 2001). Bimber (1998) argued in his seminal work on "The Internet and Political Transformation" that early examples of Internet-based issue campaigns illustrated *"an acceleration of the process of identifying and mobilizing groups into political action, outside of some of the traditional constraints in interest group formation and structure. The results may be a political system in which issues develop and move more quickly because of the quicker cycle of mobilization and response and in which government officials increasingly hear from and respond to new kinds of groups - those without large, stable memberships or affiliations with established institutions"* (Bimber, 1998:158). This notion of accelerated pluralism resonates with developments regarding social media platforms as accelerators of socio-political fragmentation resulting in dissonant and disconnected public spheres (Pfetsch, 2018). Before reviewing a short history of the Internet and politics and how we got here, it is crucial to define the understanding of the critical terms of this dissertation.

Nick Srnicek (2014) famously described platforms as digital infrastructures facilitating interactions between two or more actors. The providers of platforms position themselves as intermediaries that bring together different groups or types of users (Srnicek, 2014:43). In contrast to transaction-based platforms like Amazon, social media platforms are based on advertisements and managing their audiences' attention. Dallas Smythe (1981) called this marketing-based value an audience commodity describing how companies commodify access to and attention of audiences. This dissertation focuses on Meta and Twitter as two of the largest social media platform companies. Regarding social media platforms, Facebook, as a platform for social networking, influences whom we communicate with within our social networks and enables tailored advertisements based on detailed user profiles. Whilst Twitter's business model also builds on earnings with advertisements, the platform has become a forum for public discourse. It is linked, or even predicates, media agendas because journalists pick up what's happening on Twitter or are at least more sensitive to issues on the platform. As such, Twitter has become a platform for public attention with high importance for news media and, thus, also politicians and political

communication. Concerning the continued consumption of traditional media like broadcast radio, news outlets and especially television, political communication today takes place in a hybrid media system (Chadwick et al., 2015; Chadwick, 2017). In this hybrid media system, the mass media logic of political communication and action intertwines with a new networked media logic (Klinger and Svensson, 2015). As such social media platforms substantially change the logic and environment of political communication.

Fragmentation has at least two meanings relevant to this dissertation project. 1) As a political science definition of the ongoing fragmentation of interest groups in politics (Bimber, 1998), and 2) as the fragmentation of the media system accelerated by commercialization and the Internet, and nowadays social media platforms (Mancini, 2013). Fragmentation relates to a process of change compared to the perceived homogeneity of the public sphere during the mass media age (Strömbäck, 2008). Habermas (1962/1991) famously described the public sphere as a space for deliberating public discourse between civil society and the political system. Correspondingly, social media platforms constitute digital public spheres allowing democratic deliberation (Dahlgren, 2005; Colleoni et al., 2014). More recently, however, Habermas (2022) describes the recent fragmentation and further transformation of the public sphere and states that a "mode of semi-public, fragmented and self-enclosed communication seems to be spreading among exclusive users of social media that is distorting their perception of the political public sphere as such" (Habermas, 2022:146). Whilst the original public-sphere reference highlighted discourses' deliberative function, Habermas highlights the contemporary risks of distortion and fragmentation of democracy by social media platforms.

During the mass media age, a limited choice of newspapers or television programs existed. However, these had a similar news agenda and, most importantly, were committed to journalistic norms. Nowadays, social media platforms and digital-born news formats fragment the news sphere and may have vastly varying news agendas, including channels that purposefully with an economic or political objective and report even non-factual or counterfactual information. For instance, Russia Today and Sputnik were Russian international news channels blocked in the European Union in 2022 after the onset of the Ukraine war. They functioned for years as a soft power tool and propaganda instrument for the Russian government (Yablokov, 2015).

Nevertheless, even neglecting foreign influences, the variation of news channels and unverified information significantly increased because of social media platforms on which citizens or news consumers choose channels by their liking or following depending on the specific platform's affordances (Theocharis et al., 2022). Social media platforms bundle news content, provide an algorithmic recommendation that tends to reinforce people's existing beliefs and attitudes, and provide mechanisms of selective reinforcement of attitudes via following on Twitter or Facebook groups. Additionally, they enable bottom-up and top-to-bottom communication and create new ways of citizen-initiated campaigning (Copeland and Römmele, 2014; Gibson, 2015).

Concerning the media and the news, the development of media technologies like radio broadcasting, television and the Internet each revolutionized how people consume information about politics and correspondingly locate themselves and their identity in an increasingly mediated reality (Couldry and Hepp, 2018). Moreover, each new medium lowered access barriers to political information. Nowadays, social platforms offer a sheer indefinite choice of information, news and entertainment freely accessible and recommended by algorithms, resulting in a passive attitude to searching for political information and news online (De Zúñiga and Diehl, 2019). Social platforms might contribute to these processes in various political arenas.

The reader might wonder how we got to this point and why social media platforms remained largely unregulated since their emergence in the mid-to-late 2000s. For most of the 2000s, the Internet and early-stage social media were celebrated for their unprecedented capabilities to connect, communicate and collaborate beyond geographical or legislative borders. More recently, the broader public perception of social media platforms shifted towards a more critical and risk-sensitive stance. Especially the revelations related to the Cambridge Analytica Scandal in 2018 and the company's potential influence on the results of the 2016 US Presidential Elections and the Brexit Referendum in 2016 marked a turning point (Klonick, 2020:2442). The reader might remember Mark Zuckerberg's hearing on the Cambridge Analytica scandal by the US Senate.

Additionally, after his election, former US President Donald Trump used social media, especially Twitter, to avoid directly communicating with traditional media outlets and

journalists. Facebook also played a role as a platform for hate and calls for violence before and during the Rohingya genocide in Myanmar starting in October 2016. In January 2021, the blocking and de-platforming of former President Donald Trump as a reaction to his involvement in the storm of his supporters on Capitol Hill in Washington DC after the 2020 US Presidential elections constituted a second shift and alarmed policymakers to consider the power that companies as moderators of their communication platforms hold. Even former German chancellor Angela Merkel expressed her concerns about the decision to ban Donald Trump from Meta's services and Twitter.¹ In 2021, the "Facebook files" revelations based on information leaked by former Facebook employee Frances Haugen illustrated how Meta's management repeatedly ignored warnings by internal researchers that its recommendation algorithms may fuel hate and harm young people on Facebook and Instagram.² At the centre of the problem remains that social media platforms were not primarily designed for political debates, but in order to maximize user engagement for advertising purposes. As the Facebook files revelations showed this misalignment of design and advertisement purposes and social and political uses can have harmful consequences for individual well-being, social cohesion and the quality of democracy.

This dissertation research is motivated by debates on social media platforms and seeks to understand better and explore in which political arenas social platforms affect democracy. The dissertation research and its studies explore the following research questions:

RQ1: To what extent are social media platforms accelerators of socio-political fragmentation?

RQ2: In the arena of political campaigns and elections, to what extent do social media platforms fragment political election campaigns?

RQ3: Regarding protest movements, what factors can result in fragmenting protest movements on social platforms?

RQ4: Considering democratic institutions, how might social platforms contribute to a fragmentation of democratic institutions?

¹ Deutsche Welle, <https://www.dw.com/en/angela-merkel-calls-trump-twitter-ban-problematic/a-56197684>, November 1, 2021, accessed January 19, 2023.

² Wikipedia, "2021 Facebook leak", https://en.wikipedia.org/wiki/2021_Facebook_leak, accessed April 1, 2023.

The author investigates the research questions in three arenas of political activity: political campaigns, protest movements and democratic institutions. These arenas are each subject to one of the three case studies. The first case focuses on political discourse on Twitter before the German Federal elections in 2017 and 2021. RQ2 is investigated in the arena of political campaigns, and the respective study explores online partisans' potential influence on political election campaigns on Twitter. The second case study examines RQ3 and focuses on citizen-initiated campaigns on Twitter in the UK during the Covid-19 pandemic. During the pandemic and the so-called lockdown in the UK, groups of Twitter users promoted conspiracy narratives and anti-science discourses. They formed protest movements that my co-author and I conceptualize as disinformed social movements. The messages spread by these groups are partly antisemitic or contain dangerous false information about medical treatments, and the findings imply that they can have fragmenting results for democracy by forming disinformed fringe groups. The third case study regards RQ4 in the arena of democratic institutions and inspects the US online news discourse on Meta's Oversight Board, mimicking and potentially eroding trust in democratic institutions due to the use of constitutional metaphors. These cases were selected since all three, Germany, the United Kingdom and the United States, are established democracies and face challenges regarding democratic fragmentation that become pertinent during times of crisis, such as during the Covid-19 pandemic or election campaigns as periods of increased political conflict.

This dissertation comprises five chapters and provides a framework and discussion of the three empirical studies. Chapter 2 states the background for the dissertation research, and reviews the theoretical and empirical literature on social media platforms and politics, as well as the potential challenges that social media platforms pose in the arenas of political campaigns, protest movements and democratic institutions. Chapter 3 consists of a single-authored study on the ideological polarisation of online debates on Twitter and the potential influence of partisans on political campaigns in the case of the German federal elections in 2017 and 2021. Chapter 4 is a co-authored study investigating conspiracy narratives on Twitter in the UK during the COVID-19 pandemic and interpreting these groups as disinformed social movements. Chapter 5 is a co-authored study that assesses the potentially harmful use of constitutional metaphors to describe private-governance initiatives such as Meta's oversight board. Chapter 6 discusses the main findings of the

dissertation research and concludes with an argument of the policy relevance for regulatory measures and research data access that are part of current EU regulatory initiatives such as the Digital Services Act. Finally, the author gives an outlook to further research on the topic and concludes with final remarks on potential technological challenges in the near future.

Chapter 2 – Background

At least since the 1990s, scholars have discussed how internet-based communication technologies (ICT) may affect politics and society. While some highlight the community-enhancing networking features of the Internet, mainly social media platforms, others focus on potential adverse effects such as the risk of rising populism. The following sections first discuss empirical research on social media platforms and risks of fragmentation for democracies (2.1.) and then focus on the three political arenas as organised political campaigns (2.2.), civic participation and protest (2.3.), and the relationship of social media platforms and democratic institutions (2.4.).

2.1. Social media platforms as fragmented communication environments

An abundance of empirical studies investigated social media platforms as communication environments. A recent systematic review by Lorenz-Spreen et al. (2023) indicates that most studies examining contexts in established democracies find adverse effects of social media concerning democratic quality. However, limited and unreliable research data access hampers scientific knowledge (Tromble, 2021). Besides, methods are a further challenge for social science researchers since the empirical investigation of these research problems requires new research methods in the social and political sciences, often summarized as computational social science (Lazer et al., 2020; Theocharis and Jungherr, 2021). These research approaches are based on digital trace data, e.g., from social platforms or web-tracking data, to assess associations of online behaviour and attitudinal or behavioural changes in the "offline" world. The following paragraphs introduce a selection of research studies on news consumption, political campaign communication and partisan dynamics on social platforms relevant to the selected arenas where social platforms may contribute to democratic fragmentation.

In 2019 a report by the Scientific Foresight Unit (STOA) of the European Parliament reviewed existing empirical studies on the relation between attitudinal polarisation and a heterogenization of news production and consumption. While the authors noted a thin empirical foundation, existing empirical evidence indicated no systematic association between news media changes and polarisation of attitudes and political opinion (Fletcher et al., 2019). Since then, the findings of several widely recognized studies resonated with the call for a nuanced assessment of digital media and politics. For instance, an empirical study

based on web browsing data found that instead of seeing a filter bubble effect, social media usage increased the heterogeneity of individuals' news media diets (Scharkow et al., 2020). However, when the news media diet contains hyper-partisan news media, this may have long-term effects on individuals' trust in mainstream media sources, highlighting the importance of trust in media and institutions (Guess et al., 2021). Another study emphasized that online news consumption and distribution by news recommender systems result in an increased frequency of news topic cycles and collective attention (Lorenz-Spreen et al., 2019). This increased frequency resonates with the notion of an acceleration of political processes and fragmentation of political issue campaigns.

With regards to communication by political elite like parties, government institutions and politicians, social media may result in a change of campaign messages, e.g., more negativity or increased negative campaigning, such as attacking political opponents instead of promoting their policy ideas and political beliefs (Baranowski et al., 2022; Klinger et al., 2022a). Professional politicians and political partisans use social media as interactive communication environments to support their messages and politicians or other individuals they ideologically agree with. Because algorithms up weigh content with a high number of interactions and viewers may give messages with high interaction numbers more importance, partisan activity on social media platforms is essential for a politician's visibility. From a researcher's perspective, group dynamics and partisan activity can be analysed by examining amplification mechanisms like retweeting on Twitter or liking on Facebook. Regarding visualizing and analysing these partisan amplification dynamics, social network analysis provides a powerful methodology that also enables the identification of partisan groups by their online behaviour (Conover et al., 2011; Conover et al., 2012). Online partisans behave strategically and seek to promote political actors they support, criticize (and sometimes harmfully attack) political adversaries and coordinate to take over or hashjack hashtags of other political groups (Bode et al., 2015; Darius and Stephany, 2019; Hadgu et al., 2013; Klinger et al., 2022b; Knüpfer et al., 2020). In election campaigns, this may result in citizen- or partisan-initiated campaigning (Copeland and Römmele, 2014; Gibson, 2015).

Empirical evidence indicates that increased social media use for political news consumption and political communication does not directly result in negative consequences for

democracy but alters ways of political participation (Theocharis et al., 2022). Other reviewed literature, however, indicates that partisan groups and high political fringe activity may significantly influence who and what opinions are visible on social platforms. Moreover, recommendation systems and content moderation matter for what is visible on social media platforms and create the environment for political communication on social media platforms. The following section briefly reviews selected literature on social media platforms as environments for political campaigning.

2.2. Social media platforms as environments for political election campaigns

Social platforms allow for top-down and bottom-up communication and, thus, technically collapse the classical demand and supply side division of politics, providing a guideline to think about political communication between political elites, government, and the electorate (Bennett and Iyengar, 2008). Before the Internet and Internet-based communication technologies, politicians needed to work with media gatekeepers such as television channels or news journalists. Nowadays, individuals like former US President Donald Trump use Twitter or other social media as primary communication channels to communicate with the public. Whilst individual behaviour and communication style also depends on politicians' political ideology and personality, there is a general move towards digital communication in election campaigns (Gibson et al., 2014).

Regarding digital campaigning, Twitter has become an essential communication tool during election campaigns. In contrast to Facebook, which allows targeted political advertisements and direct communication with groups of voters in the constituencies, Twitter is, in most countries, not as widely adopted as Facebook, and has banned political advertisements in 2019.³ Politicians use Facebook to communicate to voters directly. In contrast, Twitter is used to attract the attention of journalists and influence media agendas (Stier *et al.*, 2018). In order to attract attention, party and campaign hashtags play a central role in political communication on Twitter (Bruns and Burgess, 2011). Consequently, the first paper in Chapter 3 investigates the use of political party hashtags on Twitter in the 2017 and 2021 German Federal election campaigns. Here partisan supporters play a significant role in

³ Twitter, <https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html>

amplifying and engaging with campaign content, which also relates to social media as environments for citizen participation and activism, discussed in the following section.

2.3 Social media platforms as environments for civic participation and protest

Social platforms and political communication have a Janus-faced character regarding democratic quality and participation. On the one hand, they enable deliberation and can promote political participation. On the other hand, they may contribute to increasing social fragmentation, political polarisation and extremism. Regarding the academic debate, there has been a discussion of widening the understanding of political participation to include forms of digitally networked participation (Theocharis, 2015).

Vaccari and Valeriani (2022:66) define political participation as a "fluid set of repertoires of political action spanning across face-to-face and digitally enabled activities and aiming to exercise influence on politically relevant outcomes." Political participation may aim to influence 1) policy decisions, 2) the selection of public officials, or 3) other citizens' political preferences (Vaccari *et al.*, 2021). Considering this advanced understanding of political participation, creating or sharing political messages on online social platforms may serve one of these targets. Consequently, social media platforms provide political fora that are, in reference to Habermas (1991), often called public sphere-like spaces (Colleoni *et al.*, 2014). These accounts interpret ICTs and later social media as democracy-enhancing tools that may facilitate deliberative discourses and engage citizens to participate (Ferree *et al.*, 2002). Undoubtedly, social media platforms provide communication channels and public fora for political debates and mobilization. In contrast to organised political election campaigns, social movements and protest formation are characterized by ad-hoc and grass-root, bottom-up forms of organising (Bennett and Segerberg, 2013).

Regarding this organisation feature, social media are especially prominent in the literature on protest movements of the late 2000s and early 2010 (Theocharis *et al.*, 2015). Scholars distinguish social movements from protests by a higher degree of identification with the movement's aims, a principle of opposition and the totality of beliefs (Touraine, 1995). Though, if these beliefs of identification cluster around counterfactual claims and disinformation, as politically motivated false information, like in conspiracy narratives, movements can be classified as "dark" social movements (Sternisko *et al.*, 2020). This

classification of the dark side normatively describes the ambiguity of social media platforms as potential facilitators of "good" democratic participation and the "dark" side of harmful behaviour, radicalisation, extreme and fringe political groups or conspiracy narratives that were highly present during the Covid-19 pandemic. Thus, the second study in Chapter 4 focuses on the debate on Twitter during the COVID-19 pandemic and conceptualises communities that proliferate conspiracy narratives as disinformed social movements.

2.4. Social media platforms and democratic institutions

The literature reviewed in the prior sections described how social platforms as environments for political communication affect organised and informal political communication, such as political campaigns and the formation of social movements online. Besides being in the power of social media as a communication environment, social platforms have, in recent years, created governance bodies that copy or mimic democratic institutions. Social platform companies are owners of their platforms as quasi- "public" online spaces, which gives them significant power over political communication, especially during election campaigns, when politicians, political parties, and third-party actors must reach potential voters with political messages. For too long, the gatekeeping power that comes with control of social platforms as communication and information infrastructures was not fully acknowledged, and platforms were instead understood as intermediaries (Schulz, 2018). This notion of being an intermediary defines social platforms as sole providers of a technical infrastructure but not as responsible entities for what happens on the platforms. The companies were not legally liable for the content shared and created on their platforms (Gillespie, 2018).

Whilst platform companies took a hesitant approach to moderating toxic and harmful speech, they reacted to public pressure⁴ as a reaction to major scandals and regulations⁵ by governments (Klonick, 2020). Since then, platform companies have taken a more active role in moderating harmful or illegal activities on their platforms. Moderation relates to the "governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse" (Grimmelmann, 2015:47). Firstly, this refers to social media

⁴ Scandals, such as the Cambridge Analytica issue or Facebook's role in hosting hate that fueled the genocide in Myanmar and other cases constitute a major PR problem for platform corporations (Klonick, 2020).

⁵ E.g., the Network Enforcement Act holds platforms accountable for copyright infringements and other sorts of illegal content created and shared by using their services (Gorwa, 2021).

platforms' function as sorting content and connections via recommendation algorithms to facilitate exchange and engagement. Secondly, preventing abuse means filtering content infringing on the platform's community guidelines or legislation (Grimmelmann, 2015).

The latter has taken centre stage in public debates since the 2020 US presidential election, the storm on the US Capitol Hill by Trump supporters, and the consequential deletion of Donald Trump's harmful messages and his accounts on major social platforms (Facebook, Instagram and Twitter). Platforms' power consists of two forms of gatekeeping, gatekeeping of information and gatekeeping of access. Gatekeeping of information refers to the recommendation of content for millions or billions of users, and voters, ultimately deciding who is more visible and more likely to succeed in an election. Gatekeeping of access decides who may have access to the platform and its community, consisting of hundreds of millions or billions of global users in the case of Twitter and Meta.

The Trump case happened within the first 100 days of the Meta's Oversight Board, launched on October 22nd, 2020, that decided whether Trump's accounts would remain blocked on Meta's services. The decision to confirm Trump's ban from the platform marks one of the first major decisions of the Oversight Board and constituted a test case for its procedures, decisional transparency and public reception of decisions taken. Moreover, it underlined the power of social platforms and alarmed policymakers worldwide. With the perceived policy puzzle of democratic regulation of speech or speech moderation, many policymakers focus on co-regulation approaches that describe the cooperation of regulators and, in this instance, social media platform corporations. Since policymakers seek to avoid directly accessing data and lack insight into algorithmic systems' workings, there are persistent knowledge gaps between policymakers and platforms (Cowls et al., 2020). At the same time, social media platform corporations, especially Meta, have created governance initiatives to deal with content moderation questions like redress mechanisms to account for mistakes or controversial decisions. The potentially most debated governance institution of this sort is the Oversight Board by Meta which my co-authors and I examine in the third study in Chapter 5. The following chapters comprise the three published studies, followed by a discussion of the findings in Chapter 6.

Chapter 3 – Who polarises Twitter? Ideological polarisation, partisan groups and strategic networked campaigning on Twitter during the 2017 and 2021 German federal elections 'Bundestagswahlen'

Abstract⁶

Political campaign activities are increasingly digital. A crucial part of digital campaigning is communication efforts on social media platforms. As a forum for political discourse and political communication, parties and candidates on Twitter share public messages and aim to attract media attention and persuade voters. Party or prominent candidate hashtags are a central element of the campaign communication strategy since journalists and citizens search for these hashtags to follow the current debate concerning the hashed party or political candidate. Political elites and partisans use social media strategically, e.g., to link their messages to a broader debate, increase the visibility of messages, criticise other parties or take over their hashtags (hashjacking). This study investigates the cases of the most recent German Federal Elections 'Bundestagswahlen 2021' and the 2017 German Federal elections. The investigation (1) identifies communities of partisans in retweet networks in order to analyse the polarisation of the most prominent hashtags of parties, 2) assesses the political behaviour by partisan groups that amplify messages by political elites in these party networks, and 3) examines the polarisation and strategic behaviour of the identified partisan groups in the broader election hashtag debates using #BTW17 and #BTW21 as the prominent hashtags of the 2017 and 2021 elections. While in 2017, the far-right party 'Alternative für Deutschland' (AfD) and its partisans are in an isolated community, in 2021, they are part of the same community as the official party accounts of established conservative and liberal parties. This broader polarisation may indicate changes in the political ideology of these actors. While the overall activity of political elites and partisans increased between 2017 and 2021, AfD politicians and partisans are more likely to use other party hashtags, which resulted in the polarisation of the observed parts of the German political Twitter-sphere. While in 2017, the AfD polarised German Twitter, 2021 shows a broader division along the classical left-right divide.

⁶ Darius P (2022) "Who Polarizes Twitter? Ideological Polarisation, Partisan Groups and Strategic Networked Campaigning on Twitter during the 2017 and 2021 German Federal Elections 'Bundestagswahlen.'" *Social Network Analysis and Mining* 12(1). <https://doi.org/10.1007/s13278-022-00958-w>.

3.1. Introduction

Social media communication as digital campaigning is essential to political election campaigns (Gibson and Römmele, 2004; Gibson et al., 2014; Lilleker et al., 2017). The COVID-19 pandemic further increased the importance of digital campaign tools, e.g., for the 2021 German federal elections, the 'Bundestagswahlen 2021'. For political candidates, politicians and parties, journalists, party supporters, and politically-interested citizens, Twitter is an essential platform for public debates. These public debates on Twitter organise around hashtags that link users' tweets to a particular topic. The acronyms of the German word for the federal elections' Bundestagswahlen' have been among the most frequently used hashtags⁷ (#BTW17 and #BTW21) to connect campaign tweets to the general election issue during the election cycles.

While prior work monitored the frequency of political party and candidate hashtags in German election campaigns on Twitter, previous analyses do not account for the strategic use of hashtags regarding political opponents (Stier et al., 2018a). Several scholars named this strategy hashtag-hijacking (Hadgu et al., 2013; Van Dam and Tan, 2016, Xanthopoulos et al., 2016) or, in short, 'hashjacking' (Bode et al., 2015; Darius and Stephany, 2019). In the German context, prior work showed that the German far-right party 'Alternative für Deutschland' (AfD) used hashjacking as a political communication strategy (Darius and Stephany, 2019). Additionally, politicians and partisans shared vaccination-sceptic messages during the COVID-19 pandemic and used hashtags strategically to influence the broader debate on the pandemic (Darius and Stephany, 2022). This study extends prior work by investigating the strategic hashtag use by political elites and partisans during the final phase of the 2017 and 2021 German federal election campaigns. Election campaigning comprises a unique opportunity to observe political communication efforts by parties and candidates. Correspondingly, the study examines strategic hashtag use of political party hashtags and compares the polarisation of these hashtag discourses during the 2017 and 2021 election campaigns.

The study proceeds as follows: the background (3.2.) section briefly outlines different historical phases of political campaigning and the role of the media (3.2.1). After that, it

⁷ Based on the literature and searches with a social media tool

discusses the use and functions of digital campaigning methods focusing on political communication on social media (3.2.2.). Then it discusses online polarisation and strategic political behaviour online, such as hashjacking (strategic hashtag use of other political parties' hashtags), which comprises a strategy predominantly used by right-wing populists/far-right actors (3.2.3.). Section (3.3.) formulates the research hypotheses based on existing literature and prior work. Section 4 introduces the research design based on data collection (3.4.1.), network approach (3.4.2.), and explanation of the measures (3.4.3.). Successively, the following section (3.5.) presents the analysis and results, followed by the discussion section (3.6.) that reviews the findings (3.6.1.), limitations of the study (3.6.2.), and concludes with a reflection on the broader meanings of the findings for political campaigns and society.

3.2. Background

Studying online polarisation and strategic communication as digital campaigning speaks to political science, media studies, network science, and interdisciplinary social media research. While this section introduces literature on Twitter as a campaigning tool and forum for public debate and reflects on strategic political behaviour on Twitter, it also aims to provide an interdisciplinary readership with a short introduction to the development of political campaigning and the media. Correspondingly, this section first introduces a short history of political campaigning and then reviews the most relevant literature on the strategic use of social media in political campaigning and communication.

3.2.1. Political campaigning and the media as a process

In order to grasp changes in digital political campaigning, it is crucial to contextualise the historical development and different phases of political campaigning in the past two centuries. Most of the existing literature distinguishes three or four eras of political campaigning that correspond with broader technological and societal developments (Strömbäck, 2008; Hjarvard, 2013; Esser and Strömbäck, 2014; Couldry and Hepp, 2017; Hepp, 2019). It is common to differentiate between premodern, modern, and postmodern periods of political campaigning and their typical campaign practices (Schmitt-Beck and Farrell, 2002; Norris, 2004). More recently, scholars argued that campaigning moved into a fourth phase due to the influence of digitalisation on the media and politics (Strömbäck,

2008; Blumler, 2013; Magin et al., 2017; Römmele and Schneidmesser, 2016; Römmele and Gibson, 2020). The Democratic campaign Obama in 2008 is often seen as an early exemplary case, laying out strategies of how to implement web-based communication and "big data" in a political campaign (Gueorgieva, 2008; Lilleker & Jackson, 2011; Gerodimos and Justinussen, 2015).

The catalysation of trends fuelled by technological change distinguishes the fourth era from prior eras, leading to a new form of data-driven campaigning (Römmele and Gibson, 2020). With the growing adoption of social media platforms and the Internet, web-based or digital political campaigning has spread globally (Gibson et al., 2014; Dimitrova and Matthes, 2018). Due to the wide use of social media in politics and the related power of platforms as information gatekeepers, social platform architecture and governance become crucial factors for the quality and legitimacy of democracy (Gillespie, 2018; Cowls et al., 2022; Stockmann, 2022). This increased importance of digital technologies and data analysis in campaign operations and organisation is reflected in the transition of political communication from mass media-based to a more direct, interactive, and networked type of communication with the electorate, targeting of campaign messages, and an increasingly international dimension of political campaigns with interferences by foreign actors. Moreover, a qualitative separation of digital and data-driven campaigns may occur into a rational-scientific approach on the one hand and an emotionalised subversive approach to campaigning on the other that may benefit populist parties (Römmele and Gibson, 2020). Social media provides a platform for public debate and communication beyond traditional media like print, television, or the radio. The following section will discuss Twitter's role as a forum for political debate and a political communication tool in election campaigns.

3.2.2. Twitter as a political communication tool and forum for public discourse

From the beginning of social media use in politics, academics have asked whether it would enhance direct communication between politicians, journalists, and citizens. This direct communication could indicate a more participative or, to quote Habermas, a more "public sphere-like" democratic space (Habermas, 1991; Ferree et al., 2002; Dahlgren, 2005;

Colleoni, 2014; Ekman and Widholm, 2015; Rau und Stier, 2019)⁸. In several studies, there was no consistent indication of increased communication between political 'elites' and citizens, but politicians and journalists communicate primarily with each other in public (Grant et al., 2010; Verweij, 2012; Nielsen and Vaccari, 2013; Oelsner and Heimrich, 2015; Jensen, 2017). Within these online spaces, candidates, political representatives, and parties communicate with the public, and journalists and citizens interact with these messages (Gibson, 2015). Concerning the capacities for campaigning and the central questions of which target groups political communication reaches, social media differ significantly due to mediation effects based on their varying structures, communication mechanisms, and user audiences (Bossetta et al., 2018, Stier et al., 2018b, Bronstein et al., 2018). In Germany, for instance, Facebook reaches a much broader demographic group than Twitter⁹ and parties may leverage social media to reach young voters (Copeland and Römmele, 2014). On Twitter, however, elite actors such as politicians react to trends such as rising hashtag debates and aim to influence media reporting, meaning that journalists cover their political messages, which has multiplier effects in reaching the public (Larsson, 2015; Kreiss, 2018).

In general, studies question whether social media lead to interactions between voters and politicians (Graham et al., 2013; Oelsner and Heimrich, 2015; Caton et al., 2015). During campaign periods, politicians, such as members of parliament in Germany, tend to use Twitter more actively and differently to non-campaigning times. For instance, they refer more often to the broader election topics or hashtags instead of sharing content from their personal lives (Nuernbergk and Conrad, 2016).

With regards to election campaigns, social media play a multifaceted role. Social media as digital tools for political elites affect political campaigning practices mainly by their four main functions 1) organisational structures and work routines, 2) presence in online information spaces, 3) support in resource collection and allocation, and 4) symbolic uses in the sense of political marketing (Jungherr, 2016). Symbolic uses and presence in online

⁸ While the political discourse on these platforms is public or semi-public, e.g., in large Facebook groups, the online spaces are privately owned and algorithmically governed by social platform corporations (Gillespie, 2017; Katzenbach and Ulbricht, 2019).

⁹ This representative online study by German public broadcasters ARD and ZDF summarises results of online media use for different demographics <https://www.ard-zdf-onlinestudie.de/files>

information have been the focus of scholars working on populism, extremism, and media research. For extreme parties and their political narratives, social media offer additional channels for political communication in which extreme political actors do not need to follow the values and norms of traditional mass media and are thus able to spread their respective ideologies (Engesser et al., 2017). The ideology of right-wing populist parties builds on the rhetoric construction of (1) anti-elitism/establishment, (2) anti-migration, and (3) anti-Muslim stances. Notably, these three pillars of right-wing populist rhetoric and policies polarise voters against something and, in particular, against certain groups of people (Mudde, 2004; Mudde and Kaltwasser, 2013). Social platforms and messengers are desirable for right-wing populist and radical-right parties as political challengers who often have a hostile attitude towards established media and sometimes limited access to traditional media channels (Engesser et al., 2017; Jungherr et al., 2019; Koc-Michalska and Klinger, 2021). Thus, right-wing populist actors and movements have benefited disproportionately from the emergence of social media since they can circumvent traditional media and communicate directly to their target audiences (Stier et al., 2018b; Jacobs, 2018). Besides, direct contact with political actors and the represented ideologies enables the self-socialisation of citizens into right-wing populist beliefs and worldviews (Krämer, 2017; Schumann et al., 2021).

Furthermore, social media also provides an opportunity for top-down leadership claims for populist parties and politicians. Social media provided additional channels for communication with and between political elites, partisans, and the electorate. Hashjacking of political adversaries and the strategic hashtag use of broader discourses increase the representation of populist messages on social media (Darius and Stephany, 2019, 2022). What remains unclear is whether strategic hashtag use and hashjacking also increase online polarisation during campaign times.

3.2.3. Online polarisation as a strategy?

There has been conflicting evidence on the relationships between social media and socio-political polarisation (Garimella and Weber, 2017; Bail, 2018, 2021). It is yet unclear whether social media 1) might reduce polarisation by enabling access to a more diverse set of information and news (Stier et al., 2021) or 2) might increase or accelerate polarisation

tendencies by algorithmic enforcement of opinions, e.g., by the formation of so-called echo-chambers, or 3) whether online polarisation on social media solely reflects differences in the social and political (offline) world¹⁰. Even the assumption that exposure to a more diverse set of information and political views may reduce polarisation is questionable or even counterfactual (Bail et al., 2018). Besides this unclarity about interpretations, various definitions of offline polarisation hamper academic consensus (Bramson et al., 2017; Tucker et al., 2018). Regarding polarisation, social media provide a chance to analyze political behaviour by elites and partisans (Gil de Zúñiga et al., 2020). For instance, individuals self-sort into ideologically aligned communities by their retweeting behaviour when using political party hashtags (Conover 2011; Conover 2012). This sorting happens because most users retweet in support of messages, especially regarding political hashtags and topics (Metaxas et al., 2015). Consequently, strategic hashtag use by politicians and users' retweeting behaviour may result in polarised political hashtag debates in which communities reflect camps with contrasting political ideologies.

Hashtags enable Twitter users to interact in so-called ad-hoc publics outside their follower networks¹¹ and link their tweets to a broader conversation (Bruns and Burgess, 2011). Consequently, hashtags allow for ad-hoc (political) debate on Twitter and are used frequently by politicians and journalists (Enli and Simonsen, 2018). Political elites and partisan groups employ strategic behaviour such as retweeting or hashjacking (using political opponents' hashtags) to influence these public debates. Usually, elite actors such as politicians or social media influencers issue messages using hashtags strategically, and partisans amplify their messages by sometimes excessive retweeting. These strategic expressions may increase online polarisation caused by political elites and partisans' self-sorting into ideological camps (Conover, 2011; Garimella and Weber, 2017).

Political elites strategically use party and campaign hashtags to increase their visibility and support their party in election campaigns. In contrast to traditional political communication

¹⁰ The differentiation between the online and offline world employed here is not an analytical, but rather a communicative distinction since digital media have become deeply entangled in our everyday offline practices.

¹¹ Twitter started displaying content from outside the follower networks in users' home timelines to break up so-called bubbles. <https://help.twitter.com/en/using-twitter/twitter-timeline>

via mass media, this constitutes a new form of so-called networked campaigning. Networked campaigning is characterised by the increasing importance of networked communication logic and logic of connective action (Bennett and Segerberg, 2013; Klinger and Svensson, 2015). These new logics result in equally new forms of strategic political communication behaviour. Two forms of strategic political online behaviour on Twitter that political elites and partisan groups frequently exercise are 1) "retweeting" (amplifying someone else's message) and 2) "hashjacking" (using hashtags of political opponents) as a particular form of hashtag use to take over the hashtag of political opponents (Bode et al., 2015). Regarding retweeting, in political contexts, many users only retweet messages they support, which results in ideologically aligned groups and polarised retweet networks (Conover et al., 2011). These partisan groups coordinate to use political opponents' hashtags, leading to spontaneous jumps in the polarity of hashtags or even polarising hashtags in the long term (Hadgu et al., 2013; Weber et al., 2013). In the case of Germany, hashjacking was used strategically by the far-right party AfD as a polarisation strategy (Darius and Stephany, 2019) and a disinformation strategy during the COVID-19 pandemic (Darius and Stephany, 2022). In practice, hashtag hijacking can be challenging to distinguish from strategic hashtag use, which connects individual messages to broader political debates. In the German case, however, far-right actors were much more likely to use other party hashtags, e.g., to hashjack, or link to broader COVID-19-related debates than partisans of other parties, which indicated a coordinated and strategic use by AfD politicians amplified by partisans (Darius and Stephany, 2022). This isolation of the far-right partisans differs from other partisans or party supporters in one large community with other parties and journalists. Beyond national politics, partisan groups, such as far-right partisans, may also coordinate transnationally with the aim of hashjacking hashtags of social movements such as #MeToo (Sorice, 2018; Knüpfer et al., 2020).

Regarding social movements, Twitter may also facilitate the spread of conspiracy narratives and enable the mobilisation of disinformed social movements, e.g., during the COVID-19 pandemic (Darius and Urquhart, 2021). While political elites, such as official politicians or political influencers, create content, partisans drive the dynamic by amplifying messages. This mechanism lies at the centre of the analysis of strategic hashtag use and provides a fundament for formulating the research questions in the following section.

3.3. Research questions

While prior work investigated the role of strategic hashtag use and 'hashjacking' as a strategy used predominantly by far-right actors in the German Twittersphere, this study explores the use of party, chancellor candidate, and broader election debate hashtags during the election campaigns. During campaign times, political debates intensify, which may result in hashtag debates that are more polarised in contrast to periods outside election cycles. Besides, political elites, such as parties, members of parliament or election candidates, and political partisans have higher incentives to behave actively on social media and criticise political adversaries for gaining media attention or persuading undecided voters. Revisiting prior work on 'hashjacking' (Darius and Stephany, 2019; 2022) and Twitter use in (German) political campaigns (Stier et al., 2018a) and based on the presented literature on 'hashjacking' and strategic hashtag use, the study formulates the following research questions. It is worth noting that using a party hashtag when politically opposed is not automatically hashjacking. However, messages are often issued strategically by politicians and political influencers. Thus, only if there is a significant degree of coordination, as expressed in a higher presence and likelihood of partisans to use other parties' hashtags, should this be understood as hashjacking.

(RQ1): Did AfD partisans aim to hashjack other parties' hashtags during the election campaigns in 2017 and 2021?

(RQ2): Was the average activity of partisan communities around the party and election hashtags higher in 2021 than during the 2017 federal election campaigns?

(RQ3a): Do AfD partisans appear as an isolated community in the #BTW17 retweet networks?

(RQ3b): Do AfD partisans appear as an isolated community in the #BTW21 retweet network?

(RQ4): Are right-wing partisans of the AfD more likely to engage in the macro-debate #BTW17 and #BTW21 hashtags than partisans of other parties?

3.4. Research design

This section elaborates on the data collection (4.1), introduces the network approach and community detection of partisan groups as an assessment of polarisation (4.2), and presents the measurement approach (4.3) to further examine online polarisation and strategic political online behaviour during the 2017 and 2021 German federal election campaigns. Figure 3.1 illustrates the pipeline of the analysis.

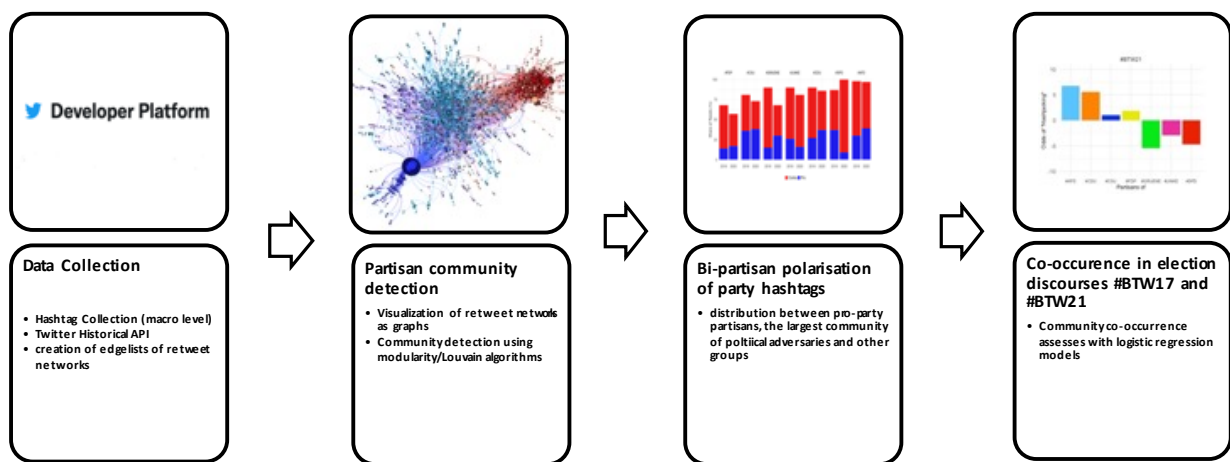


Figure 1: Pipeline of the analytical approach to assess the polarisation of hashtags within and between hashtag discourses

3.4.1. Data collection

The study collects Twitter data by accessing Twitter's application programming interface (API) for academics.¹² Twitter's historical API¹³ access allows the retrospective collection of user timelines or tweets of hashtags. For two reasons, hashtags are the macro-level selection criteria (Weller et al., 2013). First, party hashtags are important campaign goals used by politicians, supporters, journalists, and the main party accounts. Secondly, journalists and citizens might use Twitter hashtags to inform themselves about the latest news on the party. This information function may consciously or unconsciously serve as an indicator of public support for that party. This online debate is not representative but may have significant consequences, e.g., influencing media reporting. Consequently, the collection focuses on party hashtags for all parties that entered the federal

¹² Twitter, <https://developer.twitter.com/en/products/twitter-api/academic-research>

parliament (#AfD, #CDU, #CSU, #FDP, #GRUENE, #LINKE, #SPD) plus candidate hashtags as the last names of the three top candidates (#Baerbock, #Laschet, #Scholz). Additionally, the study analyses the broader debates on the elections represented by the most used hashtags regarding the Federal elections in 2017 and 2021 are #BTW17 and #BTW21 as the acronyms of the German word 'Bundestagswahlen.'

The observation period is the final week before the elections in 2017 and 2021, in which we expect intensified campaigning efforts by parties, candidates, and supporters. The study first identifies partisan communities and hashjacking efforts using party hashtags during the two observation periods and, second, investigates the role of these partisan groups within the broader political debate on the elections represented by the #BTW17 and #BTW21.

3.4.2. Network approaches for the detection of partisan communities

Political networks may represent many different relations, such as parliamentary co-sponsorship (Fowler, 2006), coalition formation (Maulana et al., 2022), organisational ties (Hafner-Burton et al., 2009), or political news diffusion on social platforms (Grinberg et al., 2016). In social network analysis, nodes represent individuals, and edges indicate relations or interactions. When aggregated, this relational data enables the graphical visualisation and statistical analysis of the structure of networks (Wassermann and Faust, 1994; Carrington et al., 2005; Scott, 2017). A network approach is favorable for the structural analysis of social media since communication on most platforms is networked by design. On Twitter, for instance, by linking to other users via @mentions, retweeting, or linking messages to a topic using hashtags (#). Retweeting creates a link (edge) between two accounts (nodes), constituting a network. These retweet networks often cluster into multiple communities, and for political hashtags, these communities may have different political party affiliations or ideologies (Conover, 2012). Thus, retweet networks provide the chance to assess ideological alignment and opinion leaders within communities where people self-sort their retweeting behavior (Conover, 2012; Bruns et al., 2016). This self-sorting into different ideologically aligned communities occurs because most users retweet in support of messages they ideologically agree with (Boyd et al., 2010; Metaxas et al., 2015). Moreover, users adopt retweets quicker than using hashtags in individual tweets

(Oliveira et al., 2021). Therefore, the analysis focuses on the retweeting networks of the chosen hashtags to identify partisan communities and assess the polarisation of the political debates on Twitter.

The analysis assumes that the network consists of two major partisan clusters. The more significant the proportion of accounts in these clusters, the more polarised is the network. In prior work on the German context, partisans of the far-right AfD constituted an isolated community in several political party hashtags (Darius and Stephany, 2019) and Covid-19-related hashtags (Darius and Stephany, 2021) retweet networks. In contrast, partisans and politicians of all other parties gather in a large community with journalists and media outlets. Consequently, AfD partisans, as supporters of a reasonably new anti-establishment party, have polarised the political discourse on Twitter as represented by retweet networks of common hashtags.

The analysis builds on the networked structure and the visualisation of retweet networks in Gephi using the Force2 layout algorithm (Bastian et al., 2009; Jacomy et al., 2014). In the first step of the analysis, the modularity-based (community detection) algorithm assigns the nodes to different communities based on the structural properties of the network graph (Newman, 2006; Blondel et al., 2008; Fortunato, 2010). Being retweeted is highly unequally distributed (Bild et al., 2015). Therefore, a qualitative content analysis of the 30 most retweeted accounts in each party network makes sense of the clustering (White and Marsh, 2006; Mayring, 2014; Krippendorff, 2018). Table A2 and Table A3 illustrate an example of the analysis of the 30 most retweeted accounts in #BTW17 and #BTW21. To account for the skewed distribution of being retweeted, the study uses a log-transformation that works for most social network data and also produces acceptable results for social media data (Broido and Clauset, 2019).

Partisan communities often center around official party accounts such as @AfD or @dieLinke (or the other party accounts). In prior studies, a closely connected far-right partisan community, whose activity was much higher than that of other communities, formed around official AfD accounts and amplified their political messages by retweeting (Darius and Stephany, 2019). Figure A1 illustrates this hashjacking mechanism. This pro-/contra-polarisation of each party retweet network and assigned community memberships

by the Louvain algorithm enables partisan groups' identification¹⁴. In a further step, the analysis assesses the occurrence of these groups in the broader debates (here #BTW17 and #BTW21).

3.4.3. Measurement of polarisation and strategic hashtag use

The investigation of the research questions builds on a network approach that enables the visualisation and analysis of the structure and identifies partisan communities in the party retweet networks via the Louvain community detection algorithm. The study uses the terms 'community,' and 'cluster' interchangeably since communities of politicians and partisans appear as clusters in the network structure. In party hashtags, the proportion of AfD partisans to the pro-party community indicates the extent of hashjacking (RQ1). The analysis measures the average activity of partisan groups (RQ2) by the weighted outdegree retweet network of nodes in the pro-party partisan community within each party hashtag. The weighted outdegree, also accounting for the number of reciprocated edges, signifies the number of times an account retweeted another account. In order to examine (RQ3a) and (RQ3b), the study assesses the assumed isolation of the AfD partisan cluster in two steps. At first, the investigation of the retweet network topology indicates the networks' potential polarisation between partisan communities, and second, a qualitative content analysis of the most retweeted accounts (with the highest in-degree) indicates the political ideology of the communities.

The analysis proceeds with logistic regression models to test each partisan group's binary likelihood of co-occurrence in one of the two main clusters on the broader #BTW17 and #BTW21 debates. The response (independent variable) of being in one of the two largest communities in #BTW17 and #BTW21 is binary since most nodes are assigned to these two communities (see Table A1).

$$y_i = \begin{cases} 1, & \text{if the } i\text{th node is co-occurring in the right-wing partisan cluster} \\ 0, & \text{otherwise} \end{cases}$$

¹⁴ To allow reproducibility of the study, network edge lists and anonymised data on network measures is publicly accessible on the author's GitHub profile: https://github.com/philippdarius/SNAM_Darius_2022

The realisation of Y_i (community membership in #BTW17 and #BTW21 retweet networks) y_i is random and binary (0,1) with the probabilities π_i and $1 - \pi_i$. Assuming no ideological sorting, observations within each partisan group should have the same probability of co-occurring in the right-wing partisan community in #BTW17 and #BTW21. The distribution of Y_i is binomial with the parameters π_i and n_i ; $Y_i \sim B(n_i, \pi_i)$. The log-odds are calculated as: $\eta_i = \text{logit}(\pi_i) = \log \frac{\pi_i}{1 - \pi_i}$ and the odds for co-occurring in the partisan cluster (in the party retweet networks) and the right-wing partisan community in #BTW17 and #BTW21.

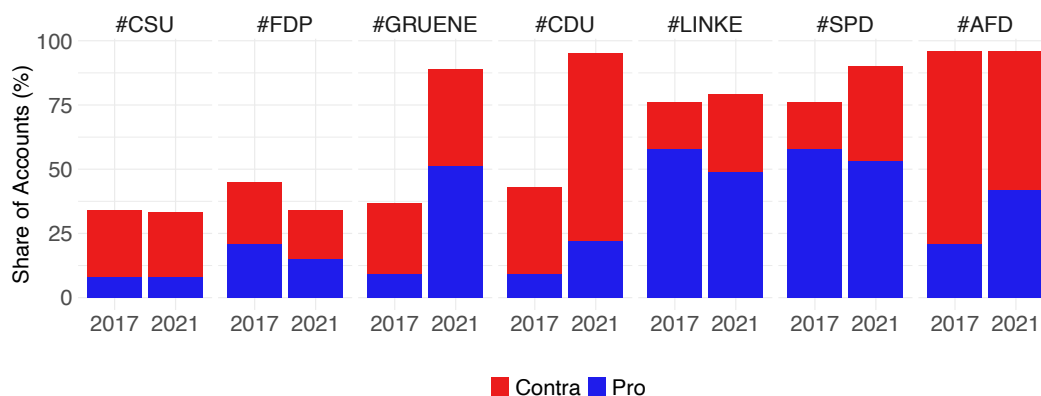
The logistic regression models assess cross-cluster homophily between partisan clusters and an identified right-wing community in the #BTW17 and #BTW21 retweet networks (RQ4). The resulting odds as outputs of the logistic regression models indicate how likely partisan groups from the party networks co-occur in the more extensive retweet networks of #BTW17 and #BTW21. A high likelihood of partisan co-occurrence in clusters of the broader electoral debate indicates the ideological closeness of these groups. Similar directions of the odds indicate membership in the same significant clusters. If partisans of one party are isolated in one community, as hypothesised for AfD partisans in (RQ3a) and (RQ3b), they should have the opposite odds direction as all other parties.

3.5. Results

This section presents the results of the analysis. The first part compares German political party hashtags polarisation between the two election periods in 2017 and 2021. The community detection identifies a pro-partisan cluster and a contra cluster as the largest community with highly retweeted members of other parties. Then, the analysis compares the activity of different partisan groups. After that, the focus lies on the structural assessment of the large-scale hashtags #BTW17 and #BTW21 and investigates the likelihood of (strategic) hashtag use of different partisan communities. The main finding is a differently polarised electoral debate on Twitter between 2017 and 2021. While the far-right party AfD was a segregated community in 2017, in 2021, the AfD is in the same cluster as the center-right, conservative parties (CDU and CSU) or the liberal party (FDP). Section (6.) further discusses this central finding and the findings regarding the individual hypotheses.

3.5.1. Partisan polarisation of party hashtags

Political hashtag debates often polarise into two or more communities. Prior research has shown that a far-right cluster led by official accounts of the AfD polarised German party hashtags (Morstatter et al., 2018; Darius and Stephany, 2019). Most of the observed retweet networks of party hashtags distribute into two central clusters, 1) the AfD community and 2) another cluster containing major news outlets and politicians of other parties. For the observation periods in 2017 and 2021, the modularity-based Louvain algorithm identifies these communities, and then a qualitative content analysis assesses the 30 most retweeted accounts. This way, AfD communities and the partisan community of the other party are identified. After applying a community detection algorithm, the analysis compares the identified communities with the share of accounts of the overall network. Figure 2 displays the proportions of the pro-party partisans and contra clusters (largest community containing partisans of other parties) for the selected hashtags of major German parties. The polarisation between AfD partisan and other partisan communities has increased for several hashtags, especially for the Greens and the CDU. Moreover, Table A1 summarises the cluster proportions of the three most significant clusters in the retweet networks during the observation periods in 2017 and 2021.



Darius, P. (2022) Partisan polarization of party hashtags

Figure 2: Column chart of pro-partisans for each party and contra-communities as members of the largest community consisting of other partisans

The only remarkable difference is the growth of the two major clusters for #GRUENE and #CDU, caused by the AfD partisan community (contra) and their partisan community's growth in the Greens case. Except for the slight decrease in bi-polar polarisation for #CSU and #FDP, there is an overall tendency of higher polarisation between the two largest

partisan communities, which relates to RQ1. The results indicate a higher level of bi-polar polarisation for party hashtags in 2021.

3.5.2. Behaviour of partisan communities

While RQ1 concerned the distribution of accounts in two partisan communities, it does not account for the activity of partisan clusters. The activity of the partisan cluster is subject to RQ2, which raises the question of whether the average activity of partisan communities was higher in the 2021 elections than in the 2017 elections. Figure 3 shows the distribution of log-transformed partisan activity (accounts that retweet) and network elites (accounts that are most often retweeted by others) for partisan communities in each retweet network.

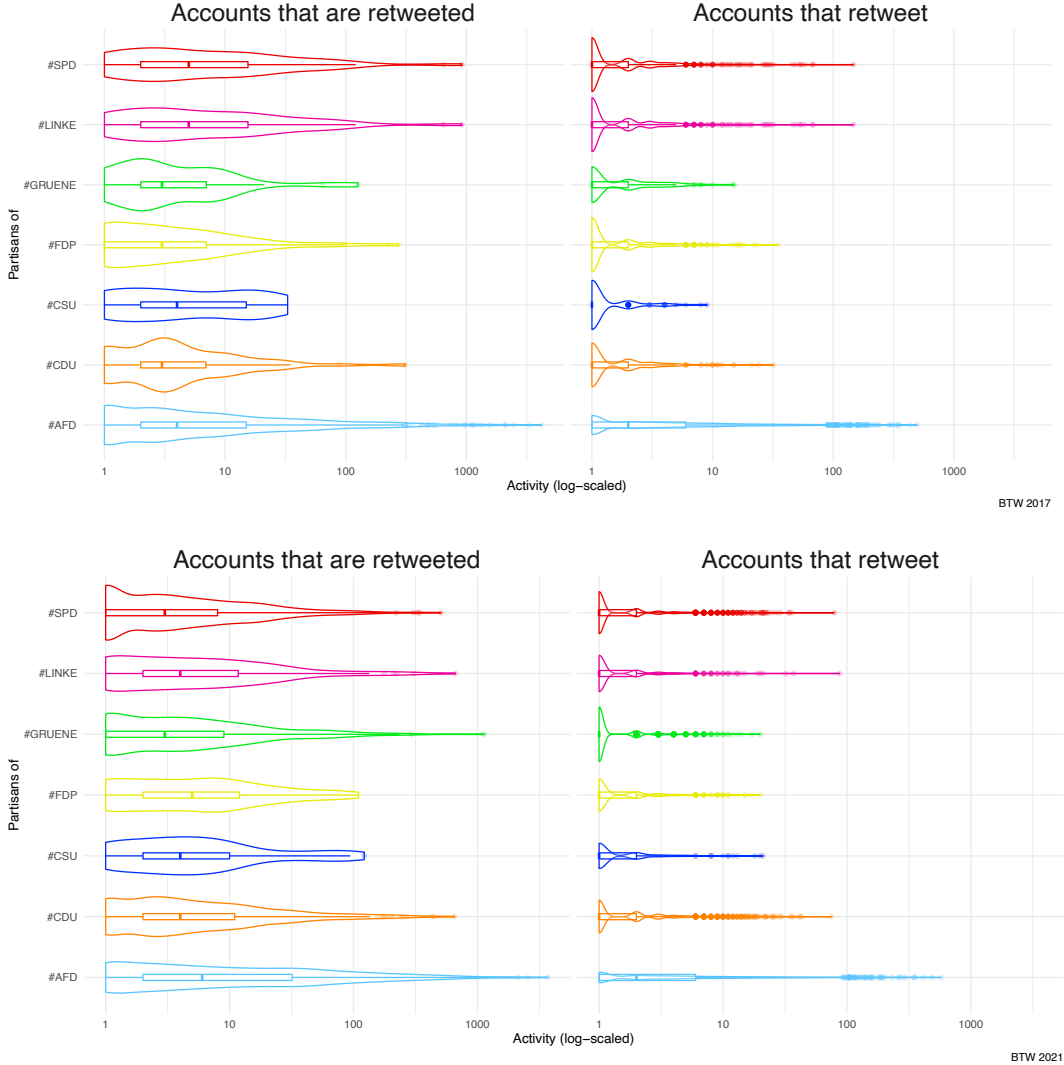


Figure 3: Activity of partisans of various parties contrasted between the final week of the election campaign in 2017 and 2021

The distribution shows differences between 2017 and 2021. At first, AfD partisans are much more active than partisans of other parties. Besides, the retweets are more unequally distributed than for other communities. AfD network elites have much higher values of weighted indegree as the number of being retweeted than the elites in other party networks. Additionally, the activity of other party supporters increased notably between 2017 and 2021, especially for the AfD, FDP, and CDU, as indicated by the means in the box-candle plots in Figure 3.

After analysing the use of party hashtags and identifying partisan communities within the retweet networks, the analysis proceeds with analysing the co-occurrence of these communities in the broader debate on the elections. This co-occurrence is an indicator of strategic hashtag use. Additionally, the co-occurrence of partisans from different parties indicates ideological closeness between groups.

3.5.3. Polarisation of the broader electoral debate (RQ3a & RQ3b)

The acronym hashtags of the *Bundestagswahlen* as *#BTW17* and *#BTW21* are common hashtags to refer to the upcoming elections on Twitter. Figure 4 represents the retweet network polarisation of *#BTW17* (top) with 72,745 nodes and 168,239 edges and *#BTW21* (bottom) with 91,789 nodes and 225,925 edges during the final week of the campaigns before the 2017 and 2021 German Federal elections. Both retweet networks are clustered and filtered to only show accounts retweeted more than 100 times during the observation period. In 2017 red cluster contains significant media outlets and all parties except the AfD (blue cluster). In 2021 red cluster contains major media outlets and left parties. However, the CDU and the main accounts of the FDP and CSU are located in the blue cluster dominated by AfD politicians. Two large clusters divide the network. The colouring indicates the community assignments of individual nodes based on the Louvain algorithm (Blondel et al., 2008; Chen et al., 2014). In contrast to earlier observation periods in 2017, official CDU, CSU, or FDP accounts are also part of the blue AfD partisan cluster in which official AfD accounts and the German tabloid BILD are the most retweeted accounts (Figures A2 and A3). A qualitative content analysis of the 30 most retweeted accounts in

Table A2 and Table A3 shows that AfD politicians are present in the blue cluster in 2017 and 2021. In 2021, however, CDU politicians were also in the same cluster. Table A4 illustrates

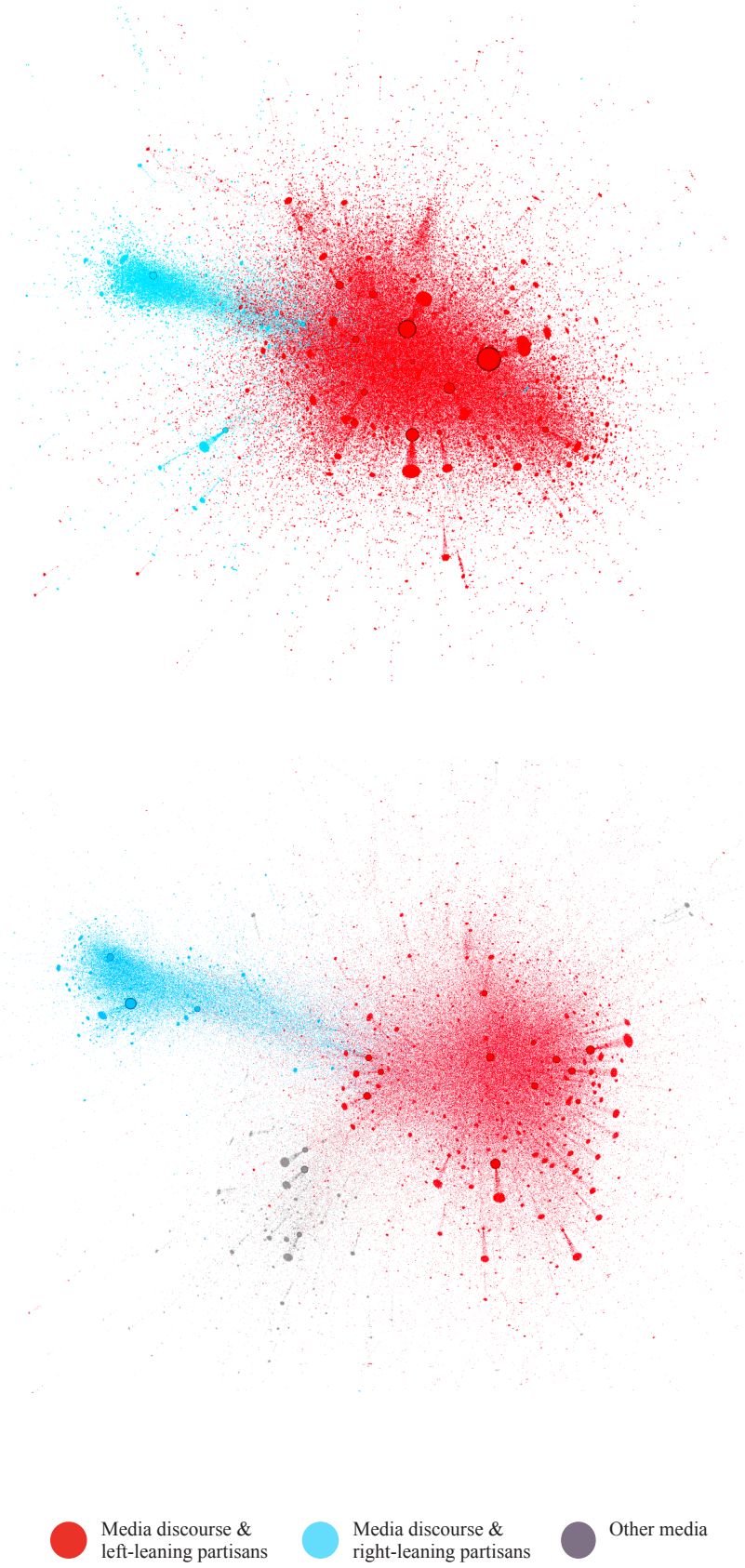


Figure 4: Retweet network polarisation of #BTW17 (top) and #BTW21 (bottom).

the changing cluster co-membership of official party accounts between 2017 and 2021. While the AfD was isolated as the only party in the blue cluster in 2017, in 2021, the AfD, CDU, CSU, and FDP are located in the blue cluster indicating a broader left-right division that the following section further investigates. The analysis proceeds by examining partisan groups retweeting political elites in the different clusters to investigate RQ4.

3.5.4. The likelihood of partisan co-occurrence in the broader electoral debate (RQ4)

The analysis investigates the likelihood of co-occurrence in partisan communities and the communities in the broader discourses as an indication of strategic hashtag use. The qualitative assessment implied that the #BTW21 debate is more broadly polarised than #BTW17. Figure 5 illustrates the log-likelihoods of the co-occurrence of partisans in the two major communities of the broader debate on the elections on #BTW17 in 2017 and #BTW21 in 2021. The column chart contrasts the univariate outputs of the logistics regression models for each partisan community and potentially co-occurring in the right-wing community (blue). The logistic regression output as the odds of co-occurrence confirms the first observation of the network analysis in the prior section that indicated a polarisation along a classical left-right party divide. Partisans of the AfD, CDU, CSU, and FDP are much more likely to occur in the right-leaning community (blue colour), whereas partisans of left parties such as the Greens, SPD, and LINKE are unlikely to appear in this right-leaning cluster. In contrast to earlier studies where AfD partisans actively polarised the Twittersphere, this is a broader polarisation between political ideologies in which users self-sort by their retweeting behaviour. Moreover, the resulting odds for the candidate hashtag in Figure A2 indicate a similar co-occurrence pattern.

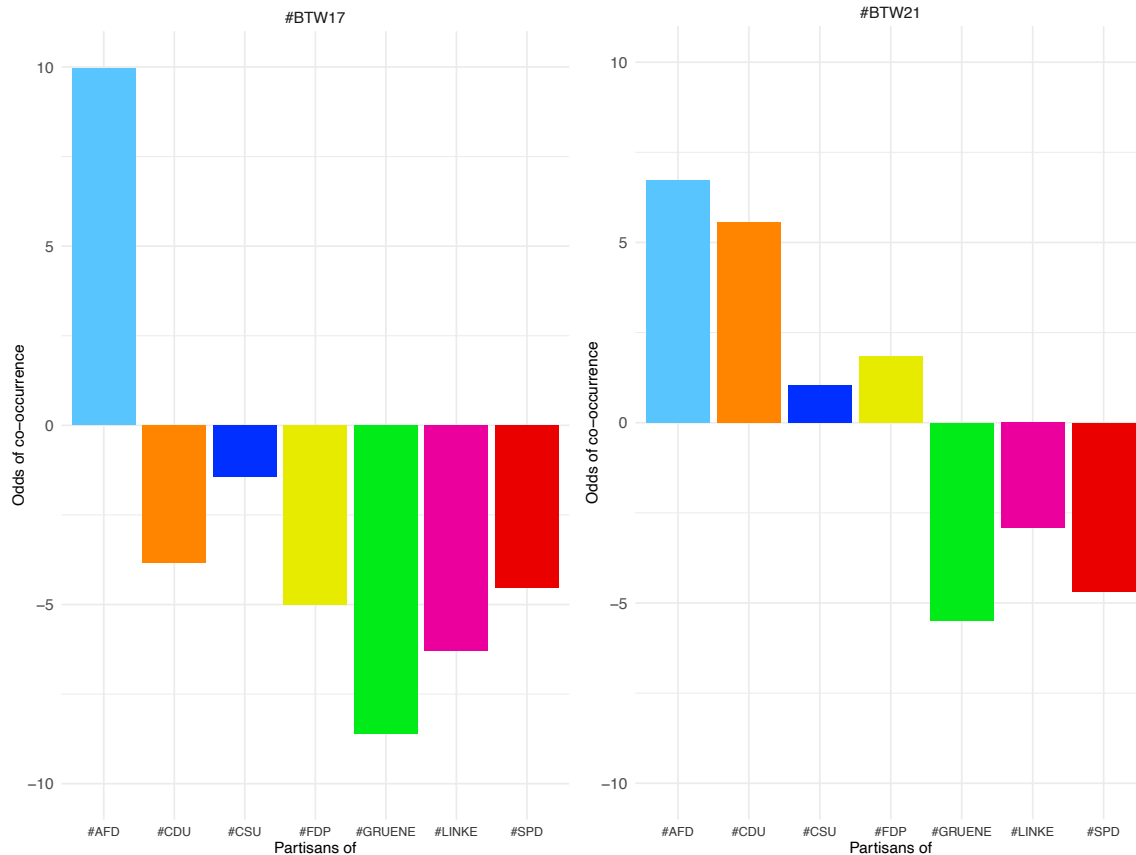


Figure 5: Log-likelihoods of the co-occurrence of partisans in the two major communities of the broader debate on the elections on #BTW17 in 2017 and #BTW21 in 2021

However, it is unclear what caused these differences in partisan retweeting behaviours. Possible explanations could be, 1) political elites retweet different accounts such as media sources, 2) retweeting users/partisans retweet more closely along classical right-left divides, 3) the AfD may have become part of the conservative mainstream and is retweeted by similar users that also retweet the other parties, or the other parties have moved towards the AfD, and AfD partisans tend to be more likely to retweet CDU, CSU and FDP accounts. The following section will discuss these potential explanations and summarise the findings and limitations of the study.

3.6. Discussion

This study investigates the structure of the German political party and candidate hashtags during the final week of the campaigns before the election. The analysis focuses on online political behaviour such as strategic hashtag use and hashjacking, a politically motivated networked communication strategy. The term refers to a coordinated use of hashtags

commonly used by or referring to a politically opposed group. While using other parties' hashtags is not per se strategic, the analyses confirm the hypotheses that partisans and politicians of the German far-right party AfD are more likely to use other parties' hashtags and appear as a partisan group in all other party hashtags displayed by the contra bars in Figure 2. Moreover, Figure 3 illustrates that AfD partisans show much higher activity than other, more heterogeneous groups. Understanding the mechanisms of "hashjacking" as a far-right communication strategy contributes to making sense of the overproportioned representation of far-right actors and opinions on social platforms, e.g., Twitter. During the election period, this may also result in multiplier effects regarding traditional media reporting and may therefore affect the media debate and voter decisions during the final week of the election period.

In contrast to the 2017 elections, however, AfD partisans in 2021 do not appear as an isolated group but in a joined community with CDU politicians and partisans. This joined community shows the higher likelihood of co-occurrence in Figure 5 and shared community membership of AfD and CDU accounts that is highly apparent during the qualitative assessment of the most retweeted accounts. The high likelihood of AfD and CDU partisan's co-occurrence in a right-wing network cluster (in #BTW21) indicates that the parties have become ideologically more similar. The following sub-sections reflect on this main finding and the study's limitations.

3.6.1. Findings

The analysis showed that the Twittersphere, represented by essential political hashtags during the German Federal election, has polarised into a so-called left-right divide between party supporters. While in 2017, AfD partisans were located in an isolated cluster and effectively polarised the hashtag discourse on the elections, the 2021 #BTW21 shows a broader polarisation, especially between the AfD and CDU on the one hand and Greens, SPD, and LINKE on the other side. Partisan clusters of the FDP and CSU are also in the right-wing cluster, with a lower likelihood of partisan co-occurrence as measured by the odd ratios in Figure 5. This assessed polarisation as the proportion of accounts represented in the AfD and the respective party community results from self-sorting of retweeting individuals and the content provided by political elites. For instance, in the CDU community,

politicians of the party's right like Friedrich Merz or former president of the internal intelligence agency "Verfassungsschutz" Hans-Georg Maaßen, were among the most frequently retweeted. The closer location of AfD and CDU in the network structure and broader polarisation may indicate a political shift of Germany's largest party, CDU. This indication of a political shift appears sensible since Friedrich Merz has become the party leader after the election and former candidate Armin Laschet stepping.

Moreover, the broader left-right polarisation and higher closeness of CDU partisans and AfD partisans may indicate an increasing online polarisation of social media and online news sharing and consumption that was visible in the US media system during the Trump campaign in 2016 (Benkler et al., 2018). While the AfD did not increase its vote share, it seems like the CDU has moved closer to the AfD and its online partisans or become more 'retweetable' for AfD partisans. This indicates an ideological shift and may have some forward indication since a right shift within the party was realised when Friedrich Merz became the new party leader after former party leader and chancellor candidate Armin Laschet stepped back after the elections.

3.6.2. Limitations of the study and methodology

The study design and methodology also come with some limitations. First, choosing hashtags as selectors also limits the analysis to the streams of debate and information (Burgess and Bruns, 2012; Weller et al., 2013). However, increasing the visibility of party hashtags is a campaign goal of parties interested in increasing the frequency of messages referring to the organisation and its election promises and policy plans. Concerning the assessed frequencies, the retrospective collection may not contain tweets that Twitter or the owners of the accounts, or accounts that deleted themselves or were deleted by Twitter, e.g., for automation or conflicting with the community guidelines. However, a study by Keller and Klinger (2019) indicated that automated accounts, so-called social bots, only played a limited role in the German political Twittersphere during the 2017 Federal elections.

Another limitation of the study is the focus on retweeting behaviour, which leaves out quote tweets and mentions. The focus on retweeting has been chosen for two reasons.

Firstly, the network structures and mechanisms differ between tweet types (Conover, 2011). Secondly, the retweet networks represent the debates sufficiently since most tweets in the selected hashtag debates are retweets.

Additionally, modularity-based community detection has some limitations. Modularity values and respective cluster detection vary when repeated. Due to this variation, the reproducibility of the research has its limitations. However, the edge lists of the analysed retweet networks and Gephi output data on measures like centrality values and community memberships are published on the authors' GitHub account¹⁵ to allow for a replication of the approach. This data allows colleagues a comparative assessment of the results, e.g., by applying various community detection algorithms to test the limits of modularity maximisation for community detection (Lancichinetti and Fortunato, 2009; 2011; Gates et al., 2016). The study recognises existing research design limitations and encourages further research based on the collected data and methodological approach.

3.6.3. Further research

The study finds a broader level of polarisation during the election campaigns in 2021, indicating an ideological shift of the CDU. Thus, it is vital to continuously investigate the political Twitter-sphere e.g., as a potential indicator of ideological shifts in the political spectrum. Due to this visualisation and analysis of political behaviour, Twitter constitutes something like a political big-data microscope. With regards to the findings of this study, further research needs to assess whether the level of polarisation into a clear left-right divide was only a side-effect of higher politicisation, activity, and, thus polarisation, during the campaign period or whether it marks a lasting shift of the German political sphere. The findings align with a move towards more conservative CDU leader Merz after the elections and may have had an indicator function for ideological shifts of political parties and represent individual and party positions (Ceron, 2016; Sältzer, 2022).

Further research should also investigate whether membership in partisan communities varies over time and, if so, precisely for what proportion of its members. Moreover, it

¹⁵ To allow reproducibility of the study, network edge lists and anonymized data on network measures is publicly accessible on the author's GitHub profile: https://github.com/philippdarius/SNAM_Darius_2022

remains under-researched to what extent community memberships in retweet networks represent differences in political beliefs or offline political behavior. The association between online and offline political behavior can be further investigated by linking social media and survey data (Beuthner et al., 2021; Karlsen and Enjolras, 2016; Stier et al., 2020; Sloan et al., 2020). Further methodological research should investigate social media data linked with panel survey data during elections to better understand to what extent the polarisation of hashtag discourses represents different political alignment and a temporal perspective also opinion dynamics. Moreover, the left-right division found in the communicative behaviours could be extended by comparing language-use in the communities (Däubler and Benoit, 2021). While further research should investigate the impact of the AfD and media outlets on the political sphere on- and offline, the observed broader polarisation could also result from higher social media efforts by all parties. In terms of a higher activity of politicians and partisans, these efforts may result in a more explicit representation of underlying ideologies in the strategic behaviour of political elites and partisans.

3.6.4. Conclusions

Networked and digital campaigning has become a crucial part of election campaigning. Concerning networked digital campaigning, this study assesses strategic communication and polarisation on Twitter during election campaigns. While politicians and partisans of all parties use hashtags strategically to link their tweets to broader discourses, AfD partisans use other parties' hashtags much more frequently. This higher frequency is a sign of a purposeful hashjacking strategy that reflects the party's anti-establishment character. From the network perspective, these AfD partisans built an isolated community due to strategic retweeting behaviour and ideological differences. While this was the case during the 2017 federal elections, this study finds a much broader left-right polarisation of the electoral debate on Twitter during the 2021 German federal elections. On the right-leaning side of the political spectrum, partisans of the AfD, CDU, FDP, and CSU are more likely to appear within the same community, whereas politicians and partisans of the SPD, GRUENE, and LINKE are more likely to appear in the broader community with major news outlets. While the AfD intended to hashjack other party hashtags in both elections, they did not hashjack the broader electoral debate on #BTW17 and #BTW21. However, the AfD and CDU have become ideologically more similar, as indicated in a shared community membership on

Twitter. This online polarisation and closeness to AfD partisans resonate with or might even foreshadow a right shift of the party leadership in the months after the elections when CDU members elected Friedrich Merz as the new party leader for times in opposition. Concludingly, the study contributes to research on online political behaviour during election campaigns and calls for further development of measurement methods of online discourses as measures of ideology, opinion dynamics, and political polarisation.

References

- Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MF, Lee J, Mann M, Merhout F, Volfovsky A (2018) Exposure to opposing views on social media can increase political polarisation. *Proc Natl Acad Sci* 115(37):9216–9221. <https://doi.org/10.1073/pnas.1804840115>.
- Bail C (2021) *Breaking the social media prism*. Princeton University Press, Princeton.
- Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. *Proc Int AAAI Conf Web Soc Media* 3(1):361–362.
- Benkler Y, Faris R, Roberts H (2018) *Network propaganda: manipulation, disinformation, and radicalization in American politics*. Oxford University Press, Oxford.
- Bennett WL, Segerberg A (2013) *The logic of connective action digital media and the personalization of contentious politics*. Cambridge University Press, Cambridge.
- Bernard H (2004) The development of social network analysis: a study in the sociology of science. *Social Netw* 27:377–84. <https://doi.org/10.1016/j.socnet.2005.06.004>.
- Beuthner C, Breuer J, Jünger S (2021) Data linking—linking survey data with geospatial, social media, and sensor data. *GESIS Survey Guidelines*. https://doi.org/10.15465/GESIS-SG_EN_039.
- Bild DR, Liu Y, Dick RP, Mao ZM, Wallach DS (2015) Aggregate characterization of user behavior in Twitter and analysis of the retweet graph. *ACM Trans Internet Technol (TOIT)* 15(1):1–24. <https://doi.org/10.1145/2700060>.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp* 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- Blumler JG (2016) The fourth age of political communication. *Politiques De Commun* 6(1):19–30. <https://doi.org/10.3917/pdc.006.0019>.
- Bode L, Hanna A, Yang J, Shah DV (2015) Candidate networks, citizen clusters, and political expression: strategic hashtag use in the 2010 midterms. *ANNALS Am Acad Polit Soc Sci* 659(1):149165. <https://doi.org/10.1177/0002716214563923>.
- Bossetta M (2018) The digital architectures of social media: comparing political campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US election. *J Mass Commun Quart* 95(2):471–496.
- Boyd D, Golder S, Lotan G (2010) Tweet, Tweet, Retweet: conversational aspects of retweeting on Twitter. *Hawaii Int Conf Syst Sci*. <https://doi.org/10.1109/HICSS.2010.412>.
- Bramson A, Grim P, Singer DJ, Berger WJ, Sack G, Fisher S, Flocken C, Holman B (2017) Understanding polarisation: meanings, measures, and model evaluation. *Philos Sci* 84(1):115–159. <https://doi.org/10.1086/688938>.

Broido AD, Clauset A (2019) Scale-free networks are rare. *Nat Commun* 10(1):1017. <https://doi.org/10.1038/s41467-019-08746-5>.

Bronstein J, Aharony N, Bar-Ilan J (2018) Politicians' use of Facebook during elections: Use of emotionally-based discourse, personalization, social media engagement and vividness. *Aslib J Inf Manag*.

Bruns A, Burgess J (2011) The use of twitter hashtags in the formation of ad hoc publics. In: *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*, edited by A. Bruns and P. De Wilde, pp. 1–9. United Kingdom: The European Consortium for Political Research (ECPR). <https://eprints.qut.edu.au/46515/>.

Bruns A, Moon B, Paul A, Münch F (2016) Towards a typology of hashtag publics: a large-scale comparative study of user engagement across trending topics. *Commun Res Pract* 2(1):20–46. <https://doi.org/10.1080/22041451.2016.1155328>.

Burgess J, Bruns A (2012) Twitter archives and the challenges of "Big Social Data" for media and communication research. *M/C J*, 15(5).

Carrington PJ, Scott J, Wasserman S (2005) *Models and methods in social network analysis*, vol 28. Cambridge University Press, Cambridge.

Caton S, Hall M, Weinhardt C (2015) How do politicians use Facebook? An applied social observatory. *Big Data Soc* 2(2):2053951715612822.

Ceron A (2016) Intra-party politics in 140 characters. *Party Polit*. <https://doi.org/10.1177/1354068816654325>.

Colleoni E, Rozza A, Arvidsson A (2014) Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data: political homophily on Twitter. *J Commun* 64(2):317–332. <https://doi.org/10.1111/jcom.12084>

Conover MD, Gonçalves B, Flammini A, Menczer F (2012) Partisan asymmetries in online political activity. *EPJ Data Sci* 1(1):1–19. <https://doi.org/10.1140/epjds6>.

Conover MD, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F (2011) Predicting the political alignment of twitter users, pp. 192–199. doi: <https://doi.org/10.1109/PASSAT/SocialCom.2011.34>.

Copeland L, Römmele A (2014) Beyond the base? Political parties, citizen activists, and digital media use in the 2009 German federal election campaign. *J Inf Technol Polit* 11(2):169–85. <https://doi.org/10.1080/19331681.2014.902783>.

Couldry N, Hepp A (2017) *The mediated construction of reality*. John Wiley & Sons, New York.

Cowls J, Darius P, Santistevan D, Schramm M (2022) Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance. *New Media Soc.* <https://doi.org/10.1177/14614448221085559>.

Däubler T, Benoit K (2021) Scaling hand-coded political texts to learn more about left-right policy content. *Party Polit.* <https://doi.org/10.1177/13540688211026076>.

Dahlgren P (2005) The internet, public spheres, and political communication: dispersion and deliberation. *Polit Commun* 22(2):147–62. <https://doi.org/10.1080/10584600590933160>.

Darius P, Stephany F (2019) Hashjacking' the debate: polarisation strategies of Germany's political far-right on Twitter. In: Weber I, Darwish KM, Wagner C, Zagheni E, Nelson L, Aref S, Flöck F (eds) *Social informatics. Lecture notes in computer science.* Springer International Publishing, Cham, pp 298–308. https://doi.org/10.1007/978-3-030-34971-4_21.

Darius P, Stephany F (2022) How the Far-Right Polarises Twitter: 'Hashjacking' as a Disinformation Strategy in Times of COVID-19. In *Complex Networks & Their Applications X*, edited by Rosa Maria Benito, Chantal Cherifi, Hocine Cherifi, Esteban Moro, Luis M. Rocha, and Marta Sales Pardo, 100–111. *Studies in Computational Intelligence.* Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-93413-2_9.

Darius P, Urquhart M (2021) Disinformed social movements: a largescale mapping of conspiracy narratives as online harms during the COVID-19 pandemic. *Online Soc Netw Media* 26:100174. <https://doi.org/10.1016/j.osnem.2021.100174>.

De Oliveira JF, Marques-Neto HT, Karsai M (2022) Measuring the effects of repeated and diversified influence mechanism for information adoption on Twitter. *Soc Netw Anal Min* 12(1):1–15. <https://doi.org/10.1007/s13278-021-00844-x>.

Dimitrova DV, Matthes J (2018) Social media in political campaigning around the world: theoretical and methodological challenges. *J Mass Commun Quart* 95(2):333–42. <https://doi.org/10.1177/1077699018770437>.

Ekman M, Widholm A (2015) Politicians as media producers: current trajectories in the relation between journalists and politicians in the age of social media. *Journal Pract* 9(1):78–91.

Engesser S, Ernst N, Esser F, Büchel F (2017) Populism and social media: How politicians spread a fragmented ideology. *Inf Commun Soc* 20(8):1109–1126.

Enli G, Simonsen C-A (2018) "Social media logic" meets professional norms: twitter hashtags usage by journalists and politicians. *Inf Commun Soc* 21(8):1081–1096.

Esser F, Strömbäck J (2014) *Mediatization of politics: understanding the transformation of western democracies.* Springer, Cham.

Ferree MM, Gamson WA, Gerhards J, Rucht D (2002) Four models of the public sphere in modern democracies. *Theory Soc* 31(3):289–324.

- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(35):75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>.
- Fowler JH (2006) Connecting the congress: a study of cosponsorship network. *Polit Anal* 14(4):456–87. <https://doi.org/10.1093/pan/ mpl002>.
- Garimella K, Weber I (2017) A long-term analysis of polarisation on Twitter. ArXiv: <http://arxiv.org/abs/1703.02769> [Cs], March 17, 2017.
- Gates KM, Henry T, Steinley D, Fair DA (2016) A Monte Carlo evaluation of weighted community detection algorithms. *Front Neuroinform*. <https://doi.org/10.3389/fninf.2016.00045>.
- Gerodimos R, Justinussen J (2015) Obama’s 2012 Facebook campaign: political communication in the age of the like button. *J Inform Tech Polit* 12(2):113–132.
- Gibson RK, Römmele A (2004) Internet campaigning around the world. *Campaigns Elect* 25(4):38–46.
- Gibson R, Römmele A, Williamson A (2014) Chasing the digital wave: international perspectives on the growth of online campaigning. *J Inf Technol Polit* 11(2):123–29. <https://doi.org/10.1080/19331681.2014.903064>.
- Gil de Zúñiga H, Koc Michalska K, Römmele A (2020) Populism in the era of Twitter: How social media contextualized new insights into an old phenomenon. *New Media Soc* 22(4):585–594. <https://doi.org/10.1177/1461444819893978>.
- Gillespie T (2018) *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, Connecticut.
- Gillespie T (2017) Governance of and by platforms. *SAGE Handbook of Social Media*, London, pp 254–78.
- Graham T, Broersma M, Hazelhoff K, van ’t Haar G (2013) Between broadcasting political messages and interacting with voters. *Inf Commun Soc* 16(5):692–716. <https://doi.org/10.1080/1369118X.2013.785581>.
- Grant WJ, Moon B, Grant JB (2010) Digital dialogue? Australian politicians’ use of the social network tool Twitter. *Austral J Polit Sci* 45(4):579–604.
- Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 US presidential election. *Science* 363(6425):374–78. <https://doi.org/10.1126/science.aau2706>.
- Gueorguieva V (2008) Voters, MySpace, and YouTube: the impact of alternative communication channels on the 2006 election cycle and beyond. *Soc Sci Comput Rev* 26(3):288–300. <https://doi.org/10.1177/0894439307305636>.

- Habermas J (1991) *The structural transformation of the public sphere: an inquiry into a category of Bourgeois society*. MIT Press, Cambridge.
- Hadgu AT, Garimella K, Weber I (2013) Political Hashtag Hijacking in the US. In: *Proceedings of the 22nd international conference on world wide web*, pp. 55–56.
- Hafner-Burton EM, Kahler M, Montgomery AH (2009) Network analysis for international relations. *Int Organ* 63(3):559–592.
- Hepp A (2019) *Deep mediatization: key ideas in media & cultural studies*. Routledge, London.
- Hjarvard S (2008) The mediatization of society: a theory of the media as agents of social and cultural change. *Nordicom Rev* 29(2):102–131.
- Howard PN (2005) Deep democracy, thin citizenship: the impact of digital media in political campaign strategy. *Ann Am Acad Pol Soc Sci* 597(1):153–170. <https://doi.org/10.1177/0002716204270139>.
- Jacobs K (2018) A populist paradise? Examining populists' twitter adoption and use, p. 17.
- Jacomy M, Venturini T, Heymann S, Bastia M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE* 9(6):e98679. <https://doi.org/10.1371/journal.pone.0098679>.
- Jensen MJ (2017) Social media and political campaigning: changing terms of engagement? *Int J Press/Polit* 22(1):23–42.
- Jungherr A (2016) Four functions of digital tools in election campaigns: the German case. *Int J Press/Polit* 21(3):358–377. <https://doi.org/10.1177/1940161216642597>.
- Jungherr A, Schroeder R, Stier S (2019) Digital media and the surge of political outsiders: explaining the success of political challengers in the United States, Germany, and China. *Soc Media* p. 12.
- Kalsnes B, Larsson AO, Enli G (2017) The social media logic of political interaction: exploring citizens' and politicians' relationship on facebook and twitter. *First Monday*. <https://doi.org/10.5210/fm.v22i2.6348>.
- Karlsen R, Enjolras B (2016) Styles of social media campaigning and influence in a hybrid political communication system: linking candidate survey data with Twitter data. *Int J Press/Polit* 21(3):338–357. <https://doi.org/10.1177/1940161216645335>.
- Katzenbach C, Ulbricht L (2019) Algorithmic governance. *Internet Policy Rev* 8(4):1–18.
- Keller, Franziska B, David Schoch, and Junghwan Yang. "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign," 2019, 26.

- Klinger U, Svensson J (2015) The emergence of network media logic in political communication: a theoretical approach. *New Media Soc* 17(8):1241–1257. <https://doi.org/10.1177/1461444814522952>.
- Knüpfer C, Hoffmann M, Voskresenskii V (2020) Hijacking MeToo: Transnational dynamics and networked frame contestation on the far right in the case of the ‘120 Decibels’ campaign. *Inf Commun Soc*. <https://doi.org/10.1080/1369118X.2020.1822904>.
- Koc-Michalska K, Klinger U (2021) Populism as communication: political campaigning on Facebook Krämer B (2017) Populist online practices: the function of the internet in right-wing populism. *Inf Commun Soc* 20(9):1293–1309. <https://doi.org/10.1080/1369118X.2017.1328520>.
- Kreiss D, Lawrence RG, McGregor SC (2018) In their own words: political practitioner accounts of candidates, audiences, affordances, genres, and timing in strategic social media use. *Polit Commun* 35(1):8–31. <https://doi.org/10.1080/10584609.2017.1334727>.
- Krippendorff K (2018) *Content analysis: an introduction to its methodology*. SAGE Publications, London.
- Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. *Phys Rev E* 80(5):056117. <https://doi.org/10.1103/PhysRevE.80.056117>.
- Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Phys Rev E* 84(6):066122. <https://doi.org/10.1103/PhysRevE.84.066122>.
- Larsson AO, Kalsnes B (2014) “Of Course We Are on Facebook”: use and non-use of social media among Swedish and Norwegian politicians. *Eur J Commun* 29(6):653–667.
- Lilleker Darren, Jackson Nigel (2013) *Political campaigning, elections and the internet: comparing the US, UK, France and Germany*, vol 4. Routledge, London.
- Lilleker DG, Koc-Michalska K, Negrine R, Gibson R, Vedel T, Strudel S (2017) Social media campaigning in Europe: mapping the terrain. *J Inf Technol Polit* 14(4):293–98. <https://doi.org/10.1080/19331681.2017.1397239>.
- Magin M, Podschuweit N, Haßler J, Russmann U (2017) Campaigning in the fourth age of political communication. A multi-method study on the use of Facebook by German and Austrian parties in the 2013 national election campaigns. *Inf Commun Soc* 20(11):1698–1719.
- Maulana, Ardian, Hokky Situngkir, and Rendra Suroso (2022) "Dynamics of Polarisation and Coalition Formation in Signed Political Elite Networks." In *Complex Networks & Their Applications X*, edited by Rosa Maria Benito, Chantal Cherifi, Hocine Cherifi, Esteban Moro, Luis M. Rocha, and Marta Sales-Pardo, 94–103. *Studies in Computational Intelligence*. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-93409-5_9.
- Mayring, P (2014) *Qualitative Content Analysis*, 2014, 144.

Metaxas P, Mustafaraj E, Wong K, Zeng L, O’Keefe M, Finn S (2015) What do retweets indicate? Results from user survey and metareview of research. *Proc Int AAAI Conf Web Social Media* 9(1):658–661.

Morstatter F, Shao Y, Galstyan A, Karunasekera S (2018) From alright to Alt-Rechts: Twitter analysis of the 2017 German federal election. *Companion Proc Web Conf* 2018:621–28.

Mudde C (2004) The populist zeitgeist. *Gov Oppos* 39(4):541–563.

Mudde C, Kaltwasser CR (2013) Exclusionary vs. inclusionary populism: comparing contemporary Europe and Latin America. *Govern Oppos* 48(2):147–74. <https://doi.org/10.1017/gov.2012.11>.

Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577–82. <https://doi.org/10.1073/pnas.0601602103>.

Nielsen RK, Vaccari C (2013) "Do People’ like’ Politicians on Facebook? Not really large-scale direct candidate-to-voter online communication as an outlier phenomenon. *Int J Commun* 7:24.

Norris P (2004) The evolution of election campaigns: eroding political engagement?, p. 27.

Nuernbergk C, Conrad J (2016) Conversations and campaign dynamics in a hybrid media environment: use of Twitter by members of the German Bundestag. *Social Media Soc* 2(1):2056305116628888. <https://doi.org/10.1177/2056305116628888>.

Oelsner K, Heimrich L (2015) Social media use of German politicians: towards dialogic voter relations? *German Politics* 24(4):451–468.

Rao A, Morstatter F, Hu M, Chen E, Burghardt K, Ferrara E, Lerman K (2021) Political partisanship and antiscience attitudes in online discussions about COVID-19: Twitter content analysis. *J Med Internet Res* 23(6):e26692.

Rau JP, Stier S (2019) Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien? *Zeitschrift für Vergleichende Politikwissenschaft* 13:399–417.

Römmele A, Gibson R (2020) Scientific and subversive: the two faces of the fourth era of political campaigning. *New Media Soc* 22(4):595–610. <https://doi.org/10.1177/1461444819893979>.

Römmele A, von Schneidmesser D (2016) Election campaigning enters a fourth phase: the mediatized campaign. *Zeitschrift Für Politikwissenschaft* 26(4):425–42. <https://doi.org/10.1007/s41358-016-0070-z>.

Sältzer M (2022) Finding the bird’s wings: dimensions of factional conflict on Twitter. *Party Politics* 28(1):61–70. <https://doi.org/10.1177/1354068820957960>.

Schmitt-Beck R, Farrell DM (2002) Studying political campaigns and their effects. Do political campaigns matter? Campaign effects in elections and referendums. Routledge, England, pp 1–21.

Schumann S, Thomas F, Ehrke F, Bertlich T, Dupont JC (2021) Maintenance or change? Examining the reinforcing spiral between social media news use and populist attitudes. *Inf Commun Soc*. <https://doi.org/10.1080/1369118X.2021.1907435>.

Scott J (2017) *Social Network Analysis*. SAGE Publications Ltd, London. <https://doi.org/10.4135/9781529716597>.

Sloan L, Jessop C, Al Baghal T, Williams M (2020) Linking survey and Twitter data: informed consent, disclosure, security, and archiving. *J Empir Res Human Res Ethics* 15(1–2):63–76. <https://doi.org/10.1177/1556264619853447>.

Sorce G (2018) Sounding the alarm for Right-Wing# metoo: ‘120 Dezibel’ in Germany. *Fem Media Stud* 18(6):1123–1126.

Stier S, Bleier A, Bonart M, Mörsheim F, Bohlouli M, Nizhegorodov M, Posch L, Maier J, Rothmund T, Staab S (2018a) Systematically monitoring social media: the case of the German federal election 2017. Preprint SocArXiv, March 9, 2018a. Doi:<https://doi.org/10.31235/osf.io/5zpm9>.

Stier S, Bleier A, Lietz H, Strohmaier M (2018b) Election campaigning on social media: politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Polit Commun* 35(1):50–74. <https://doi.org/10.1080/10584609.2017.1334728>.

Stier S, Breuer J, Siegers P, Thorson K (2020) Integrating survey data and digital trace data: key issues in developing an emerging field. *Social Sci Comput Rev* 38(5):503–516. <https://doi.org/10.1177/0894439319843669>.

Stier S, Mangold F, Scharkow M, Breuer J (2022) Post post-broadcast democracy? News exposure in the age of online intermediaries. *Am Polit Sci Rev* 116:768–774. <https://doi.org/10.1017/S0003055421001222>.

Stockmann D (2022) Tech companies and the public interest: the role of the state in governing social media platforms. *Inf Commun Soc*. <https://doi.org/10.1080/1369118X.2022.2032796>.

Stoetzer L, Munzert S, Lowe W, Çalı B, Gohdes A, Helbling M, Maxwell R, Traunmueller R (2021) Affective partisan polarisation and moral dilemmas during the COVID-19 pandemic.

Strömbäck J (2008) Four phases of mediatization: an analysis of the mediatization of politics. *Int J Press/Polit* 13(3):228–246. <https://doi.org/10.1177/1940161208319097>.

Tianlong C, Singh P, Bassler K (2017) Network community detection using modularity density measures. *J Stat Mech Theory Exp*. <https://doi.org/10.1088/1742-5468/aabfc8>.

Tucker JA, Andrew G, Pablo B, Cristian V, Alexandra S, Sergey S, Denis S, Brendan N (2018) Social media, political polarisation, and political disinformation: a review of the scientific literature.

Van Dam C, Pang-Ning T (2016) Detecting hashtag hijacking from Twitter. In Proceedings of the 8th ACM conference on web science, pp. 370–71.

Verweij P (2012) Twitter Links between politicians and journalists. *Journal Pract* 6(5–6):680–691.

Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Social network analysis: methods and applications. Cambridge University Press, New York. <https://doi.org/10.1017/CBO9780511815478>.

Weber I, Garimella VRK, Teka A (2013) Political hashtag trends. European conference on information retrieval. Springer, Berlin, Heidelberg, pp 857–860. https://doi.org/10.1007/978-3-64236973-5_102.

Weller K (ed) (2013) *Twitter and Society*. Digital Formations, vol 89. Peter Lang, New York.

White MD, Marsh EE (2006) Content analysis: a flexible methodology. *Library Trends* 55(1):22–45. <https://doi.org/10.1353/lib.2006.0053>.

Xanthopoulos P, Panagopoulos OP, Bakamitsos GA, Freudmann E (2016) Hashtag hijacking: What it is, why it happens and how to avoid it. *J Digit Social Media Market* 3(4):353–362.

Yan P (2019) Information bridges: understanding the informational role of network brokerages in polarised online discourses. In: Greene Taylor N, Christian-Lamb C, Martin MH, Nardi B (eds) *Information in contemporary society*. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp 377–88. https://doi.org/10.1007/978-3-030-15742-5_36.

Chapter 4 – Disinformed social movements: A large-scale mapping of conspiracy narratives as online harms during the COVID-19 pandemic

Abstract¹⁶

The COVID-19 pandemic caused high uncertainty regarding appropriate treatments and public policy reactions. This uncertainty provided a perfect breeding ground for spreading conspiratorial anti-science narratives based on disinformation. Disinformation on public health may alter the population's hesitance to vaccinations, counted among the ten most severe threats to global public health by the United Nations. We understand conspiracy narratives as a combination of disinformation, misinformation, and rumour that are especially effective in drawing people to believe in post-factual claims and form disinformed social movements. Conspiracy narratives provide a pseudo-epistemic background for disinformed social movements that allow for self-identification and cognitive certainty in a rapidly changing information environment. This study monitors two established conspiracy narratives and their communities on Twitter, the anti-vaccination and anti-5G communities, before and during the first UK lockdown. The study finds that, despite content moderation efforts by Twitter, conspiracy groups were able to proliferate their networks and influence broader public discourses on Twitter, such as #Lockdown in the United Kingdom.

¹⁶ Citation: Darius P, Urquhart M (2021) "Disinformed Social Movements: A Large-Scale Mapping of Conspiracy Narratives as Online Harms during the COVID-19 Pandemic." *Online Social Networks and Media* 26: 100174. <https://doi.org/10.1016/j.osnem.2021.100174>.

4.1. Introduction

Throughout 2020 and well into 2021, the COVID-19 pandemic dominated news headlines and public conversation, resulting in a communication environment saturated with confusing, contradictory, and often outright false information. In this context, social media platforms host alternative and competing modes of online participation in civil discourse, from honest engagement in debate to disruptive or subversive communication. As governments and public health bodies battle this growing 'infodemic' while simultaneously dealing with the actual epidemiological emergency, conspiracy theory threatens to disrupt the effective transfer of information and erode popular confidence in public institutions. Conspiracy narratives or theories have a long pedigree in the psychological, sociological and philosophical disciplines. Common to most interpretations is an emphasis on the pursuit, possession and denial of knowledge in the face of an epistemic adversary, usually in a position of power (Imhoff et al., 2018). Combined with a perception of "nefarious intent", the result is often an intense scepticism towards figures of authority (Lewandowsky and Cook, 2020). Recent concern about conspiracy theory is pressing for two main reasons. First, it seems to be enjoying a favourable tailwind. A mainstreaming effect is edging conspiracy theories further into the public domain on the current of celebrity gossip, academic opinion and foreign state interference. Second, this phenomenon is taking place at an opportune moment. Various opinion polls indicate a general crisis of trust in governments and social institutions exacerbated by the COVID-19 pandemic¹⁷. Trust, especially in communication, has direct implications for the perceived legitimacy of institutions, which can deteriorate rapidly in times of pressure (Schmidtke, 2019). Conspiracy communities mobilising to disrupt the digital landscape may be in a position to benefit from this momentum and accelerate the effect, and therefore present a significant challenge to policymakers and communications professionals with an interest in public trust.

¹⁷ COVID-19: government handling and confidence in health authorities", YouGov, <https://today.yougov.com/topics/international/articles-reports/2020/03/17/perception-government-handling-covid-19>

The purpose of this study is to explore the principles of social movement theory applied to conspiracy narrative with methods from social network analysis to examine the structure of relevant conspiracy communities online. Understanding the scale, density and interconnectedness of these communities will allow us to investigate our hypothesis that conspiracy groups are mobilising and increasingly interwoven during the pandemic. To facilitate that understanding and frame our analysis in a wider discussion around government intervention, we compare two time periods before and after the introduction of so-called lockdown measures in the UK on March 23, 2020. The following pages explore a conceptual framework, rooted in social movement theory, before reviewing some of the relevant empirical literature. After a description of our methodology, we present our results and discuss the findings. The evolving situation threatens to make the practice of strategic communication during the time of COVID-19 an altogether more complex and difficult task. With this understanding, the study concludes with a theoretical reflection of disinformed social movements and a discussion of policy implications for public institutions and governments.

4.2. Background

In the case of COVID-19, an existing lack of trust in governments' and public health bodies' handling of the pandemic appears to have become localised in the expression of specific misinformation and disinformation narratives, notably around anti-5G activism and the vaccine resistance movement. An anecdotal observation of these groups suggests a surge in activity, with existing communities being rejuvenated by increasing attention during the 'infodemic'. In finding new avenues to express dissent, these communities are increasingly adopting a style of engagement that is characterised by militancy, which is antithetical to the norms of civil society discourse.

4.2.1. Theoretical Framework

Given this recent mobilisation of conspiracy communities evident in their offline protest action, the comparison with more traditional social movements is becoming harder to

ignore. While the theoretical concept often has positive connotations driven by a tendency to work for progressive social goals, conspiracy theory communities have been interpreted as the 'dark side of social movements' (Sternisko et al., 2020), an understanding that we have found insightful in our analysis. For Touraine (1995), social movements involve the combination of a principle of identity, a principle of opposition and a principle of totality - mapping these three characteristics onto the conspiracy theory literature is a useful exercise.

Community identity: In the case of conspiracy theory, community identity depends upon the pursuit, possession and denial of knowledge. Acting as membership criteria, a shared belief in specific truth claims helps to build a kind of imagined community that is united in disposition if not in history or tradition (Anderson, 2006). Common accusations of paranoia or delusion also help define the community, as conspiracy belief - understood as 'stigmatised knowledge' - can lead to a minority status that consolidates a sense of belonging, even as it satisfies a narcissistic desire for uniqueness (Cichocka et al., 2016).

The principle of opposition: Reflecting the principle of opposition, this feeling of stigma or persecution is compounded by the perception of 'nefarious intent' that describes the motivation for the architects of any given conspiracy (Touraine, 1995; Lewandowsky and Cook, 2020). Trust is a fundamental component in the infrastructure of knowledge production in modern societies, which depends on things like thorough peer review in the academic discipline, a reliable free press, and the integrity of a just government; where these institutions are perceived to have a malign agenda, trust naturally deteriorates (Keeley, 1999; Aupers, 2012; Einstein and Glick, 2015).

The principle of totality: Well-wrought findings in empirical research show that people who follow one conspiracy theory are highly likely to believe in others and that conspiracy theorists are generally resistant to evidence that contradicts their truth claims (Edy and Risley-Baird, 2016; Sternisko et al., 2020). Indicating how conspiracy thinking can grow to be all-encompassing in a person's worldview, this speaks to Touraine's principle of totality. While this all-encompassing effect is persuasive, we share the view that conspiracy theory belief can be accelerated in times of uncertainty when the volume and velocity of information increase significantly (Marchlewska et al., 2018). Our framing of conspiracy theory communities as a type of social movement resonates with recent insights from social

psychology that indicate early conspiracy belief (until a certain threshold) may positively affect political engagement (Imhoff and Lamberty, 2020). During the COVID-19 emergency, this clouding of the information space has been characterised as an 'infodemic'¹⁸, which allows misinformation and disinformation to penetrate social discourse more effectively than in ordinary circumstances. One concerning symptom of this phenomenon has been the increasing visibility of the anti-vaccination and anti-5G movements, which have generated heightened attention following their vocal participation in the global counter-lockdown movement. Since these conspiracy narratives are much talked about, but rarely explained, the following section briefly summarises their formation in recent decades.

4.2.2. The Anti-5G and Anti-Vaccination Movements

To give readers that are unfamiliar with the particular conspiracy theories some background we summarise the evolution of the narratives and the respective movements.

4.2.2.1. The anti-vaccination movement

Popular resistance to vaccination against disease has a long pedigree. Almost immediately after the introduction of vaccination acts passed in Victorian Britain, anti-vaccination leagues began to challenge the laws on legal, ethical and religious grounds (Wolfe and Sharp, 2002). These prejudices and the concomitant belief that vaccines cause more harm than good - especially in children - have persisted to the present day. The modern 'antivax' movement is commonly said to have been reanimated by former British physician, Andrew Wakefield, who in 1998 published an article in the medical journal *The Lancet* claiming to draw a connection between the Measles, Mumps & Rubella (MMR) vaccine and autism. The study - since retracted and disproved - was later declared "utterly false" by the journal

¹⁸ 1st WHO Infodemiology Conference, WHO, <https://www.who.int/newsroom/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>

editor and Wakefield was barred from practicing medicine in the UK¹⁹, but not without energising a new generation of anti-vaxxers (Hussain et al., 2018).

During the COVID-19 pandemic, vaccine resistance has had a peculiar focus on Microsoft founder and philanthropist Bill Gates, whose Foundation “aspires to advance public goods for global health through technological innovation... by accelerating the development and commercialisation of novel vaccines” (among other goals)²⁰. Seemingly provoked by Gates’ comments on the relationship between vaccinology and world population, and motivated by a surging interest in vaccines during the current epidemiological emergency, anti-vaccination activists have mobilised in protest action the world over. Significantly, the movement has developed an even more conspiratorial character, expressing a visceral reaction to the perceived financial/knowledge elite represented by Bill Gates, and “branching out into various crazy tributaries”²¹ including fears of embedded microchips, thought manipulation and population control.²²

4.2.2.2. The anti-5G movement

Given its emphasis on a more modern technology, anti-5G activism has followed a much shorter timeline than that of the vaccine resistance community, although the development of these two conspiracy theories appear to follow a similar structure. While anti-vaccination sentiment resurfaces as new vaccines are made available, recurring waves of technological advances that have produced microwaves, mobile phones, WiFi, and now 5G bring new impetus to fear and scepticism for those who worry about the risks of electromagnetic radiation.²³ Depending on who you ask, 5G technology is part of a plan to weaken the

¹⁹ Bosely (2010), “Andrew Wakefield struck off register by General Medical Council, The Guardian

²⁰ Bill and Melinda Gates Foundation, Strategy overview, <https://www.gatesfoundation.org/what-we-do/global-health/vaccine-development-and-surveillance>

²¹ Lowe, (2020), “Vaccine Derangement”, Science Translational Medicine, <https://blogs.sciencemag.org/pipeline/archives/2020/06/25/vaccine-derangement>

²² Evstatieva, M. (2020), “Anatomy of a COVID-19 Conspiracy Theory”, NPR, <https://www.npr.org/2020/07/10/889037310/anatomy-of-a-covid-19-conspiracy-theory?t=1595434430326>

²³ Tiffany (2020), “The Great 5G Conspiracy”, The Atlantic

immune system, making people more susceptible to the virus; is an actual tool of disease transmission; or is causing direct harm through electromagnetic radiation that has required the creation of a COVID-19 hoax to cover up the real threat to human life. Claims that 5G poses a threat to human well-being have been disregarded by the World Health Organisation, and yet there continues to be serious resistance among certain communities, with anti-5G activism accelerating during the COVID-19 emergency. Arson attacks against telecoms masts and verbal and physical confrontations with telecoms engineers have been reported throughout the UK and internationally. As with vaccine resistance, anti-5G activism is extremely hostile to elites, who are often said to be pursuing a malevolent agenda that threatens the wellbeing of the general population. The assumption of “nefarious intent” in the role of various actors including the Chinese government, Chinese telecoms firm Huawei, and the World Health Organisation²⁴ is often extreme in nature - a clear expression of the “principle of opposition” that characterises conspiracy theories and social movements more broadly.

4.2.3. Prior empirical research and knowledge gap

Regarding the prominence in the popular conversation, the academic literature around specific conspiracy theories is further developing in the time of COVID-19. Researchers in the field of communication studies have conducted necessary early analyses of the virality of misinformation and the associated implications for public health, while others have begun the important work of outlining the rise in xenophobic and racist attitudes apparently motivated by the origins of COVID-19 (Budhwani and Sun, 2020; Kata, 2010; Larson, 2018; Pennycook et al., 2020). On specific conspiracy theories, some have explored the link between the anti-5G worldview and recent violence and the apparent resilience of the anti-vaccination movement during this time of heightened attention to the necessity of vaccines (Jolley and Paterson, 2020; Megget, 2020). With the surge of conspiracy narratives and adjacent mis- and disinformation during the Coronavirus pandemic, there has been work

²⁴ “5G mobile networks and health”, WHO, <https://www.who.int/news-room/q-a-detail/radiation-5g-mobile-networks-and-health>

investigating the proliferation of messages, communities and narratives on online social networks and media (Ahmed et al., 2020a; Bruns et al., 2020; Chen et al., 2020; Rao et al., 2020; Shahi et al., 2021). Bruns et al. (2020), for instance, focus on the proliferation of the 5G conspiracy in relation to the Coronavirus pandemic and examine Facebook data collected via Crowdtangle,²⁵ underlining the importance of investigating the dynamics on other social platforms, such as the micro-blogging platform Twitter. Using data from Twitter, Ahmed et al. (2020b) analysed social networks of 5G activists, identifying the lack of a clear authority contradicting anti-5G truth claims by examining profiles using the “5GCoronavirus” keyword and #5GCoronavirus. However, limiting the analysis to only these keywords make it difficult to draw conclusions about how conspiracy narratives and their communities might affect broader Twitter discussions such as on the so-called lockdown measures. Moreover, existing literature has not examined the intersections between different conspiracy narrative communities and the principles of social movements, which we relate to our theoretical background of conspiracy narrative communities as disinformed social movements. Given the centrality of political figures and public health bodies in active social networks contemplating COVID-19 and our chief concern with trust in public institutions, we have instead opted for wider selection criteria when building our network for analysis, including a greater variety of relevant hashtags (Yum, 2020). In this way, we hope to posit a broader set of claims about the changing structure of conspiracy theories, the implications for institutions that place a high value on public trust, and the ultimate potency of conspiracy narratives and disinformation during crises such as the COVID-19 pandemic.

4.3. Methodology

Twitter has become a crucial tool in public and political communication since it is used globally by politicians, journalists and citizens who interact in open and direct conversation; however, it is also recognised as a major forum for health misinformation with regards to conspiracy theory during COVID-19 (Ahmed et al., 2020a; Chen et al., 2020; Darius and Stephany, 2020; Kouzy, 2020; Rao et al., 2020; Shahi et al., 2021). Given this two-sided

²⁵ Crowdtangle, <https://www.crowdtangle.com/>

characteristic and our interest in the congruence of misinformation and public health communication, our analysis will focus on Twitter as a major platform for public communication by government institutions and international organisations. Our approach combines quantitative assessments based on message frequencies and network properties with qualitative content analysis to get a qualified impression of content shared in the examined networks. We also introduce material such as screenshots of content to give the reader an impression of the characteristics of different content types and messages referring to broader conspiracy narratives.

Our research benefits from prior studies that analysed the networked architecture of social platforms to determine political polarisation strategies (Darius and Stephany, 2019) or the dynamics of conspiracy narratives and the communities that are active in spreading misinformation (Ahmed et al., 2020a; Bruns et al., 2020; Rao et al., 2020). Social network analysis provides a tool kit to visualise and analyse the connections between social media users that allow to inspect to the formation of network clusters as online communities (Conover et al., 2011; Yang et al., 2016; Scott, 2017). Working from this perspective, we explore the scale, density and interconnection between communities before (T₁) and after (T₂) the lockdown policy measures that were communicated by the UK government on March 23, 2020. We aim to draw conclusions about the evolution of the target conspiracy narratives on Twitter and the potential threat that these pose in the current communication environment. Our methodological approach is similar to Bruns et al. (2020) who use a mixed methods approach based on qualitative and quantitative analysis such as time series analysis, network analysis and a qualitative in-depth reading of messages and profiles to determine content differences between network clusters. While their approach examines the evolution of the 5G conspiracy only, our analysis assesses the growth of the 5G and anti-vaccination conversations in relation to COVID-19 and the relation to broader discourses on UK policy measures referred to as 'lockdown' (#lockdown). Figure 6 summarises our methodological approach and the workflow, which is further elaborated in the following paragraphs.

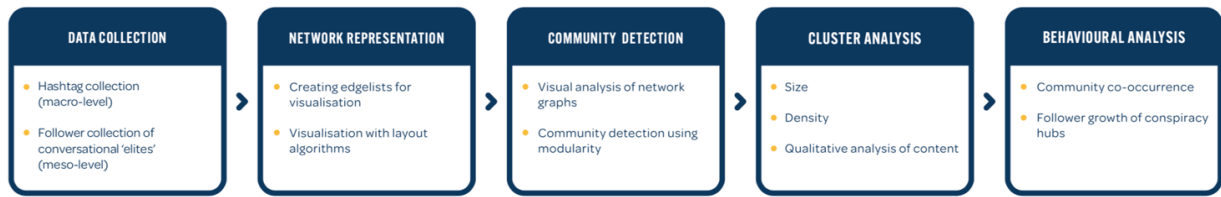


Figure 6: Schematic diagram of the research design

Datasets	Unique accounts	Total Tweets	Avg. Tweets	Avg. RTs
#Lockdown	676,059	1,606,908	2.377	0.643
#5G	34,777	79,558	2.288	0.764
#Vaccines	29,354	50,540	1.554	0.784
#WHO	25,848	45,607	1.764	0.936
#BillGates	15,105	28,997	1.920	0.937
#Plandemic	13,001	18,845	1.450	0.641
#DavidIcke	4,810	7,137	1.484	0.811

Table 1: Summary statistics for the selected hashtag discourses

In order to properly frame our research in the current health crisis and in the hope of drawing conclusions about the impact of COVID-19 on the structure of sceptic communities and conspiracy theories, we focused our analysis on a set of hashtags relevant to our research objective. Widely reproduced on Twitter, the performative function of slogans such as “5G Kills” and “Plandemic” is most notable where they circulate as hashtags, signalling identification with a particular sentiment or community. Since it takes considerably less effort to retweet a post than it does to paint a sign and attend a demonstration, these digital placards lower the barrier of participation in protest action. Table 1 gives an overview of the samples of our chosen hashtags collected during the observation period between 1st January and 10th June 2020. These seven hashtags were selected because of their varying emphases, covering the broad discourse around lockdown as a public health intervention (#Lockdown); targeted discussion topics around technologies that form a core part of notable conspiracy theories (#Vaccines and #5G); specific references to individuals and institutions that are prolific in international efforts to combat COVID-19 and consequently subject to allegations that resonate with our chosen conspiracy theories (#WHO and #BillGates); and two narrower examples with a direct connotation of conspiracy theory (#Plandemic and #DavidIcke) for comparison.

For the operationalisation of the data collection, we accessed Twitter's application programming interfaces (API), mainly with the commercial media listening software Meltwater that also allows the retrospective collection of hashtag discourses. We also used Rtweet and TwitteR packages for the programming software/language R to collect the followers of selected accounts by accessing Twitter's free REST API access.²⁶

4.3.1. Network representation

Many processes and connections can be modelled as networks and in particular, the growth of internet-based communication technologies during the past two decades produces vast amounts of networked data. Our methodology, for instance, makes use of the fact that communication on Twitter is networked by design, since features like retweeting or mentioning or following a profile creates a link between two users. In social network analysis this link is called an edge, while the individuals are called nodes, which when aggregated enables graphical visualisation and statistical analysis of the emerging networks (Yang et al., 2016; Scott, 2017).

To answer our research questions the collection approach focuses on the retweet networks illustrated as the middle level in Figure 7. Despite a trend for proclaiming "RT != endorsement" in Twitter bios, a RT is more likely to represent support or advocacy than an @mention or QT, both of which may express criticism, and we focus on RTs as these often indicate agreement and, thus, in aggregate illustrate a hierarchy of information diffusion and ideological support (Conover et al., 2011; Metaxas, 2017). Consequently, the network representation enables an analysis of discourse changes after the initiation of the lockdown measures and facilitates the automated detection of communities that will be introduced in the following section.

²⁶ Twitter Developer, <https://developer.twitter.com/en/docs/twitter-api>

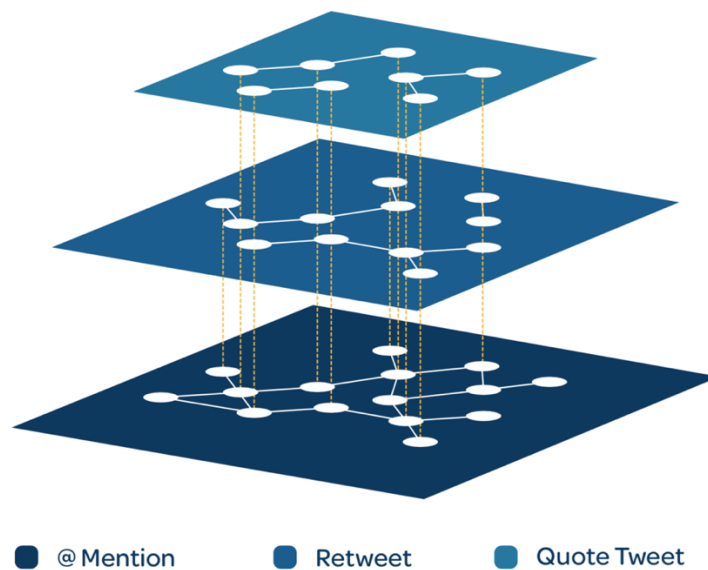


Figure 7: Conversational network types on Twitter

4.3.2. Community detection in network graphs

In networked communication on social platforms such as Twitter, communities form due to different retweeting or mentioning behaviour of users, since not everybody retweets everybody (Boyd et al., 2010). Recent research has underlined the importance of conducting disaggregated analysis of different message types and we decide to focus on retweeting, as the most commonly used messaging type on Twitter (Shugars et al., 2021). Distinct communities in hashtag discourses and in particular retweet networks can represent ideological alignment or at least shared opinions on a political topic such as the so-called lockdown measures designed to reduce the spread of COVID-19 (Conover et al., 2011). The networked structure of the data allows the interpretation and visualisation of centrality measures such as indegree and outdegree (Wasserman and Faust, 1994). The number of times an account is retweeted is represented as incoming edges and measured by weighted indegree, whereas the opposite, outgoing edges are the number of times an account retweeted another account (Bastian et al., 2009). For the visualisation we use Gephi to apply a network graph layout algorithm, ForceAtlas2, and a modularity-based community detection, that assigns each node to a community (Blondel et al., 2008; Bastian et al., 2009; Jacomy et al., 2014). We select modularity-based community detection, since the method is computationally inexpensive and integrated in Gephi (Newman, 2006; Bassett et al., 2013).

We also opt for a resolution value of 5.0 to highlight larger communities in our network visualisation (Chen et al., 2014). This modularity group assignment is used to colour nodes based on their community membership in the network graphs and to visualise and investigate the clustering in more detail for the next steps of analysis.

4.3.3. Qualitative content analysis of detected communities

After the automated detection of communities based on the modularity algorithm, we apply qualitative content analysis (White and Marsh, 2006; Krippendorff, 2018; Mayring, 2014; Selvi, 2019). In conversational networks from social media data, activity is often unevenly distributed, thus, we focus on the network hubs that others most often engage with to get an impression of the content and ideological alignment of different clusters. To give the reader an impression of the sort of content shared in conspiracy communities we include screenshots of messages and describe our observation of differences between communities that are visualised in the network graphs.

4.3.4. Co-occurrence between hashtag discourses

While some accounts might only appear in one hashtag discourse, others can appear in multiple debates and sometimes even in very similar communities, which can be an indication of strategic behaviour and a high level of ideological alignment in these groups (Conover et al., 2011; Darius and Stephany, 2019; Knüpfer et al., 2020). This co-occurrence in retweet network clusters can happen when accounts are retweeting the same account or accounts from the same cluster. We investigate the co-following behaviour of hubs that act as conspiracy theory influencers or super-spreaders and assess whether some groups are likely to appear in multiple conspiracy communities in the selected hashtags, which would strongly indicate an involvement in supporting the propagation of these trust-undermining narratives. The evolution of this “hard core” of overlapping interests forms a central part of our hypothesis that the communications challenge posed by conspiracy theories online is increasing.

4.3.5. Assessing follower growth of network hubs

In order to retrospectively assess the following behaviour of accounts on Twitter we partly adapted an innovative research method that makes use of the chronological order of lists of followers and friends on Twitter (Garimella and Weber, 2017). Using this approach, we were able to determine the absolute and relative follower growth of the identified network hubs not just during the complete observation period, but also before and after the introduction of the UK lockdown on March 23, 2020.

4.4. Analysis and findings

In our literature review, we began to frame our understanding of conspiracy theory from the existing literature in a broader conversation around social movements (specifically on social networking sites) to reflect a concern about the disruptive influence of these networks in the current communication environment. This section presents the results and findings of our analysis in the context of this theoretical framework.

4.4.1. Exploring a spectrum

While a social movement is a neutral theoretical description, the notion is often taken to be inherently positive due to the progressive impetus that many social movements express in their efforts to change society. The line between scepticism and conspiracy thinking is admittedly unclear and accusations against elites, public health organisations and other public institutions run to varying degrees arranged on a spectrum, from contrarian opinion or dissent to more radical, and in our view more dangerous, truth claims. At one end of the spectrum, we find that many social media users endorse and promote the theory of alleged WHO complicity in a cover up of COVID-19 in the virus' early stages Figure A3 (Post1), which is representative of a broad narrative circulating online. Further, many of the top profiles in the various networks often publish content that contains reasonable criticism of, for example, the update to UK guidelines on mandatory face masks Figure A4 (Post 2). However, this kind of libertarian reflex is often framed alongside more outlandish claims about "nefarious intent" on the part of global financial and knowledge elites, such as Bill Gates Figure A5 (Post 3). Highlighting the principle of totality in the conspiracy theory belief system, the connections drawn between various conspiracy theories can run to an

impressive degree: Figure A6 (Post 4) covers New World Order, Bill Gates, 5G, George Soros, and the Epstein scandal, all under the cover of concerns regarding chemtrails. Similarly, Figure A7 (Post 5) combines anti-5G and anti-vaccination rhetoric with an extra layer of anti-elite or “deep state” style conspiracy as expressed in mentions of WHO, Bill Gates and George Soros. We recognise this potential “dark side of social movements,” of which conspiracy theory is one example, and work from the principle that entering particular belief systems may lead to an increased intolerance towards epistemic adversaries, naturally undermining trust in the public institutions that represent the consensus view. The following section outlines an interpretation of the data, drawing on qualitative, quantitative and mixed-methods to locate our findings from multiple perspectives and building to a discussion in which the implications of this research is considered in a wider context of policy and strategic communications during the time of COVID-19.

4.4.2. Mapping networks before (T₁) and after (T₂) the introduction of the 2020 UK Lockdown

Due to the rapidly changing context following the UK government lockdown, we decided to compare data sets before and after 23rd March, giving us Time 1 (T₁) of 1st January to 22nd March 2020, and Time 2 (T₂) of 23rd March to 10th June 2020. We found a dramatic acceleration in the activity around our selected hashtags. Figure A8 illustrates the very large-scale Lockdown retweet networks that represent the conversation in T₁ and T₂. The RT network grew from 36,702 to 267,770, increasing by 630%. We also observed that the average activity of users increased, rising from 1.07 tweets per account in T₁, to 2.94 tweets per account in T₂. Moreover, we named the clusters based on a qualitative content assessment of most retweeted messages and accounts in each of the network clusters as stated in Table 3. Comparing the time periods before and after the UK lockdown was announced on 23rd March yields further interesting results about the changing proportion of conspiracy theory clusters relative to the wider hashtag discourse in which they sit. In T₁, the #5G network was largely controlled by a coalition of two dominant clusters (see Figure 8). The largest community, accounting for 33.4% of the RT network and highlighted in teal, contained mostly tech enthusiasts and researchers and showed some intermingling with a cluster (13.5% and highlighted in red) labelled UK Telecoms.

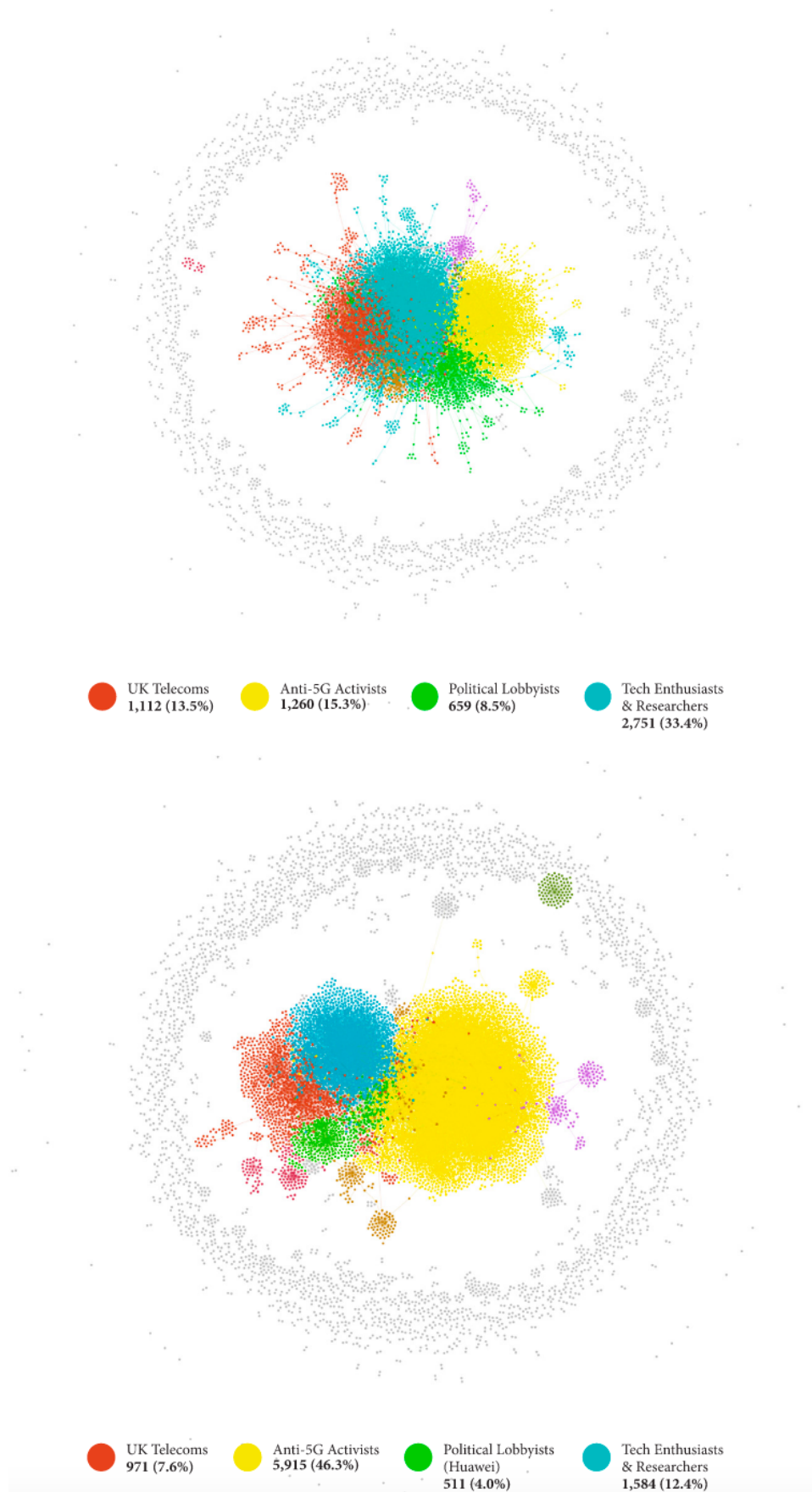


Figure 8: Retweet networks of #5G in T1 and T2 with cluster names

Together, these formed a professional community that appeared to be relatively coherent and organised in the network visualisation. These prominent civil society voices promise to be able to develop a credible response to misinformation and disinformation. The clusters characterised by a professional interest in 5G included active Twitter profiles that have been playing that role with some degree of success, acting as network hubs around which a lively community has developed. However, while we expected to find that the absolute volume of profiles advocating conspiracy theory may have grown, it was surprising to find that conspiracy clusters also grew as a proportion of their respective networks.

Cluster	Mean in T1 (Median)	Mean in T2 (Median)	Mean increase in T2
Conspiracy	3,877 (2,034)	12,133 (5,007)	313%
Civil Society	2,905 (1,118)	7,214 (4,085)	248%

Table 2: Mean follower numbers and increase of conspiracy and civil society network hubs

This suggests that representatives of the civil society discourse that opposes controversial truth claims have been less successful in articulating their position, which has resulted in both #Vaccines and #5G being further penetrated and increasingly characterised by conspiracy theory. This metric can only serve as a proxy, but the fact that conspiracy theory communities are growing at this rate suggests that widespread trust in the social narrative has been lacking, otherwise we would expect to see a similar effect within civil society clusters, such as the pro-vaccine lobby. That being said, an influx of new participants in the conversation are unlikely to have strong connections to any one group. As we show later, community cohesion within conspiracy groups tended to deteriorate later in the period, implying that while controversial truth claims may be an attractive 'hook' for social media users engaging in new topics, they may not be sufficient to build a lasting affiliation.

4.4.3. Network hubs and follower growth

We speculated that one potential reason for this growth in relative size was the significant increase in public attention to these discussion topics, but also that key network hubs in the conspiracy theory clusters have been adept in reaching new audiences. Earlier in this section we have seen that most selected debates and hashtags emerging during the pandemic increased in activity after the UK lockdown was introduced; it was also interesting to assess how many followers the identified conspiracy hubs gained during that period. When a new

hashtag related to a socio-political issue emerges on Twitter, accounts of all sizes seek to capitalise on the momentum derived from increased popular attention and try to link to the discourse to get attention and increase their follower numbers.

Hashtag	Conspiracy	Civil Society
#Lockdown	4.8	3.2
#5G	2.3	4.2
#Vaccines	3.9	1.4
#BillGates	2.6	1.2
#WHO	1.7	1.3
#Plandemic	2.0	1.0
#DavidIcke	1.6	1.1

Table 3: Average weighted indegree for main pro-conspiracy and civil society clusters in the RT networks

The results of our analysis are displayed in Table 3 and underline the increased attention to these accounts, or conspiracy hubs, after the introduction of the lockdown on 23rd March. For key conspiracy theory hubs, the follower growth rate increased by 313% after the initiation of the lockdown, significantly more so than the comparable acceleration for civil society representatives. For both groups we saw quite large differences between the average and median, indicating the presence of outliers such as large news outlets; we therefore focused on accounts with fewer than 100,000 followers at the beginning of 2020. Consequently, the increased visibility of conspiracy theory and scale of the conspiracy clusters can partially be explained by a growth in supporters/followers of conspiracy content creators. These findings are in line with our expectations and contribute to the overall impression that conspiracy theories have significantly gained momentum on Twitter, and potentially other social platforms, during the Coronavirus pandemic.

4.4.5. Network interaction and community building

One of our central research objectives was to assess the strength of the relationships within relevant conspiracy groups. Since community cohesion naturally amplifies content by building a pseudo-echo chamber and simultaneously increasing the opportunity for ideas to circulate, this contributes to a broader picture about the evolution and potency of target conspiracy theories online. The same holds true for the civil society response: an interactive approach to sharing and promoting content is a necessary means of constructing a

community which can offer a strong foundation for coherent, impactful messaging campaigns. After running a community detection algorithm to differentiate between clusters of individual profiles in the RT network for each relevant hashtag, we used average weighted indegree to assess the density of each cluster. A higher average indegree - a proxy for density that represents rate of incoming RTs - indicates stronger or more frequent connections between individual accounts; conversely a lower average indegree points to weaker ties binding each cluster together. The results for the both periods are shown in Table 4. The #5G network stood out as being the only monitored hashtag discourse in which the main civil society community was more cohesive than the conspiracy cluster. As discussed above, #5G includes notable representation from professionals working in the telecoms industry and tech enthusiasts with an observable passion for new technologies, such as 5G. Crucially, these groups of accounts appear to have been successful in generating community cohesion, a likely result of the high degree of interaction that occurs between individuals with a shared interest online.

#5G∩#Vaccines	T1	T2	Difference
Conspiracy	23.48	31.57	8.09
Civil Society	0.57	1.67	1.10

Table 4: Proportion (in %) of conspiracy and civil society cluster members that used #5G and #Vaccines between T1 (January 1, 2020 – March 22, 2020) and T2 (March 23, 2020 – June 10, 2020)

However, the standout trend shows that most of the pro-conspiracy clusters recorded a higher average indegree than their respective civil society clusters, meaning that pro-conspiracy communities tend to be more tightly linked than the communities of researchers and civil society representatives offering a viable information response. The largest anti-vaccination community in the Vaccines discourse, for example, was just under 3 times more densely connected than the pro-vaccination opposition. This is further illustrated in the T1 #Vaccines RT network, shown in Figure A9. The largest clusters representing the anti-vaccination movement (in yellow) and the main pro-vaccine lobby (in green) are almost identical in terms of the proportion of the RT network, accounting for 20.9% and 20.4%, respectively. However, note that the anti-vaccination cluster is dense and tightly organised, whereas the main pro-vaccine group is comparatively sparse and disconnected.

Examining the change between T₁ and T₂, we found that density often decreased in the conspiracy communities as a large influx of new participants in the conversation resulted in a diluting effect on cohesion within clusters. This is in stark contrast to the effect on civil society clusters, most of which recorded a slight increase in density. Without a full qualitative assessment of content, it is difficult to be certain about the reasons for these trends. However, an ad hoc reading of the posts in our data sets suggests that the increasing volume and proportion of the conspiracy element had something of a mobilising effect among researchers, politicians, policymakers and public health officials. It appears that these network hubs in the civil society clusters rallied to promote the scientific consensus and rebut the conspiracy theory truth claims that had begun to benefit from surging public attention.

When exploring the relevant hashtag discourses and analysing the network clusters, we had the impression of a high level of interconnectedness and interaction, especially since some key accounts were recorded among the most retweeted profiles across several networks. Consequently, we next examined the actual overlap between the networks to expose the proportion of users that co-occur in the conspiracy or civil society clusters, which would indicate a polarisation of the observed debates on the pandemic and policy measures. Key profiles across the target networks retweet and interact with one another regularly, undergirding a community of advocates across various conspiracy theories and hashtag networks. One of the main discussion topics which acts as a vector linking various discourse is Bill Gates' work in developing vaccine technology. Bill Gates' designation as the "voodoo doll" of conspiracy theorists during COVID-19 has been well-documented and while the proportion of the RT network using #BillGates in a conspiratorial sense is surprisingly large, this particular hashtag seems likely to attract a specific category of truth claim that features across various different discourses in a kind of cross-pollination driven by network hubs, or profiles that amplify material within the wider network. This kind of interaction and cross-pollination of ideas between different hashtag discourses represents a quantitative relationship between various communities that we explored by establishing the degree of overlap, or the proportion of accounts that feature in two or more of our target RT networks. In order to maintain focus on profiles that are proponents of conspiracy theory, we contained this portion of the analysis to RT networks on the assumption that these

would be more likely to elicit support (whereas mentions and QTs allow the possibility to express criticism).

Figure 9 shows the proportion of Twitter users in each hashtag discourse that co-occur in the #Lockdown discourse. In other words, it represents the degree of overlap between the discussion around lockdown and the narrower conversations we tracked alongside.

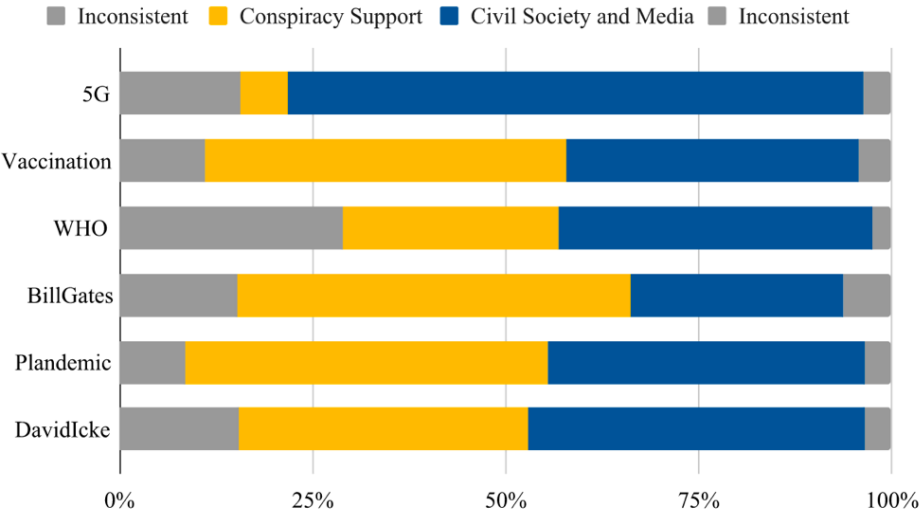


Figure 9: Proportion of individual cluster nodes co-occurrence

A strong connection between WHO and lockdown measures is to be expected, given that WHO data is often a reference point for governmental decision-making. However, it is striking that almost two thirds of those Twitter profiles that have posted #WHO also posted #Lockdown during the reporting period, which implies a high level of perceived association between the WHO and governments’ interventions. It was also striking to find that almost one-third of 5G users also feature in the Lockdown network, as these keywords are not obviously related in the scientific consensus view. While the network overlaps indicate a familiarity of discourses due to similar participants, using both hashtags does not necessarily indicate a belief in conspiracy theory. Since we already identified and labelled clear communities in the networks, we were able to use this information on cluster membership to determine whether an account appears consistently in the conspiracy cluster across various hashtags. We found that the consistent assignments far outweighed

the inconsistent assignments, which means that the Twitter profiles captured in our data set tended to be ideologically coherent in their hashtag use. This holds true both for profiles belonging to a conspiracy theory community and to a civil society community online. We began this report with the expectation that the anti- 5G and anti-vaccination communities have become more connected during the COVID-19 pandemic, which has dominated public discourse in the first half of 2020. For this to be proved true, we would have to have found a higher proportion of co-occurrence between the conspiracy theory clusters in each of the #5G and Vaccines networks later in the period. In T₁, there were 256 accounts in the conspiracy theory overlap between #5G and #Vaccines. In T₂, the absolute number grew to 1,719, an increase of 571%. In relative terms, the overlapping portion between conspiracy clusters in the #5G and #Vaccines networks increased by 8 percentage points. By comparison, the overlap in the civil society clusters grew by just 1pp, suggesting that new hashtag users in the observation period were far more likely to participate in the conversation from a conspiratorial perspective than they were to promote the scientific or civil society consensus.

The data outlined above shows a notable increase in the proportion of profiles engaging in both anti-5G and anti-vaccination discussion on Twitter. The fact that this overlap is growing at a faster rate than the comparable civil society element in these networks is alarming. Together, these findings are representative of the mobilising front of conspiracy theory belief that we recognised in anecdotal terms in our motivation to conduct this research. It is problematic to find that the natural dynamics involved here tend towards a growing and synthesising conspiracy theory community characterised by a fundamental mistrust, rather than a united civil society promoting the scientific consensus. In the concluding section that follows, we summarise our report, link our findings to the theoretical framework, and posit a strategy for consolidating a civil society response that may help to undermine controversial truth claims and shore up trust in public institutions as a result.

4.5. Discussion

This section discusses our main findings, limitations and avenues for further research. We were careful in our analysis not to project too much agency onto these developments. Highlighting the evolution of various conspiracy theory communities on Twitter is not to say that recent developments have been orchestrated or that the resulting community is ideologically coherent. Rather, the changes we have described in scale, density and interconnectedness are likely to be an organic response to the uncertainty that characterises much of public discourse in relation to COVID-19. However, it is problematic to find that the natural dynamics involved here tend towards a growing and synthesizing conspiracy theory community characterised by a fundamental mistrust, rather than a united civil society promoting fact-based arguments. This section discusses our main findings, limitations and avenues for further research.

4.5.1. Discussion of the results

One of the main research objectives guiding this study was to explore how the scale of relevant conversation has changed since the start of 2020. This covers various specific metrics, but the main takeaway suggests that the problem has indeed grown significantly, and has accelerated since the announcement of a UK lockdown on 23rd March. Part of this growth was to be expected, as for example the dramatic increase in activity around the broad hashtag discourse of Lockdown. In some cases, the proportion of key hashtags that is controlled by communities of conspiracy theory believers was alarming. In #BillGates, for example, the vast majority of profiles in the network belonged to a single large conspiracy theory cluster pushing material that asserted the existence of a deep state, elite agenda designed to promote vaccine technology at all costs. In terms of the evolving scale of the problem, the proportion of each network classified as conspiracy theory increased significantly after the 23rd March: the relative size of the anti-5G activist group, for example, grew to almost half of the entire #5G network. Meanwhile, the rate at which key conspiracy theory accounts accumulated new followers also increased, at an even faster rate than many of the profiles offering a response to extravagant truth claims. All of the above points to the evolution of various conspiracy communities during the time of COVID-19 and suggests a kind of organic mobilisation in response to the pandemic. On the

question of density, we found that conspiracy theory communities on Twitter are likely to be far more cohesive than their respective opponent groups. A significant cluster of anti-vaccination activists, for example, was found to be 3 times more likely to engage in interactive behaviour within their cluster than the pro-vaccination lobby against which they have mobilised during the COVID-19 pandemic. Although the density of the conspiracy theory clusters in both the #5G and #Vaccines networks decreased following the implementation of a UK lockdown, this is to be expected given the influx of new users engaging in these discussion topics. Reason for more concern is the apparent sparseness of groups of researchers and other representatives of civil society, which are best placed to form a united response to misinformation and disinformation on social media channels. The key learning from the professional network of 5G advocates and researchers is to build a culture of interaction and reciprocity to underline and amplify messaging that is more friendly to the pro-science, pro-evidence worldview. In our interpretation, generating a cohesive community response to questionable truth claims is an important strategy to help build trust in public health messaging online. Despite the overall disparity evident in our analysis, it has been encouraging to find that civil society communities were successful in maintaining and even increasing cohesion across the time period. An information frontier, composed of interactive relationships between a variety of actors including organisations and individual voices promoting 'good information', is necessary to counteract the 'bad information' that currently benefits from a high degree of cohesion. Lastly, we found that discourses are linked by network hubs who, on the conspiracy theory side, introduced conspiracy theory material to new audiences by transplanting hashtags into new conversations on Twitter. This tendency to penetrate new debates is representative of the potential for conspiracy clusters to polarise broader social and political discussion in charged or controversial communication environments, and is therefore an important insight and strategic consideration with regards to trust in public institutions. We also highlighted an increasing interconnectivity between the target groups of 5G and anti-vaccination activists. This development confirmed our hypothesis, which speculated the emergence or consolidation of a popular front of conspiracy theory, characterised by a fundamental mistrust in public institutions, uniting in response to the COVID-19 pandemic.

4.5.2. Discussion of conspiracy theories as social movements

Our conceptual framework started with the literature on social movements, working on theoretical contributions which suggest that there may be a “dark side” to some such movements. In the context of our research into digital conversations around COVID-19, we were keen to explore the different styles of online engagement that are expressed by conspiracy theory communities versus the civil society groups that contradict their truth claims. Throughout our analysis sections, we described how three key characteristics of social movements described by Touraine (1995) are reflected in the behaviour of selected conspiracy theory communities.

Principle of identity: recasting this slightly as a measure of community identity, we found that 5G and anti-vaccination conspiracy groups have been adept at consolidating their ‘ingroup’ during the COVID-19 pandemic. This was evident as a measure of a) scale, as conspiracy theory clusters grew over the reporting period, and b) density, as these clusters were consistently more cohesive than their civil society counterparts, with the notable exception of an active professional network of tech enthusiasts and telecommunications industry representatives in the 5G network.

Principle of opposition: the RT networks explored in this report all contained a significant conspiracy theory element that was factually opposed to one or more civil society clusters. We interpreted this as being representative of epistemic polarisation, which in anecdotal terms was found to feed a sense among conspiracy theory elites that their truth claims were therefore validated. Where conspiracy theory elites were found to have breached community guidelines and were removed as a result, their supporters simply interpreted this as further evidence that they were ‘onto something’.

Principle of totality: this framing, or tendency to fit all new information to a very rigid worldview, shows that conspiracy theory belief can come to dominate an individual’s interpretation of the world around them. In our section on interconnectedness, we described how some profiles were highly likely to participate in multiple conspiracy theory discourses; this strength of connection between (as well as within) different conspiracy theories highlights the risk that this mode of thinking can overlap across various different discussion topics, with the potential to seed new dissent and mistrust.

Given these observations, we conclude that the conspiracy theory reaction to the COVID-19 emergency is emblematic of a broader epistemic crisis, in which normal contrarian opinion has been appropriated and accelerated by conspiracy communities. The civil society response to these developments has been inconsistent, partly because of the difficulty in coordinating such a response in an organic fashion. However, some of the key learnings from this analysis point to a central and necessary communication principle that should guide messaging strategy in relevant organisations.

2.5.3. Limitations and further research

Our methodological approach builds on modularity-based community detection and the qualitative assessment of the content to examine differences between the detected communities come as any method with a number of limitations. For instance, modularity values and respective cluster detection vary slightly when repeated. Due to this variation the reproducibility of the research has its limitations. However, we have publicised the edge lists of the analysed retweet networks on one of the authors' GitHub account¹² to allow for the replication of our approach and potentially a comparative analysis of the results for various community detection algorithms to test the limits of modularity maximisation for community detection (Lancichinetti and Fortunato, 2009, 2011). Regarding the qualitative assessment, the results of our interpretation could be biased by the authors' opinions. Consequently, we integrated some content examples in the article, but due to resource restrictions and a relatively clear separation of community ideologies we refrained from having annotators double-check the community assessments. Moreover, a number of accounts and especially anti-vaccination influencers were deleted or deleted themselves during the observation period. This might be a result of Twitter's actions against social bots and public health misinformation on their network as indicated by prior research (Davis et al., 2016; Ferrara et al., 2016; Ferrara, 2020). The activity of social bots could be a confounder to our results or interpreting them as a representation of human behaviour. However, Twitter turned more active against inauthentic behaviour, especially with regards to health misinformation on vaccinations and the Coronavirus. While any approach comes with inherent limitations, we would like to emphasise the benefit of our mixed methods approach.

More research needs to be conducted on links between large social platforms and messenger services, since after the deplatforming of influential figures their supporters have often transferred to messenger services like Telegram, Signal or WhatsApp as a reaction to the increased content moderation (Jhaver et al., 2021). We hope our approach may support the identification of network hubs and conspiracy narrative communities in research and enable debunking conspiracy narratives directly and effectively in practice (Vijaykumar et al., 2021). These direct debunking campaigns can help to foster trust into public and multilateral institutions and consequently strengthen the basis for effective public crisis communication.

2.5.4. Conclusions

As of June 2021, the COVID-19 pandemic has resulted in over 176 million cases and upwards of 3,800,000 fatalities worldwide; upsetting the global economy, it has amounted to a downturn of historic proportions. In these turbulent times, much emphasis has been rightly placed on the importance of clear and effective public health communication amidst surging levels of information. This article aimed to better understand one of the main factors disrupting this delicate information environment. Conspiracy theories or narratives as social movements - specifically the anti-5G and anti-vaccination movements - contradict social narratives with spurious truth claims, undermine public health messaging, and ultimately play a role in deteriorating public trust in the institutions whose role it is to safeguard citizens' well-being and navigate our societies through the current epidemiological crisis. Anecdotally, this appeared to be driven by hostility to elites and the institutions they represent - a fundamental mistrust emerged as the key uniting factor in this particular community. Our analysis showed a notable increase in the proportion of profiles engaging in both anti-5G and anti-vaccination discussion on Twitter. The fact that this overlap is growing more than the comparable civil society element in these networks is concerning and might help to make sense of sometimes violent street protests against Coronavirus restrictions. Together, these findings illustrate the mobilising front of conspiracy theory belief that we recognised in anecdotal terms in our motivation to conduct this research.

References

Ahmed W, Vidal-Alaball J, Downing J, Seguí FL (2020) COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *Journal of Medical Internet Research* 22, e19458. Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. Available at: <https://www.jmir.org/2020/5/e19458>, doi:10.2196/19458.

Anderson B (2006) *Imagined communities: Reflections on the origin and spread of nationalism*. Verso books.

Aupers S (2012) 'Trust no one': Modernization, paranoia and conspiracy culture. Available at: <https://journals.sagepub.com/doi/abs/10.1177/0267323111433566>.

Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, Mucha PJ (2013) Robust detection of dynamic community structure in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23, 013142. American Institute of Physics AIP. Available at: <https://aip.scitation.org/doi/abs/10.1063/1.4790830>, doi:10.1063/1.4790830.

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open-source software for exploring and manipulating networks, in: *Third international AAAI conference on weblogs and social media*.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, P10008. Available at: <http://arxiv.org/abs/0803.0476>, doi:10.1088/1742-5468/2008/10/P10008. arXiv: 0803.0476.

Boyd D, Golder S, Lotan G. (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in: *2010 43rd Hawaii International Conference on System Sciences*, pp. 1–10. doi:10.1109/HICSS.2010.412.

Bruns A, Harrington S, Hurcombe E (2020) 'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia* 177, 12–29. SAGE Publications Ltd. Available at: <https://doi.org/10.1177/1329878X20946113>.

Budhwani H, Sun R (2020) Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. *Journal of Medical Internet Research* 22. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7205030/>, doi:10.2196/19301.

Chen E, Lerman K, Ferrara E (2020) Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance* 6, e19273. Publisher: JMIR Publications Inc., Toronto, Canada.

- Chen M, Kuzmin K, Szymanski BK (2014) Community Detection via Maximization of Modularity and Its Variants. *IEEE Transactions on Computational Social Systems* 1, 46–65. doi:10.1109/TCSS.2014.2307458.
- Cichocka A, Marchlewska M, Golec de Zavala A, Olechowski M (2016) 'They will not control us': Ingroup positivity and belief in intergroup conspiracies. *British Journal of Psychology* 107, 556–576. Wiley-Blackwell Publishing Ltd, United Kingdom. doi:10.1111/bjop.12158.
- Conover M, Goncalves B, Ratkiewicz J, Flammini A, Menczer F (2011) Predicting the Political Alignment of Twitter Users, pp. 192–199. doi:10.1109/PASSAT/SocialCom.2011.34.
- Darius P, Stephany F (2019) Twitter "Hashjacked": Online Polarisation Strategies of Germany's Political Far-Right. Technical Report. SocArXiv. Available at: <https://osf.io/preprints/socarxiv/6gbc9/>, doi:10.31235/osf.io/6gbc9.
- Darius P, Stephany, F (2020) How the Far-Right Polarises Twitter: 'High-jacking' Hashtags in Times of COVID-19. arXiv:2010.05686 [cs] URL: <http://arxiv.org/abs/2010.05686>. arXiv: 2010.05686.
- Davis CA, Varol O, Ferrara E, Flammini A, Menczer F (2016) Botornot: A system to evaluate social bots. *Proceedings of the 25th international conference companion on world wide web*, pp. 273–274.
- Edy JA, Risley-Baird EE (2016) Rumor Communities: The Social Dimensions of Internet Political Misperceptions. *Social Science Quarterly* 97, 588–602. Available at: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ssqu.12309>.
- Einstein KL, Glick DM (2015) Do I think BLS data are BS? The consequences of conspiracy theories. *Political Behavior* 37, 679–701. doi:10.1007/s11109-014-9287-z. Springer, Germany.
- Ferrara E (2020) What types of covid-19 conspiracies are populated by twitter bots? arxiv. Available at: <https://arxiv.org/abs/2004.09531>.
- Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Communications of the ACM* 59, 96–104. Publisher: ACM New York, NY, USA.
- Garimella VR K, Weber I (2017) A Long-Term Analysis of Polarisation on Twitter, 4. *Proceedings of the International AAAI Conference on Web and Social Media*. Available at: <https://ojs.aaai.org/index.php/ICWSM/article/view/14918>.
- Hussain A, Ali S, Ahmed M, Hussain S (2018) The Anti- vaccination Movement: A Regression in Modern Medicine. *Cureus* 10. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6122668/>, doi:10.7759/cureus.2919.

Imhoff R, Lamberty P (2020) Conspiracy theories as psycho-political reactions to perceived power. *Routledge handbook of conspiracy theories*, 192–205. Routledge Oxon, UK.

Imhoff R, Lamberty P, Klein O (2018) Using power as a negative cue: How conspiracy mentality affects epistemic trust in sources of historical knowledge. *Personality and Social Psychology Bulletin* 44, 1364–1379. Sage Publications Sage, Los Angeles, CA, USA.

Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one* 9. Public Library of Science.

Jhaver S, Boylston C, Yang D, Bruckman A (2021) Evaluating the effectiveness of deplatforming as a moderation strategy on twitter.

Jolley D, Paterson JL (2020) Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology* 59, 628–640. Available at: <https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjso.12394>.

Kata A (2010) A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 28, 1709–1716. doi: 10.1016/j.vaccine.2009.12.022.

Keeley BL. (1999) Of Conspiracy Theories. *The Journal of Philosophy* 96, 109–126. *Journal of Philosophy, Inc.* doi:10.2307/2564659.

Knüpfer C, Hoffmann M, Voskresenskii V (2020) Hijacking MeToo: transnational dynamics and networked frame contestation on the far right in the case of the '120 decibels' campaign. *Information, Communication & Society* 0, 1–19. <https://doi.org/10.1080/1369118X.2020.1822904>.

Cureus | COVID-19 Coronavirus Goes Viral: Quan- Misinformation Epidemic. Available at: <https://www.cureus.com/articles/>

Kouzy R (2020) Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. 28976-coronavirus-goes-viral-quantifying-the-covid-19-misinformation-epidemic-on-twitter.

Krippendorff K (2018) *Content analysis: An introduction to its methodology*. Sage publications.

Lancichinetti A, Fortunato S (2009) Community detection algorithms: A comparative analysis. *Physical Review E* 80, 056117. American Physical Society. doi:10.1103/PhysRevE.80.056117.

- Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Physical Review E* 84, 066122. American Physical Society. doi:10.1103/PhysRevE.84.066122.
- Larson HJ (2018) The biggest pandemic risk? Viral misinformation. *Nature* 562, 309–309. Nature Publishing Group. doi:10.1038/d41586-018-07034-4.
- Lewandowsky S, Cook J (2020) *The Conspiracy Theory Handbook*, 12.
- Marchlewska M, Cichońska A, Kossowska M (2018) Addicted to answers: Need for cognitive closure and the endorsement of conspiracy beliefs. *European Journal of Social Psychology* 48, 109–117. doi:10.1002/ejsp.2308.
- Mayring P (2014) *Qualitative Content Analysis*, 144.
- Megret K (2020) Even covid-19 can't kill the anti-vaccination movement. *BMJ* 369, m2184. British Medical Journal Publishing Group, Section: Feature. doi:10.1136/bmj.m2184.
- Metaxas P (2017) Retweets indicate Agreement, Endorsement, Trust: A Meta- Analysis of Published Twitter Research, 20.
- Newman ME J (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 8577–8582. National Academy of Sciences Section: Physical Sciences. doi:10.1073/pnas.0601602103.
- Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG (2020) Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention, 11.
- Rao A, Morstatter F, Hu M, Chen E, Burghardt K, Ferrara E, Lerman K (2020) Political partisanship and anti-science attitudes in online discussions about covid-19. ArXiv preprint. arXiv:2011.08498.
- Schmidtke H (2019) Elite legitimation and delegitimation of international organizations in the media: Patterns and explanations. *The Review of International Organizations* 14, 633–659. doi: 10.1007/s11558-018-9320-9.
- Scott J (2017) *Social Network Analysis*. SAGE Publications Ltd, London, UK. doi: 10.4135/9781529716597.
- Selvi AF (2019) Qualitative content analysis, *The Routledge Handbook of Research Methods in Applied Linguistics*. Routledge.

Shahi GK, Dirkson A, Majchrzak TA (2021) An exploratory study of COVID-19 misinformation on Twitter. *Online Social Networks and Media* 22, 100104. doi: 10.1016/j.osnem.2020.100104.

Shugars S, Gitomer A, McCabe S, Gallagher RJ, Joseph K, Grinberg N, Doroshenko L, Foucault Welles B, Lazer D (2021) Pandemics, Protests, and Publics: Demographic Activity and Engagement on Twitter in 2020. *Journal of Quantitative Description: Digital Media* 1. doi:10.51685/jqd.2021.002.

Sternisko A, Cichocka A, Van Bavel JJ (2020) The dark side of social movements: social identity, non-conformity, and the lure of conspiracy theories. *Current Opinion in Psychology* 35, 1–6. doi: 10.1016/j.copsy. 2020.02.007.

Touraine A (1995) *Beyond Social Movements?* Lyman, S.M. (Ed.), *Social Movements: Critiques, Concepts, Case-Studies*. Palgrave Macmillan UK, London. *Main Trends of the Modern World*, pp. 371–393. doi:10.1007/978-1-349-23747-0_16.

Vijaykumar S, Jin Y, Rogerson D, Lu X, Sharma S, Maughan A, Fadel B, de Oliveira Costa MS, Pagliari C, Morris D (2021) How shades of truth and age affect responses to COVID-19 (Mis)information: randomized survey experiment among WhatsApp users in UK and Brazil. *Humanit Soc Sci Commun* 8(88). <https://doi.org/10.1057/s41599-021-00752-7>.

Wasserman S, Faust K (1994) *Social network analysis: Methods and applications*. *Social network analysis: Methods and applications*, Cambridge University Press, New York, NY, US. doi:10.1017/CBO9780511815478.

White MD, Marsh EE (2006) Content analysis: A flexible methodology. *Library trends* 55, 22–45.

Wolfe RM, Sharp LK (2002) Anti-vaccinationists past and present. *BMJ* 325, 430–432. doi:10.1136/bmj.325.7361.430. British Medical Journal Publishing Group Section: Education and Debate.

Yang S, Keller FB, Zheng L (2016) *Social network analysis*. SAGE Publications, Inc, <https://doi.org/10.4135/9781071802847>.

Yum S (2020) Social Network Analysis for Coronavirus (COVID-19) in the United States. *Social Science Quarterly* 101, 1642–1647. doi: <https://doi.org/10.1111/ssqu.12808>.

Chapter 5 – Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance

Abstract ²⁷

Who governs — and who should govern — online communication? Social media companies, international organisations, users, or the state? And by what means? A range of rhetorical devices have been used to simplify the complexities associated with the governance of online platforms. This includes 'constitutional metaphors': metaphorical allusions to traditional political concepts such as statehood, democracy, and constitutionalism. Here we empirically trace the ascent of a powerful constitutional metaphor currently employed in the news media discourse on platform governance: characterisations of Facebook's Oversight Board (OB) as a 'supreme court'. We investigate the metaphor's descriptive suitability and question its normative and political ramifications. We argue that uncritical characterisations of the OB as Facebook's 'supreme court' obscure its true scope and purpose. Additionally, we argue that appropriating the socio-cultural symbolism and hence political legitimacy of a supreme court, and mapping it onto a different type of actor, poses a threat to responsible platform governance.

²⁷ Cows J, Darius P, Santistevan D, Schramm M (2022) "Constitutional Metaphors: Facebook's 'Supreme Court' and the Legitimation of Platform Governance." *New Media & Society*, April 5, 2022, 14614448221085560. <https://doi.org/10.1177/14614448221085559>.

5.1. Introduction: Platform governance, the language of statehood and 'constitutional metaphors'

An outside observer of the discourses on platform governance might think that the baton of governing the internet has already been passed on to companies—so common are references to 'statehood', 'constitutions', and other allusions to the government-like role and scope of large technology companies in managing online platforms. For example, Facebook's CEO Mark Zuckerberg has described Facebook²⁸ as 'more like a government than a traditional company' concerning its governance role (Foer, 2019). And Facebook's Oversight Board (OB) is, as we will see, frequently dubbed Facebook's 'supreme court'. The OB consists of twenty members contracted by an independent trust established and funded by Facebook, supported by full-time staff. It became operational in 2020, and mostly reviews user appeals against Facebook's content moderation decisions. The OB's scope and powers are established by its charter. It is tasked with interpreting Facebook's existing community standards in light of earlier decisions and the human rights implications of content removal. Although Facebook is undoubtedly a frontrunner owing to its scale, the language of statehood and constitutionalism arises in broader debates over the private governance of digital technology. In 2018 the normative practices of social media companies were characterised as 'platform law' by a United Nations Special Rapporteur (Kaye, 2018).

But can we simply map the language, practices, and institutions associated with statehood, constitutionalism, and the rule of law to the digital sphere—particularly since many of the most powerful 'governing' actors in this space are private companies (Gillespie, 2018b; van Dijck, Poell and De Waal, 2018; Gorwa, 2019)? And what might be the unintended consequences of doing so? Drawing on metaphor theory (Lakoff, 1992; Lakoff and Johnson, 2003) and theories of legitimacy and legitimation (Suchman, 1995), this article identifies *constitutional metaphors* as a powerful rhetorical device in the media discourse about platform governance—shedding light on the use of one particular metaphor, that of a 'supreme court', to characterise the Facebook Oversight Board. Taking a constructionist approach, we investigate how constitutional metaphors iteratively construct new meaning

²⁸ During the drafting of this article, the parent organisation of Facebook (and subsidiaries including Instagram and WhatsApp) changed its name to Meta Platforms. We retain references to "Facebook" throughout for parsimony and historical accuracy. The OB decides 'cases' related to Instagram and Facebook.

and, in turn, how they may even alter the underlying meaning of existing constitutional concepts, in the context of platform governance. Previous work has critically investigated the use of metaphors to characterise and understand various digital phenomena (Lindh and Nolin, 2017; Ganesh, 2020; Simon and Camargo, 2021; Slupska, 2021). Separately, scholars have also assessed metaphorical understanding of laws and constitutions (Tribe, 1987; Brooks and Gewirtz, 1996; Berger, 2009, 2013; Cloutier, 2019), yet this existing work looks at the metaphorical narration of actual constitutions—not at the use of constitutional language to characterise private governance. Here, by contrast, we empirically trace and critically reflect on the ascent of one particular constitutional metaphor, that of a supreme court, in the context of Facebook’s private governance power.

The social and political stakes of constitutional metaphors in the context of private ‘platform’ governance are high. If the OB continues to be characterised publicly as a ‘supreme court’, this may accord it the legitimacy associated with a conventional supreme court, as one branch of government in a democratic system. This legitimacy may, in turn, extend to Facebook itself, reinforcing the perception of Facebook as, in the words of its founder, ‘like a government’ (Foer, 2019)—with weighty implications for democracy and the rule of law. Although different people may mean different things when referring to such notions, it is commonly understood that constitutionalism and the rule of law in principle refer to authority that has been legitimated through public consent (Taylor, 2021). Constitutional metaphors in platform governance, however, establish a novel connection between old concepts and new practices, with the effect of legitimating private institutions through association with public governance institutions. Moreover, an equivalent phenomenon may emerge in reverse: eventually, the metaphors’ underlying constitutional concepts may be a ‘re-conceptualised’ and common understanding of the metaphor’s referent—in our example, supreme courts—may change in turn. Thus, the question at hand is whether it could—or should—be accepted that the notion of a supreme court extends beyond democratically legitimated public authority.

The article progresses in three overarching steps. Firstly, we lay theoretical and conceptual foundations and provide background to our case study. In Section 5.2.1., we root our research in metaphor theory and define constitutional metaphors as an analytical concept. In Section 5.2.2., we introduce the constitutional metaphor at the heart of this study—the

metaphorical depiction of Facebook's Oversight Board as a 'supreme court'—and explain the dilemma that its use creates. And in Section 5.2.3., we use political legitimacy and legitimation theories to sketch out the normative and sociopolitical significance of constitutional metaphors, noting their potential to bolster the reputation of platform companies' governance. Secondly, in Section 5.3., we undertake our empirical analysis. Here, we empirically trace the rise of the OB-as-supreme-court metaphor in media discourse. We find that the use of this metaphor in the media has grown over time. It has become somewhat more reflective, yet significant ambiguities concerning its underlying meaning remain. Thirdly, in Section 5.4. we consider the normative and performative implications of our empirical findings for platform governance, expounding the possible consequences of such metaphors' discursive dominance if used uncritically. We conclude with final remarks in Section 5.5. considering the risks of constitutional metaphors blurring public perception and constructing new meanings, and identify areas for future research.

5.2. Background

This section provides the background and underlines the importance of metaphors in public discourses, especially with regards to democratic institutions and constitutional democracy.

5.2.1. Metaphors and the language of constitutional democracy

The essence of metaphors is 'understanding and experiencing one thing in terms of another' (Lakoff and Johnson, 2003:5). Metaphors map 'one conventional image onto another ... lead[ing] us to map knowledge about the first image onto knowledge about the second' (Lakoff, 1992:27–28). In doing so, however, metaphors always highlight some aspects while hiding others. Metaphors can be particularly influential when our knowledge about one concept—e.g., how a complicated piece of software functions—is minimal and our understanding of the second concept that is 'mapped' onto the first is particularly vivid. In such cases, 'metaphors are not only used to make sense; they also function as constitutive' (Lindh and Nolin, 2017:164). This may be particularly impactful in the case of novel digital phenomena, where, as Lindh and Nolin have argued, the 'functioning and usage [of new technologies] are frequently quite abstract', and where we may therefore 'lack an existing vocabulary to fall back on' (2017:166). Metaphors may thus fill this vacuum of vocabulary.

Examples abound of the use of metaphor for making sense of emerging digital technology and its societal impacts. Scholars have explored the use of metaphors in cyberspace governance (Slupska, 2021); the characterisation of autonomous machines as 'computational brains' (Ganesh, 2020), and of Covid-19 misinformation as an 'infodemic' (Simon and Camargo, 2021); and naturalistic references to forms of internet-based software services such as 'cloud' computing or technological 'ecosystems' (Lindh and Nolin, 2017, p. 7). Meanwhile, metaphors are also commonly used to characterise otherwise-abstract legal and political concepts; for instance, Montesquieu's characterisation of judges as 'the mouth of the law' (Montesquieu, 1748:180), or that of the US constitution as a 'living' thing (McBain, 1927).

This brings us to our more specific consideration of the point of collision between metaphors for digital technology and metaphors about the law and constitutions. We define 'constitutional metaphors' as metaphorical allusions to concepts, institutions, or practices of constitutional democracy in discourse about issues other than constitutional democracy itself. Here, our focus is on using constitutional metaphors in discourses about platform governance (Gorwa, 2019). We argue that the use of constitutional metaphors in the platform governance context warrants particular attention, given that the governance practices of platform companies have often been described as untransparent, harmful, or even outright dangerous (Frenkel and Kang, 2021; York, 2021; 'Statement of Frances Haugen', 2021). As we argue, the use of constitutional metaphors to characterise platform governance may unwittingly consolidate this form of private power. Of course, we do not argue that all references to constitutions or statehood should cease. That would be naïve. In fact, legal and political theory holds much potential for platform governance (Kadri, 2018; Klonick, 2020:2457–2466; De Gregorio, 2021). However, transplanting such theories to a new context presupposes detailed analysis and apt terminology. The *uncritical* use of constitutional metaphors provides neither. It is in this context that we undertake our investigation of depictions of Facebook's Oversight Board as a 'supreme court'.

5.2.2. The emergence and spread of 'Facebook's Supreme Court'

The initial idea for—and first metaphorical allusion to—a Supreme Court for Facebook²⁹ stemmed from Noah Feldman, a Harvard Law professor specialising in 'constitutional studies, ... free speech, constitutional design, and the history of legal theory'.³⁰ From the outset, the idea for the OB involved platforms 'setting up their own quasi-legal system'³¹ and even envisaged 'Facebook as quasi-sovereign'.³² Although the charter largely withholds such vocabulary, trivialising its use as owing simply to the early drafting process's euphoria would be inaccurate. The Oversight Board's 'Bylaws', for example, do not refer to customers or users, but to 'People',³³ inevitably alluding to the phrase 'We the People...' which famously opens the US Constitution. Such amalgams of the public and the private, in language and in substance, are no coincidence but programmatic with the Oversight Board, and bring to mind Gillespie's (2010:348) observation that platform companies work 'not just politically but also discursively to frame their services and technologies'.

None of this is to suggest, however, that there is no need for an independent body to undertake an impartial review of Facebook's content moderation decisions. New bodies like the OB that are expert-led and, to a contested degree, independent and accountable (Klonick 2020:2475–86), represent progress away from entirely 'authoritarian' (Douek 2019:9–24; Pozen 2018) modes of platform governance. And such progress is necessarily informed by the political and legal theories that guide democratic governance. Thus, when the executives and legal departments of social media companies consider how best to improve their governance structures in response to the growing challenges of content moderation, they did not see the need to reinvent the wheel, but instead turned to existing models.

Nevertheless, the way things are spoken about affects the way they are understood. And the use of 'supreme court' as a descriptor for the OB—the origin and prevalence of which we

²⁹ See Noah Feldman, A Supreme Court for Facebook, 30 January 2018, p 101; published in *Global Feedback and Input on the Facebook Oversight Board*, Appendix D (p 100-115), see <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf> (last accessed 8 September 2021).

³⁰ See: <https://hls.harvard.edu/faculty/directory/10257/Feldman/> (last visited 22 April 2021).

³¹ See Feldman, Fn 1.

³² See Feldman, Facebook Supreme Court: A Governance Solution, March 2018, p 112 – see Fn 1.

³³ Art. 3, The Oversight Board Bylaws, January 2021, <https://www.oversightboard.com/sr/governance/bylaws> (22 September 2021) (emphasis added).

investigate in Section 5.3.—inevitably brings with it the weighty social, cultural, and political capital that attaches to supreme courts, particularly in the United States, where Facebook was founded and where its parent company Meta is headquartered. The concept of the United States Supreme Court draws its meaning from an intricate mix of political beliefs and socio-cultural practices forged over centuries, from the 1789 Bill of Rights and *Marbury v. Madison* (1803)³⁴ to more recent clashes over civil rights or abortion. The Supreme Court consistently ranks as the most trusted institution of government in the US (Gallup, 2018). And it is made up of individuals who command professional respect. Thus, mapping common knowledge of the US Supreme Court onto the blank canvas of the OB may bolster its image and ultimately, as we suggest in the following section, enhance the legitimacy of the Board and of Facebook itself.

Thus, we face a dilemma. On the one hand, managing content and regulating behaviour on social media requires enforceable rules and, in turn, new mechanisms of governance. On the other hand, although these new mechanisms emulate existing ones, they often lack their archetypes' theoretical grounding, practical safeguards, and legal process. In particular, platform governance has considerably less democratic input legitimation by way of user participation as compared with traditional forms of public governance that rely on some form of public consent (Taylor, 2021). Facebook's only real experiment with mass democracy, which started in 2009, ended in ignominy in 2012 when it was voted out of existence (Leetaru, 2019). Today, the average Facebook user has negligible influence on the site's governance or the content of the Community Standards, which provide the basis for the OB's decision-making. In traditional systems of judicial control over administrative norm-enforcement, democratic influence (or at least accountability) over norm-setting is considered a premise for legitimate rule. Therefore, metaphorical allusions to a supreme court extend to allusions to democracy, which in the case of Facebook are misplaced.

³⁴ Some scholars make the case that the OB engages in 'constitutional review in full swing', hailing its first decisions as 'the *Marbury v Madison* of platform governance' (Gradoni, 2021). We disagree. *Marbury v Madison* deals with whether a court may exercise constitutional review over acts of a democratically legitimated legislative branch. The latter does not even remotely exist in the case of Facebook.

5.2.3. Legitimacy and legitimation in platform governance: the influence of constitutional metaphors

In the empirical analysis that follows, we explore references to the Oversight Board as 'Facebook's Supreme Court' (this metaphor is hereafter referred to as 'OB:SC'). Before turning to this empirical analysis, we sketch the socio-political and normative significance for platform governance of constitutional metaphors such as OB:SC. In addition to *being governed* by states, platform operators themselves also *govern* the content shared and behaviour exhibited on their platforms (Gillespie, 2018b; Klonick, 2018; Gorwa, 2019). 'Private governance' is not novel to social media or the internet; private companies perform governance roles in many other domains (Graz and Nölke, 2007). It is not, therefore, inconceivable that social media companies could play a similar role with respect to platform governance. Yet the use of constitutional metaphors suggests the emergence of something new and distinct: the ascription of a 'quasi-sovereign' role *to* and/or the assertion of a 'quasi-sovereign' role *by* platform operators (Pohle and Thiel, 2020:7). By adopting the functions, practices, and language of public governance, platform operators blur the boundary between private and public governance, and, ultimately, may lay claim to transformed notions of power (Sharon, 2020), which may be exercised in a less democratic or accountable manner (Taylor, 2021).

Such developments have implications for the political legitimacy and legitimation of platform governance. Political legitimacy is a widely debated concept, and several approaches to it have been advanced (Weber, 1978; Buchanan, 2002; Pettit, 2012). These typically fall into one the two camps: normative approaches seek to define criteria by which the political legitimacy of an actor can be judged. In contrast, descriptive approaches portray legitimacy as something that obtains to actors who are seen as having the right or capacity to govern (Peter, 2017). Thus, legitimacy as a concept incorporates both normative considerations about which actors *ought* to be regarded as legitimate (Nagel, 2005; Buchanan and Keohane, 2006; Hassoun, 2012), and empirical considerations about which actors *are* seen as such (Weber, 1994; Suchman, 1995; Tallberg, Bäckstrand and Scholte, 2018). This unresolved conceptual tension is significant when we consider constitutional metaphors in the context of platform governance. It is possible both to debate the criteria against which a platform operator could be outwardly seen as a legitimate 'governor', and

to document attempts by these operators to seek to be seen as legitimate—a process we can refer to as legitimation (Beetham, 2013; Tallberg, Bäckstrand and Scholte, 2018).

Several internet law and communications scholars have explored the possibility and implications of 'legitimate' governance by technology companies from a normative standpoint. Working within the framework of 'digital constitutionalism' (Celeste, 2019; De Gregorio, 2021), Suzor³⁵ and colleagues (Suzor, Van Geelen and Myers West, 2018) adopt a *rights-based* approach, by identifying 'the human rights values that might form an index of the legitimacy of governance [by] online intermediaries'. Suzor also separately (2018a:4) focuses on *procedural values* inspired by the rule of law within states, such as consent and predictability. He argues that although 'the way these principles have historically been applied has been state-centric', these values can be usefully applied to assess the [legitimate] governance of digital media'. Haggart and Keller (2021), meanwhile, propose a framework based on *democratic* legitimacy, which applies Schmidt's (2013) tripartite framework of 'input', 'output', and 'throughput' legitimacy, developed for the European Union, to several instances of private digital governance, including Facebook's Oversight Board. To varying degrees, each of these examples makes certain assumptions about political legitimacy as it relates to private technology companies that may not be warranted. An inclusive human rights-based standard for legitimacy leaves unanswered questions about how tension between different rights should be resolved, as well as failing to consider 'the issue of *which actors* possess rule-setting legitimacy' (Haggart and Keller 2021:4). As a consequence, this risks 'taking global platforms' current private-ordering regime as a given [leaving] no room to consider even a theoretical role for the state' (2021:4). Yet Haggart and Keller's own proposition—to appropriate Schmidt's tripartite framework for private forms of platform governance—rests on the strength of the analogy between private platforms and the EU, which risks conceptual confusion. Further, these contributions share the implicit assumption that private actors can or could hold political legitimacy at all—departing from the common understanding of the state as the sole 'locus of political legitimacy' (Nagel, 2005:113). Those scholars that do accept the possibility of non-state entities, including some private corporations, of being legitimate tend to attach considerable obligations and moral responsibilities to this designation (e.g. Hassoun, 2012).

³⁵ Suzor is a member of the OB.

Therefore, although it is beyond the present scope to resolve whether private companies, as non-state entities, can normatively be considered politically legitimate, it seems that, at a minimum, a cautious stance to the question is warranted.

Descriptive approaches to conceptualising legitimacy, meanwhile, see it as 'open to political manipulation' (Tallberg, Bäckstrand and Scholte, 2018:6). Here, scholars have explored 'the ways in which organisations instrumentally manipulate and deploy evocative symbols in order to garner societal support'; legitimacy in this context is an 'operational resource ... that organisations extract—often competitively—from their cultural environments and ... employ in pursuit of their goals' (Suchman, 1995:572–576). And although descriptive approaches to political legitimacy have tended to focus mostly on nation-states and international and multilateral organisations like the EU and IMF, other scholars have identified similar dynamics in the private sector. Building on Jasanoff and Kim's pathbreaking notion of 'sociotechnical imaginaries' (2009), several recent efforts have assessed the 'imaginaries' constructed by big tech (Mager and Katzenbach, 2021a), in particular by Facebook (Hoffmann, Proferes and Zimmer, 2018; Rider and Murakami Wood, 2019; Haupt, 2021).

This brief review advocates caution regarding the normative application of political legitimacy to platform operators and highlights the possibility for legitimacy to be 'extracted' by platform operators through the appropriation of socio-cultural symbols. We will shortly see that the socio-cultural symbolism of supreme courts has been appropriated in just such a way, to confer legitimacy on the OB itself and, by extension, on Facebook's own role and status as a 'governor'. However, whereas existing studies have analysed corpora consisting solely of the statements of corporate leaders like Mark Zuckerberg, in what follows, we explore a broader database of media coverage of the OB. As a result, our focus is less on a (potentially deliberate) construction, by Facebook, of the idea that the OB is akin to a 'supreme court', and more on the consequences for platform governance when this metaphor reaches a broader news public.

5.3. Documenting the rise of the 'supreme court' metaphor

Based on the theoretical framework developed above, we turn to empirical analysis. We outline our methodology and data selection, then analyze the evidence for the ascent of the supreme court metaphor in the American news discourse about the Oversight Board, identifying several landmark events which shaped the metaphor's proliferation.

5.3.1 Methodology

The empirical focus of our work is concerned with the development and usage of the 'supreme court' metaphor in reference to the OB (hereafter, we refer to the use of the metaphor as 'OB:SC'). In particular, we seek to assess *whether* and *to what extent* the metaphor is widely used, and *how* its use has evolved over time. In this sense, our inquiry is serial. We first establish the existence of the metaphor and, in doing so, produce a dataset, then follow with an analysis of the metaphor's development, paying specific attention to its implications for legitimacy. Our approach resembles an 'inductive' content analysis, meaning our categories are constructed and contextualised iteratively, always oscillating between units of analysis—or in our adopted nomenclature, between the 'micro' and the 'macro' (Hsieh and Shannon, 2005; Mayring, 2014). This distinction does not produce a clean partition; rather, as an analytic tool, it highlights characteristics of the news discourse in general and recurring and particularly notable substantive and linguistic aspects of specific news pieces.

As noted above, our focus is on legitimacy and legitimation. We seek to understand how authors and other speakers respond to—or in some cases, entirely skirt—the "why" question (Van Leeuwen, 2007:93): why use the OB:SC metaphor at all? Though inductive, we make no claims of initial 'neutrality'. Our construction of categories was also suffused by the inquiry of similar studies into metaphors and discourse, all notably acknowledging the role of authority in the construction and use of language (Hoffmann, Proferes and Zimmer, 2018; Ganesh, 2020; Haupt, 2021; Simon and Camargo, 2021; Slupska, 2021). Specific categories are outlined in more detail below. Still, briefly, at the macro level, analysing the dataset as a totality, we asked what the development of the metaphor's usage could tell us about constitutional metaphors in the news media more generally. Here, context is paramount. Discourse is only intelligible within the complexities of the social context in

which the language is formed and communicated. With this in mind, questions of motivation were the first to confront us in the dataset—why did media outlets consider a private decision-making body like the OB newsworthy at all? How did the metaphor differ in use between various news outlets? How are authors responding to and reproducing discourse? At the micro-level of individual articles, we asked how authors framed OB:SC. Did the author qualify its usage with quotation marks or other forms of distancing? Was the metaphor used in the article’s title? Did the author note Facebook’s status as a private company and its relationship to the OB? Questions such as these both emerged from and guided our iterative coding process.

5.3.1.1. Data selection and cleaning

We began by collecting a dataset of online news articles to capture the development and usage of OB:SC—meaning we intentionally only gathered news articles that met our criteria of invoking the constitutional-metaphorical terminology of the OB as a ‘supreme court.’ As such, we excluded earlier coverage that used the term ‘supreme court’ to refer to Facebook’s initial plans to introduce a form of independent oversight that preceded the OB as the entity established to perform this function. Such plans remained vague until Facebook published the OB’s draft charter in January 2019, and it was only thereafter that the nascent body’s institutional contours became apparent. Although the metaphorical term (‘supreme court’) remained the same, its functionality changed once it was linked to the OB as a specific institution. Originally, the term outlined an idea of what *ought* to be done, reflecting the vague objective outlined by Zuckerberg. But once the OB’s institutional structure became apparent from January 2019 onwards, the metaphor morphed into a descriptor of work underway in earnest. Only then could the term supreme court be ‘mapped’ (Lakoff, 1992:27) onto that of the OB as a specific named institution—instead of creating an imaginary standing on its own. Although both ways that the metaphor was used illustrate constitutional metaphors’ performative potential, only the latter offers a concrete illustration of the possibly legitimising effects of constitutional metaphors vis-à-vis the specific institution of the OB.

To assemble our dataset of media references to the OB as a supreme court, we used Media Cloud. This platform provides a database of over 1.7 billion news articles gathered since

2008 (Roberts et al., 2021:1). Researchers can query the database with search terms and Boolean operators; searches produced similar results to Google Trends in analogous research (Simon and Camargo, 2021:3). The Media Cloud Topic Mapper allows users to search for articles with search operators. After iterative testing, we decided on the following search terms: *'facebook' AND 'oversight board' AND 'supreme court' AND NOT ('privacy and civil liberties')* (specifically excluding 'privacy and civil liberties' to avoid irrelevant articles referencing the US Privacy and Civil Liberties Oversight Board).³⁶ It should be noted that the body's official name changed from 'Facebook Oversight Board' to 'Oversight Board' in its late planning process in early 2020. Although we use the latter term throughout the article, the Boolean operators we used for data collection incorporated both versions. To retrieve a representative sample of the US online news discourse on the topic, we queried Media Cloud's established "United States – National" collection. Our query resulted in a dataset of 522 articles spanning from the first public mentions of the OB in April 2018 to the coverage of the May 2021 decision on banning Donald Trump from Facebook. After reading each article to verify that all mentions were valid invocations of the constitutional language about the OB,³⁷ we were left with a dataset of 389 articles, whose size alone speaks to the prevalence of the metaphor's use in mainstream media discourse. The size of outlets spans from globally consumed news media like the New York Times to more tech-focused outlets like TechCrunch, and the collection also varies ideologically, encompassing right-wing outlets like Breitbart and Fox News and centrist and left-of-centre entities like Reuters and CNN. While the articles' authors are almost exclusively journalists, some articles contain direct quotations or interviews by OB members, academics, or politicians, which were also considered in the analysis.

To get a better impression of the representativeness of the usage of OB:SC with respect to the overall media attention that was paid to the OB during the same period, we additionally searched Media Cloud for *all* articles mentioning the OB, regardless of whether they

³⁶ Though we checked a sample of the excluded articles to verify that this decision did not wrongly exclude relevant articles, we acknowledge the possibility that some valid samples may have been omitted from our final dataset as a consequence of this choice.

³⁷ The individual search tokens, e.g. 'oversight board' or 'facebook', in our Boolean expression were matched without case sensitivity—and in our case all tokens only included alphabetic characters—to the text in the Media Cloud database (Roberts *et al.*, 2021). We used speech marks in our query in order to return references to a 'supreme court' specifically, which yielded both articles that used the phrase in speech marks, and articles which used it without speech marks. We then read each article to check whether the author used speech marks when invoking the OB:SC.

employed OB:SC,³⁸ and found similar peaks and troughs over time, with peaks coinciding with major news events associated with the OB.

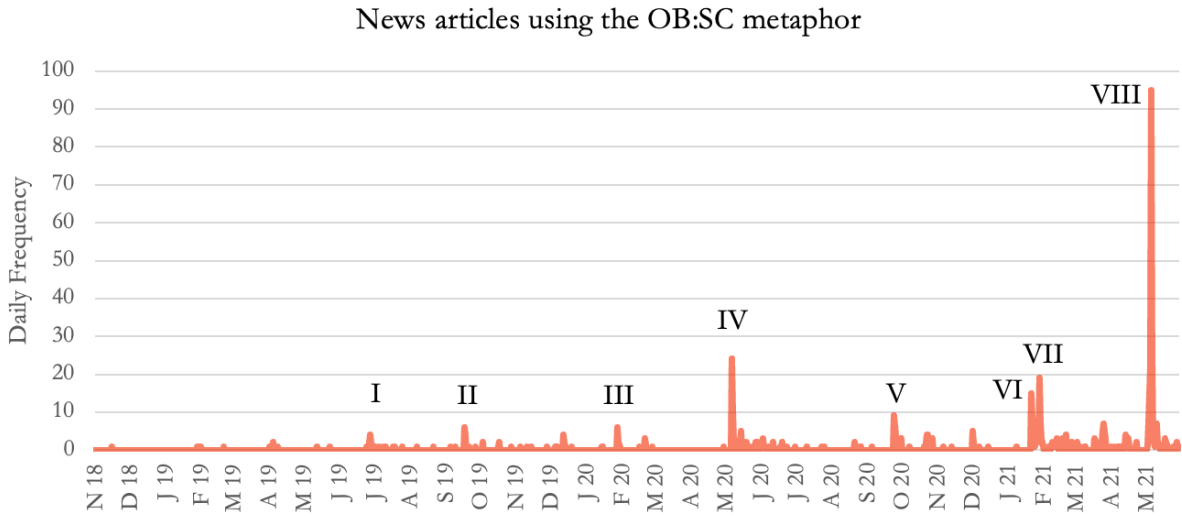


Figure 10: Number of US online news articles using the OB:SC metaphor per day, October 2018 to May 2021 (Media Cloud 'United States – National' Collection)

In Table 5 and Figure 10, we identify the main newsworthy events in the OB’s early existence. Figure 10 illustrates the prevalence of OB:SC references in media articles across the period of our study. The numerals overlaid in Figure 10 indicate major spikes in media coverage following key events in the OB’s chronological development, which we detail in Table 5.

5.3.1.2 Macro-level analysis: the overall spread of the OB:SC metaphor

We found that the majority of stories involved direct reporting on Facebook’s corporate communications about the OB (see Figure 10 and Table 5). From our coding, 218 of the 389 articles were responses to identifiable events, with the rest being broader opinion pieces, tangential mentions of the Board, or miscellany such as Noah Feldman’s appearance at Donald Trump’s first impeachment hearing.³⁹

³⁸ This search was the same as our previous search but did not require a ‘supreme court’ token match: ‘facebook’ AND ‘oversight board’ AND NOT (‘privacy and civil liberties’). This search resulted in 2046 articles.

³⁹ Breitbart, <https://www.breitbart.com/tech/2019/12/05/democrat-impeachment-witness-noah-feldman-behind-idea-for-facebook-oversight-board/>

Event	Date(s) of media coverage peak	Event and relevant corporate communication
I	June 27 2019	Release of global feedback and input on OB https://about.fb.com/news/2019/06/global-feedback-on-oversight-board/
II	September 19 2019	Announcement of OB structure https://about.fb.com/news/2019/09/oversight-board-structure/
III	January 28 2020	Announcement of OB bylaws https://about.fb.com/news/2020/01/facebooks-oversight-board/
IV	May 6-9 2020	Announcement of OB board members https://oversightboard.com/news/327923075055291-announcing-the-first-members-of-the-oversight-board/
V	October 22 2020	OB starts to accept cases https://oversightboard.com/news/833880990682078-the-oversight-board-is-now-accepting-cases/
VI	Jan 21 2021	FB announces intention to refer Trump suspension to the OB https://about.fb.com/news/2021/01/referring-trump-suspension-to-oversight-board/
VII	Jan 29 2021	OB announces its first decisions https://oversightboard.com/news/165523235084273-announcing-the-oversight-board-s-first-case-decisions/
VIII	May 3-8 2021	OB announces decision on Trump ban https://oversightboard.com/decision/FB-691QAMHJ/

Table 5: Chronology of main events in the media discourse surrounding use of OB:SC

While reports of announcements made by Facebook and its leadership during the initial creation of the Oversight Board was minimal, they established a meaningful 'precedent' to which later authors would often refer. The first notable event in the creation of what would become the OB was the initial description of the idea by Mark Zuckerberg on April 6, 2018: *'You can imagine some sort of structure, almost like a Supreme Court, that is made of independent folks who don't work for Facebook, who ultimately make the final judgment call*

*on what should be acceptable speech in a community that reflects the social norms and values of people all around the world.*⁴⁰

Zuckerberg's initial public articulation of OB:SC would become a meaningful referent for articles using the 'supreme court' metaphor. Feldman introduced the metaphor in January 2018 in a white paper produced for Facebook executives Mark Zuckerberg and Sheryl Sandberg.⁴¹ Later that year, Zuckerberg shared his vision of the future Oversight Board, though few news outlets adopted the OB:SC metaphor immediately. Then, during the first half of 2019, Facebook made two major announcements regarding the Board—one releasing the first draft of the Board's charter, and another announcing consultation feedback received on the Board's proposed remit and structure—in which the use of 'Supreme Court' was still present (Peak I in Figure 10).

It appears that when describing these initial milestones in the OB's early development, journalists reached for language already present, whether from previous articles or Zuckerberg's own words—or, as in the following example, both. Here, *Fox Business* cites Zuckerberg as the source of the metaphor and links (see underline) to an earlier article by *The Verge*, also in our dataset.⁴² *'Nevertheless, the Oversight Board created in September of the be used for exactly that: oversight. The company's CEO Mark Zuckerberg compared the group to a Supreme Court for Facebook.'*⁴³

As the OB continued to ossify into a more tangible entity, we saw a large overall increase in news coverage that invoked its characterisations as a 'Supreme Court,' specifically in May 2020 (Table 5:IV). The reason for this was the announcement of the Board's first 20 members, generating a slew of articles, 34 of which were captured in our dataset. Even an article from newswire Reuters—usually known of the sparseness of its characterisation of

⁴⁰ We include this interview despite it not mentioning the 'Oversight Board' because of its frequent reference by articles in our dataset. <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge>

⁴¹ See Facebook, *Global Feedback & Input on the Facebook Oversight Board for Content Decisions*, 27 June 2019, p 8 - <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-2.pdf> (last accessed 7 September 2021). For Feldman's memos see Fn 1.

⁴² <https://www.theverge.com/2019/9/17/20870827/facebook-supreme-court-mark-zuckerberg-content-moderation-charter>

⁴³ <https://www.foxbusiness.com/technology/the-facebook-oversight-board-first-case-announcements-draw-skepticism>

news events—deployed the metaphor, albeit in the third person: *'The independent board, which some have dubbed Facebook's "Supreme Court," will be able to overturn decisions by the company and Chief Executive Mark Zuckerberg on whether individual pieces of content should be allowed on Facebook and Instagram.'*⁴⁴

Notably, after its early announcements, Facebook representatives scarcely used the OB:SC metaphor, save for sporadic instances in media appearances.⁴⁵ The term does not appear in the OB's normative framework, the so-called 'Oversight Board Charter'.⁴⁶ The OB itself has not used it in any official publications. Nonetheless, many journalists, politicians, columnists, scholars, and others continue to use it frequently when speaking about the OB. Discursively this is significant. By first introducing the metaphor and then stepping back from it, Facebook stands to have benefitted from its perpetuation by journalists and others—in terms of its potentially legitimating effect—while no longer facing accountability for its continued use, let alone face questions about the (in)appropriateness of the metaphor itself. This may be the greatest effect of the metaphor: swaying the public thought process by tilting the discourse towards a path that may lead to almost unquestioned legitimacy of the OB's decision-making.

5.3.1.3. Micro-level analysis: how was the metaphor deployed?

Through fine-grained analysis of each article, several distinct categories in how the OB:SC metaphor was used emerge. In what follows, we identify the most common characteristics of OB:SC deployment.

First, we looked carefully at the use of punctuation when the OB as a 'supreme court' was introduced to readers. Some authors used quotation marks around 'supreme court' as a means of distancing themselves from the denoted meaning and/or attributing the metaphor to someone else, for example: *'The Facebook oversight board, commonly known as*

⁴⁴ <https://www.reuters.com/article/us-facebook-oversight-idUSKBN2212LQ>

⁴⁵ While our dataset supports the scarcity of OB:SC use by employees, it was brought to our attention by Thomas Kadri that other references were made, for example, by Zoe Darmé, then manager of Facebook's Global Affairs and Governance team, in a Lawfare Podcast interview, and by Dexter Hunter-Torricke, Head of Communications for the Oversight Board, in a recent article not published by an outlet in our dataset (PR News).

⁴⁶ https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf (last accessed 16 September 2021).

*the 'Facebook supreme court,' which was established by the social network as a quasi-independent body to hear appeals on Facebook's decision to censor or ban content, has now been empowered to demand more censorship from Mark Zuckerberg's platform.*⁴⁷

The use of passive voice here in 'known as' separates the author from the term, though the use of 'commonly' nonetheless strengthens it by suggesting it is widespread. Other authors noted (correctly) that the terminology was first used by Facebook. Overall, about 61% of all articles used quotation marks around 'supreme court', and 30% referred to Facebook executives as the source of the metaphor. While these two categories may appear to convey similar ideas and are by no means mutually exclusive, we make a distinction between them because we interpret them as having significantly different roles in the development of discourse. Quotations directly attributed to Facebook and its executives do perpetuate the metaphor, but *incorporations*—use of the metaphor without direct attributions to Zuckerberg, for example—adopt this language less critically, masking its historical lineage. From our analysis, the latter proved as twice as common.

Second, as the articles using the OB:SC began to increase, third-person attribution became increasingly frequent as well. Here is a notable example asserting that even some *outside* the company had used the metaphor: *'The board, which was first announced in 2018, is made of the 20 journalists, lawyers, activists, and even a former prime minister, and has been labeled by some outside the company as 'Facebook's Supreme Court.'*⁴⁸

However, beyond these different methods of positioning and attributing the metaphor, others eschewed qualifications or caveats altogether, embracing the OB as 'Facebook's Supreme Court', sometimes even without quotation marks.⁴⁹ We also recorded whether articles used the OB:SC metaphor in the title of the article, which would be especially consequential for those readers who only saw a headline and therefore didn't have this description contextualised by further details about the board's actual function and scope.

⁴⁷ <https://www.breitbart.com/tech/2021/04/13/facebook-supreme-court-empowered-to-demand-more-censorship/>

⁴⁸ <https://www.engadget.com/facebook-oversight-board-take-appeals-facebook-and-instagram-155533590.html>

⁴⁹ <https://www.forbes.com/sites/rebeccabellan/2020/05/06/facebook-announces-members-of-its-supreme-court/?sh=1cdf28253f12>

Of all the articles using the metaphor, about 20% employed OB:SC in the title, though it is worth noting that this distribution skews towards earlier articles, before the official launch of the board in October 2020, where almost 30% of the articles used OB:SC in the title.

We also considered more interpretive qualities regarding whether authors reflected critically about describing the OB as a 'supreme court.' Though 'criticality' is less analytically concrete—and for this reason evades a more quantitative presentation—than our other measures, we sought to operationalise it by asking whether the author made a note of the descriptive shortcomings of OB:SC, 1) because Facebook and the OB are private entities and/or 2) because of the limited overlap between the OB's scope and function and that of a supreme court.

Several articles criticised the Board's makeup. This was only after the OB's members were announced, and after the Board had officially launched. Therefore, they were less focused on the OB's (lack of) legitimacy as an institutional entity per se than on its (in)ability to execute specific functions. Most articles of this type noted how the board was unable to carry out its duties more objectively, many of which hailed from right-wing publications, such as Breitbart, with a history of criticising Silicon Valley for being too 'left-wing'. These were less relevant to our thesis concerning constitutional metaphors and legitimacy. However, the shifting nature of critiques of the OB after it was established, from big-picture criticisms of the Board itself, to more fine-grained engagement with specific appointments or decisions, serves to highlight the increasing 'incumbency advantage' of an institution once it is established, which is itself relevant to our discussion of legitimacy.

Taken together, our micro-level analysis indicates the widely uncritical use of the metaphor. Although many articles put the metaphor in quotation marks, only very few articles question its descriptive accuracy and potential political consequences. This was to change, however, with the events of January 2021.

5.3.1.4. External Shock: The Trump Case

The characterisation of the OB as a 'supreme court' only became scrutinised at significant scale after an external shock: the Capitol Hill insurrection on 6 January 2021, which caused

Facebook to ban President Donald Trump. Trump appealed the decision in February 2021, at which point the company delegated the decision on whether the measure was in accord with Facebook's normative framework, namely its Charter, of the OB. The OB upheld the ban in May 2021, while requiring Facebook to revisit the case at a later date. Our dataset allowed us to explore the media discussion concerning the case (peaks VI and VIII in Figure 10) and made clear that the emboldened critique of the OB as a supreme court was highly determined by the weighing-in of high-profile public figures. Many news articles employing the OB:SC metaphor—now more critically—in discussion of these news events relied on public voices such as those of senators Elizabeth Warren, Ted Cruz, and Bernie Sanders, to juxtapose the legitimacy of democratically elected officials with the legitimacy of the OB. For example, *NPR* draws on the following quote from Senator Warren: *'We need to break up these giant tech companies, and Facebook is one of them. They are crushing competition and in cases like Facebook, they're acting like they're bigger than government.'*⁵⁰

As this quote exemplifies, statements by public figures increasingly framed the relationship between large technological companies and public authority as an agonistic struggle. Many public officials held steadfast of the assertion that the OB was *not* a supreme court. Nonetheless, between event VII on 29 January 2021, when the Board issued its first decisions, and event VIII on 8 May 2021, when the board issued its decision in the Trump case, the use of the metaphor almost quintupled (see chart). Therefore, the Trump case triggered a double effect. On one hand, it prompted increased scrutiny by several public figures infusing criticism towards the metaphor. On the other hand, the Trump case only made the OB—and with it its metaphorical portrayal as a supreme court—more visible. It is an open question as to which of these two aspects affects the Board's public perception and legitimacy more strongly: criticism by politicians, or the continued prevalence of a possibly legitimising constitutional metaphor propelled by the media.

5.3.1.5. Discursive dominance?

In sum, our empirical analysis demonstrates that media discourse surrounding platform governance is extremely fluid, and suggests that, at least of the present case study, the news media existed almost as a stage on which executives, scholars, politicians, and the

⁵⁰ <https://www.npr.org/2021/05/07/994436847/what-we-learned-about-facebook-from-trump-decision>

like—in concert with journalists themselves—negotiated the concepts with which an understanding of the phenomenon was formed. Especially noteworthy here is the role of Facebook itself: although the metaphor originated from Noah Feldman and Mark Zuckerberg, the company then largely ceased using it in its public pronouncements. Neither the Oversight Board, to our knowledge, used the metaphor in its official publications. Instead, the metaphor seemingly developed ‘a life of its own’ through its use by journalists and others, earning a degree of discursive dominance in depictions of the OB, and perceptions of its legitimacy in turn. For example, Oversight Board member Julie Owono appeared in a podcast titled “Digital Rights and Facebook’s Supreme Court with Julie Owono” on 1 September 2021, however, without herself using the metaphor when describing the Board.⁵¹

This prompts a final set of evaluative questions as to the metaphor’s role in the discourse: the issues of intent, planning, and persuasion. Did Facebook do this on purpose? And if so, to achieve what goal? Related scholarship has noted how the executives of technology companies ‘utilise’ metaphorical devices as ‘vehicles of persuasion’ (Lindh and Nolin, 2017:1–6). Metaphorical imagery may also form part of broader efforts to ‘discursively construct’ platforms through the use of future imaginaries (Hoffmann, Proferes and Zimmer, 2018; Haupt, 2021). Lindh and Nolin, for example, argue that the overarching aim of large technology companies ‘seems to be disconnected from business models and instead concerned with the improvement of people’s lives, making the world a better place. However, this is, of course, a strategic choice of narrative.’ (Lindh and Nolin, 2017:19).

This rationale seems applicable to the supreme court metaphor as well. A multimillion-dollar initiative like the Board established by a multibillion-dollar company like Facebook is not framed in one way or the other by mere chance. And Facebook still utilises such language in the context of the OB, most strikingly in the Board’s bylaws⁵², alluding to ‘*We the People*’, the phrase founding the American republic. Even more bluntly, Feldman, the Board’s creator, argued in internal memos that “*the [aim] of the create a durable institution to deliver principled, reasoned decision-making that would be widely understood as legitimate*

⁵¹ The Priv8 Podcast with Derek E. Silva, 1 September 2021, see here:

<https://www.youtube.com/watch?v=fSEEv6IRBPO> (last accessed 16 February 2022).

⁵² Art. 3, The Oversight Board Bylaws, January 2021, <https://www.oversightboard.com/sr/governance/bylaws> (22 September 2021) (emphasis added).

*[... and ...] to capture the legitimacy benefits of decisional independence while maintaining the Facebook courts' association with Facebook.*⁵³

However, despite these indications of Facebook's motivations, we must dissect the company's possible internal decision-making process from the media discourse about the result of said decision-making process. Whereas the latter is well documented in our data set and intelligible via our content analysis, assertions about the former would remain mainly speculative as they are not covered by our data set. Rather, substantiating claims about the company's intent and possible planning would require a different methodological approach, such as qualitative interviews with company executives (Klonick 2020:2426–27). Whereas this constitutes a promising avenue for further research, it goes beyond this article's scope and focus. What our analysis does show, however, the news discourse played a major contributory role in adorning a novel institutional entity with the sociopolitical legitimacy of hundreds of years of political and legal thought and practice.

5.4. Discussion of constitutional metaphors and the need for critical reflections

In the previous section, we found that constitutional metaphors like OB:SC can be understood as, at least at their outset, an attempt by social media companies to legitimate their private platform governance power—and that the use of a metaphor can persist even after its creator ceases to employ them. In this section, we explore the implications of the OB:SC metaphor and explain our opposition to its use. First, why the term 'supreme court' is not an apt descriptor the OB. Second, we explain how a metaphorical allusion like OB:SC may legitimate the actors to whom it is applied, and why this is normatively unwarranted. Third, we investigate how constitutional metaphors may reflexively construct a new meaning of the constitutional concepts on which they originally rely.

5.4.1. What is a supreme court anyway?

At first sight, the answer of the question of why the OB does not qualify as a supreme court seems simple. A court is commonly understood as a branch of government that

⁵³ Feldman, Facebook Supreme Court: A Governance Solution, March 2018, p 104 and 110—see fn1.

independently and impartially applies the law when adjudicating disputes among citizens and between a citizen and the state. The OB is not a branch of government, but instead emanates from a privately run trust which is controlled by a private company. Its independence has been contested (Douek, 2019). It does not apply democratically enacted laws, but its scope is instead circumscribed by a 'charter' crafted by Facebook itself. Rather than adjudicating disputes among citizens and states, the OB decides whether Facebook's content moderation efforts vis-à-vis its users comply with norms set by the company. It should be noted that in most of its decisions the Board focused mainly on international human rights as the normative yardstick. This, however, is a choice by the Board—and not 'bindingly' reflected in its charter, the bylaws, or the trust agreement.

A supreme court, on the other hand, is typically the highest court within the legal system of a given jurisdiction, meaning its decisions cannot be appealed. In many systems, it is also the sole court which has jurisdiction over matters of constitutional law. Many courts are also empowered to perform the function of judicial review, assessing acts of legislation for their constitutionality, and, if necessary, declaring them unconstitutional. Today, most supreme courts or constitutional courts in democratic systems exercise judicial review. The OB does not possess the power to review the terms of service of Facebook as inconsistent with the charter (though it may recommend changes to these terms). That would reminisce judicial review in the constitutional law sense of the word.

The picture looks bleaker still when we turn to the 'charter', the OB's normative framework. It tells us almost nothing about the 'rights' the OB ought to apply. Its sparse references to 'human rights norms' and 'freedom of expression' remain vague and tentative.⁵⁴ The charter neither expressly establishes new rights, nor incorporates existing ones. It remains silent on structural issues like norm-setting or democratic participation. Equating the charter to a constitution is therefore misleading. Thus, the OB evidently does not meet the definition of a supreme court as commonly understood.

Stopping here, however, would risk over-simplification. It is true that the OB engages in norm-based third-party review of whether Facebook's actions as an institution infringe on individual(s) 'rights'. This is exactly what courts do. And, although there exists no universal

⁵⁴ Preamble and Art. 2 Sec. 2 *Oversight Board Charter*, September 2019, https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf (last visited 21 September 2021).

definition of what exactly a (supreme) court is, the OB's practices resemble many features typically encountered in courts. It receives written submissions, deliberates, and gives judgments and reasons for these judgments. It interprets norms. As Shapiro might describe it, it is—or at least may yet become—a 'social controller' (Shapiro, 1981:24). Further, in recent decades, the role of courts and judges has transformed from a pacifier of disputes between individuals into that of a 'manager' of complex, long-lasting issues converging law and policy (Chayes, 1976; Resnik, 1982). Such a role is perhaps the most important aspect of the Oversight Board's practice as the OB issues, in its decisions, 'policy advisory statements' that 'request' or 'recommend' structural, procedural, or normative changes to Facebook's content moderation regime. Further, at least in theory, the OB bases its decisions only on the normative framework it is set to apply. Also, just as we have already seen that private governance exists, so too do courts rule on this private governance beyond public authority. Think of, for example, the peculiar area of sports law and the Court of Arbitration for Sport in Lausanne, Switzerland. There are all kinds of international tribunals and courts. So why not a supreme court for Facebook?

The picture appears more nuanced the closer one looks. This is perhaps the reason the supreme court metaphor gained so much traction. It boils down various complex and interdependent questions into an easy-to-digest and publicly established narrative. However, we argue that this constitutional metaphor forecloses, rather than foregrounds, open debate about the adjustment to power relations that were, for decades, dominated by the notion of the state as the locus of authority. The uncritical use of constitutional metaphors is imprecise and even potentially dangerous, insofar as it may be taken to imply that the entities being 'constitutionalised', such as Facebook, have the status of legitimate governors.

5.4.2. The legitimating effect of the supreme court metaphor

If the Oversight Board is not a supreme court, then what is meant by claims that it is? Of course, one could simply say that such claims are false. But that would fall short of its conveyed meaning. The statement establishes a connection between a commonly known concept (a supreme court) and a different context (a private oversight body). It ties much of the former's socio-cultural significance and political legitimacy to the latter. As Lindh and

Nolin argue, technology companies might utilise metaphors as 'persuasive devices' to advance their agenda (Lindh and Nolin, 2017, p. 164). At this point, whether or not Facebook itself uses the metaphor is no longer decisive given its repeated use (sometimes critically) in reports, op-eds, podcasts, speeches, and interviews. Ultimately, constitutional metaphors legitimate certain choices regarding platform governance for their mere appearance instead of their substantive functionality, let alone normative desirability. As Klonick says, 'imagining the Oversight Board as a court is to invite debate' (Klonick, 2020:2476; Griffin, 2021). The Board obviously 'falls short of formal definitions as a court or a legal system' (Klonick, 2020:2476). Such language's effects, however, go far beyond 'formal definitions'. Nobody knows precisely what shape the OB will ultimately take or what its role will become, in the midst of epistemic progress wherein meaning-construction is still under way. The OB represents a step into uncharted territory. As such, to the extent that it continues to be portrayed as a supreme court, the OB draws on the legitimating power of century-old ideas of rights-based judicial control over administrative action. And rightly so, one might say, given the success of these ideas in the context of the state, constitutional democracy, and its rule of law. However, concepts carry their own baggage. In the case of the term 'supreme court', this baggage is brim-full of positive notions of democracy, individual rights, the rule of law, voice, and justice. The OB is not (yet) deserving of such acclaim-by-association.

5.4.3. The construction of meaning through constitutional metaphors

On top of their legitimating effect vis-à-vis certain institutions in the context of platform governance, the use of constitutional metaphors in the platform governance discourse may, in turn, reflexively modify the meaning of constitutional concepts in general. This may sound hyperbolic—but consider that meaning itself is fluid, perpetually modified and advanced by among other things science and culture. As such, meaning never unravels an ideal or universal truth. Instead, it is perpetually reconstructed by references to socio-cultural contexts which again are saturated with personal histories and peculiarities; in Lock and Strong's phrasing, 'meaning is not immanent in how things are, it is immanent in how we talk about them' (2010:148). Modifying language is therefore the first step in modifying meaning. The relationship between language and context resembles a reflexive symbiosis. Societies and cultural practices require communication and thus language, whereas

language requires a socio-cultural framework to function as a system of communication. Therefore, altering concepts through the repeated uncritical use of metaphors affects the socio-cultural context in which said concepts operate.

Consequently, new metaphors alter our individual as well as our socio-cultural conceptual systems. Such a change inevitably affects the actions people take and perceptions people hold, since both are to a large extent based on our conceptual system. Especially in the context of scholarly uncertainty on the OB's nature, the metaphorical use of the term supreme court evokes a performative spin. Once the very potent image of 'supreme court' is mapped onto the rather befogged image of the OB, the constitutional metaphor's entailments outdazzle the remaining questions regarding the metaphor's adequacy. Iteratively, speech once perceived as metaphorical may become performative.

Simply put: perpetual, uncritical utterances referring to the OB as a 'supreme court' may ultimately make it so. This raises serious questions regarding legitimisation and perceptions of the OB, and by extension Facebook's, legitimacy. Given recent regulatory efforts mandating online platforms to establish quasi-judicial bodies to deal with user complaints, we may be in the midst of a fundamental reconceptualisation of the rule of law in the digital sphere. The European Union, for example, recently proposed to subject large online platforms to so-called private 'out-of-court dispute settlement' bodies.⁵⁵ If passed, this would effectively establish privately-run, amorphous, quasi-judicial, administrative court-inspired institutions 'adjudicating' vast swaths of the disputes arising between platforms and their users (Holznagel, 2021) while the centrality of platforms to regulate communication only grows. These developments underline that the struggles of who shall govern our online lives, and by what means, have only just begun.

5.5. Conclusion

In this article we have documented characterisations of Facebook's Oversight Board as a 'supreme court', and situated these characterisations within both the socio-political context of private platform governance as well as within theoretical-legal debates over the status and role of courts. These strands dovetail, we have argued, in the concept of political

⁵⁵ See Article 18 of the Commission's 2020 proposal for a Digital Services Act, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0825&from=en> (last accessed 7 September 2021).

legitimacy. The Janus-faced understanding of legitimacy as simultaneously firm and normative, and fluid and descriptive, provides an apt theoretical frame for the empirical focus of our study. Viewed in this way, the OB seems to have obtained what can perhaps best be described as 'quantum legitimacy': it both is and is not legitimate, a state informed by the manner in which it is observed. Because constitutional metaphors—in this case, characterising the OB as a 'supreme court'—provide a particular lens through which such an observation occurs, we argue that they are themselves a politically important object of study.

Several considerations that we were unable to sufficiently cover here would benefit from further research. This includes the US-centricity inherent to both platform governance (Bloch-Wehba, 2019; Arun, 2021) and to the referents of constitutional metaphors—in this case a 'supreme court', but also proposed 'bills of rights' for internet users—which seem to further reinforce the centrality of American notions of constitutionality to how the internet and social media are governed.

Constitutional metaphors such as those invoking a 'supreme court' carry socio-cultural and political connotations, and their discursive ascent maps these connotations onto the discussed object. To the extent that this eventually increases the object's legitimacy and power (at least as viewed from certain vantage points), and vice versa, this might alter not only what is meant by certain terminology such as 'supreme court'. It is too soon to conclude whether 'Facebook's Supreme Court' will remain the dominant terminology by which the Oversight Board is referred to. However, we argue that constitutional metaphors such as this do not merely describe, but also shape broader societal shifts in the balance of power between entities like states and large technology companies. Therefore, constitutional metaphors may have unintended and undesirable normative consequences. Only further reflections on and critique of the discourse can advance us towards more apt terminology, and with it, epistemic progress.

References

- Arun C (2021) Facebook's Faces. ID 3805210, SSRN Scholarly Paper, 15 March. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3805210.
- Beetham D (2013) *The Legitimation of Power*. Macmillan International Higher Education.
- Berger LL (2009) How Embedded Knowledge Structures Affect Judicial Decision Making: An Analysis of Metaphor, Narrative, and Imagination in Child Custody Disputes. *Southern California Interdisciplinary Law Journal* 259.
- Berger LL (2013) Metaphor and Analogy: The Sun and Moon of Legal Persuasion. *Journal of Law and Policy* 22(1).
- Bloch-Wehba H (2019) Global Platform Governance: Private Power in the Shadow of the State. *SMU Law Review* 72(1): 27.
- Brooks P and Gewirtz P (eds) (1996) *Law's Stories: Narrative and Rhetoric in the Law*. New Haven; London: Yale University Press.
- Buchanan A (2002) Political Legitimacy and Democracy. *Ethics* 112(4). The University of Chicago Press: 689–719. DOI: 10.1086/340313.
- Buchanan A and Keohane RO (2006) The legitimacy of global governance institutions. *Ethics & international affairs* 20(4). Wiley Online Library: 405–437.
- Cappelletti M (1989) *The Judicial Process in Comparative Perspective* (eds PJ Kollmer and JM Olson). Oxford; New York: Clarendon Press.
- Celeste E (2019) Digital constitutionalism: a new systematic theorisation. *International Review of Law, Computers & Technology* 33(1): 76–99. DOI: 10.1080/13600869.2019.1562604.
- Chayes A (1976) The Role of the Judge in Public Law Litigation. *Harvard Law Review* 89(7): 1281. DOI: 10.2307/1340256.
- Cloutier É (2019) A Tale of Two Metaphors: A Narrative Take on the Canadian Constitution. *McGill Law Journal* 64(3): 447. DOI: 10.7202/1071747ar.
- De Gregorio G (2021) The rise of digital constitutionalism in the European Union. *International Journal of Constitutional Law* 19(1): 41–70. DOI: 10.1093/icon/moab001.
- Douek E (2019) Facebook's 'Oversight Board:' Move Fast with Stable Infrastructure and

Humility. North Carolina Journal of Law & Technology Volume 21(Issue 1): 1.

Foer F (2019) Facebook's War on Free Will. The Guardian, 19 September. Available at: <https://www.theguardian.com/technology/2017/sep/19/facebooks-war-on-free-will>.

Frenkel S and Kang C (2021) An Ugly Truth: Inside Facebook's Battle for Domination. First edition. New York, NY: Harper, an imprint of HarperCollinsPublishers.

Gallup (2018) Americans' Trust in the Three Branches of the U.S. Federal Government (Trends). Timberline 937008. Available at: <https://news.gallup.com/poll/243296/americans-trust-three-branches-federal-government-trends.aspx>.

Ganesh MI (2020) The ironies of autonomy. Humanities and Social Sciences Communications 7(1): 1–10. DOI: 10.1057/s41599-020-00646-0.

Gillespie T (2010) The politics of 'platforms'. New Media & Society 12(3). SAGE Publications: 347–364. DOI: 10.1177/1461444809342738.

Gillespie T (2018) Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press.

Gorwa R (2019) What is platform governance? Information, Communication & Society 22(6). Routledge: 854–871. DOI: 10.1080/1369118X.2019.1573914.

Gradoni L (2021) Constitutional Review via Facebook's Oversight Board. Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung. DOI: 10.17176/20210210-235949-0.

Graz J-C and Nölke A (2007) Transnational Private Governance and Its Limits. Routledge.

Griffin R (2021) Metaphors matter: Why we shouldn't call the Facebook Oversight Board a court. In: Hertie School Blog. Available at: <https://www.hertie-school.org/en/digital-governance/research/blog/detail/content/metaphors-matter-why-we-shouldnt-call-the-facebook-oversight-board-a-court>.

Haggart B and Keller CI (2021) Democratic legitimacy in global platform governance. Telecommunications Policy 45(6): 102152. DOI: 10.1016/j.telpol.2021.102152.

Hassoun N (2012) Globalization and Global Justice: Shrinking Distance, Expanding Obligations. Cambridge University Press.

Haugen F (2021) Statement of Frances Haugen. Washington D.C. Available at: <https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49>.

Haupt J (2021) Facebook futures: Mark Zuckerberg's discursive construction of a better world. *New Media & Society* 23(2). SAGE Publications: 237–257. DOI: 10.1177/1461444820929315.

Hoffmann AL, Proferes N and Zimmer M (2018) "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society* 20(1). SAGE Publications: 199–218. DOI: 10.1177/1461444816660784.

Holznagel D (2021) The Digital Services Act wants you to "sue" Facebook over content decisions in private de facto courts. *Verfassungsblog: On Matters Constitutional. Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung*. DOI: 10.17176/20210624-193507-0.

Hsieh H-F and Shannon SE (2005) Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15(9): 1277–1288. DOI: 10.1177/1049732305276687.

Jasanoff S and Kim S-H (2009) Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva* 47(2). Springer: 119.

Kadri T (2018) How Supreme a Court? *Slate Future Tense*. Available at: <https://slate.com/technology/2018/11/facebook-zuckerberg-independent-speech-content-appeals-court.html>.

Kaye D (2018) Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. A/HRC/38/35, 6 April. United Nations Human Rights Council.

Klonick K (2018) The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review* Vol 131: 1598–1670.

Klonick K (2020) The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal* Volume 129(Number 8): 2232–2605.

Lakoff G (1992) The Contemporary Theory of Metaphor. In: *Metaphor and Thought* (Ed. Andrew Ortony). 2nd ed. Cambridge: Cambridge University Press.

Lakoff G and Johnson M (2003) *Metaphors We Live By*. Chicago: University of Chicago Press.

Leetaru K (2019) Facebook Was A Democracy 2009-2012 But We Didn't Vote So It Turned Into A Dictatorship. Available at: <https://www.forbes.com/sites/kalevleetaru/2019/04/13/facebook-was-a-democracy-2009-2012-but-we-didnt-vote-so-it-turned-into-a-dictatorship/> (accessed 23 July 2021).

Lindh M and Nolin JM (2017) GAFA speaks: metaphors in the promotion of cloud technology. *Journal of Documentation* 73(1): 160–180. DOI: 10.1108/JD-03-2016-0039.

Lock A and Strong T (2010) *Social Constructionism: Sources and Stirrings in Theory and Practice*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511815454.

Mager A and Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society* 23(2). SAGE Publications: 223–236. DOI: 10.1177/1461444820929321.

Mayer F and Gereffi G (2010) Regulation and economic globalization: Prospects and limits of private governance. *Business and Politics* 12(3). Cambridge University Press: 1–25.

Mayring P (2014) Qualitative content analysis: theoretical foundation, basic procedures and software solution. Available at: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>.

McBain HL (1927) *The Living Constitution: A Consideration of the Realities and Legends of Our Fundamental Law*. The Workers Education Bureau Press.

Montesquieu C de S (1748) *The Spirit of Laws*. Batoche Books Kitchener (2001).

Nagel T (2005) The problem of global justice. *Philosophy & public affairs* 33(2). JSTOR: 113–147.

Peter F (2017) Political Legitimacy. In: Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2017. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/sum2017/entries/legitimacy/> (accessed 11 May 2021).

Pettit P (2012) Legitimacy and Justice in Republican Perspective. *Current Legal Problems* 65(1): 59–82. DOI: 10.1093/clp/cus016.

Pohle J and Thiel T (2020) Digital sovereignty. *Internet Policy Review* 9(4). Available at: <https://policyreview.info/concepts/digital-sovereignty> (accessed 2 March 2021).

Pozen D (2018) Authoritarian Constitutionalism in Facebookland: Introduction to Kate Klonick's 'Facebook v. Sullivan'. In: Knight First Amendment Institute at the Columbia University. Available at: <https://knightcolumbia.org/content/authoritarian->

constitutionalism-facebookland.

Resnik J (1982) Managerial Judges. *Harvard Law Review* 96(2): 374. DOI: 10.2307/1340797.

Rider K and Murakami Wood D (2019) Condemned to connection? Network communitarianism in Mark Zuckerberg's "Facebook Manifesto". *New Media & Society* 21(3). SAGE Publications: 639–654. DOI: 10.1177/1461444818804772.

Roberts H, Bhargava R, Valiukas L, et al. (2021) Media Cloud: Massive Open-Source Collection of Global News on the Open Web. arXiv:2104.03702 [cs]. Available at: <http://arxiv.org/abs/2104.03702> (accessed 9 January 2022).

Schmidt VA (2013) Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'. *Political Studies* 61(1). SAGE Publications Sage UK: London, England: 2–22.

Shapiro MM (1981) *Courts, a Comparative and Political Analysis*. Chicago: University of Chicago Press.

Sharon T (2020) Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*. DOI: 10.1007/s10676-020-09547-x.

Simon FM and Camargo CQ (2021) Autopsy of a metaphor: The origins, use and blind spots of the 'infodemic'. *New Media & Society*. SAGE Publications: 14614448211031908. DOI: 10.1177/14614448211031908.

Slupska J (2021) War, Health and Ecosystem: Generative Metaphors in Cybersecurity Governance. *Philosophy & Technology* 34(3): 463–482. DOI: 10.1007/s13347-020-00397-5.

Suchman MC (1995) Managing legitimacy: Strategic and institutional approaches. *Academy of management review* 20(3). Academy of Management Briarcliff Manor, NY 10510: 571–610.

Suzor N (2018) Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. *Social Media + Society* 4(3): 205630511878781. DOI: 10.1177/2056305118787812.

Suzor N, Van Geelen T and Myers West S (2018) Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. *International Communication Gazette* 80(4). SAGE Publications Ltd: 385–400. DOI: 10.1177/1748048518757142.

Tallberg J, Bäckstrand K and Scholte JA (2018) *Legitimacy in Global Governance: Sources, Processes, and Consequences*. Oxford University Press.

Taylor L (2021) *Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector*. *Philosophy & Technology*. DOI: 10.1007/s13347-020-00441-4.

Tribe LH (1987) The Idea of the Constitution: A Metaphor-morphosis. *Journal of Legal Education* 37(2). Association of American Law Schools: 170–173.

van Dijck J, Poell T and De Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford, New York: Oxford University Press.

Van Leeuwen T (2007) Legitimation in discourse and communication. *Discourse & Communication* 1(1): 91–112. DOI: 10.1177/1750481307071986.

Weber M (1978) *Economy and Society: An Outline of Interpretive Sociology*. Univ of California Press.

Weber M (1994) *Weber: Political Writings*. Cambridge University Press.

York J (2021) *Silicon Values: The Future of Free Speech under Surveillance Capitalism*. London/New York: Verso.

Chapter – 6 Discussion and Outlook

This dissertation investigates potentially fragmenting phenomena in relation to social media platforms. Platforms change their algorithmic systems, policymakers try to keep up in regulating business models and their social and legal implications, and citizens and organisations using the platform services adapt their behaviour to new functions and mechanisms. With all involved actors' decisions and behavioural adaptations, social media platforms as communication environments are constantly changing. Correspondingly change the environments in which citizens communicate, inform themselves, debate, work and love and adapt their behaviour according to the changes in platform affordances and opportunity structures (Theocharis et al., 2022).

This chapter discusses the findings of the presented studies and relates them to the research questions stated in Chapter 1. Additionally, this chapter evaluates the findings' implications with respect to recent policy debates about social platform governance and regulation. The chapter determines by providing an outlook for further research and conclusions.

6.1 Discussion of findings and implications

Concerning the overarching research question RQ1 (*"To what extent are social media platforms accelerators of socio-political fragmentation?"*), the findings imply that social media platforms contribute to democratic fragmentation in all three investigated political arenas. The empirical studies draw on this by showing potentially fragmenting results of social media platforms with respect to political campaigns, protest movements, and democratic institutions. The findings resonate with the notion of fragmentation and concerns of more individualised issue campaigns and investigate how social media platforms contribute to the process. The studies' findings illustrated how social media platforms may function as accelerators of fragmentation. Chapter 3 illustrated how social media platforms as digital campaign spheres might allow online partisans and extreme actors to influence mainstream parties and politics. The study implied partisans' potential influence on political election campaigns online in the case of the 2017 and 2021 German federal elections. The study in Chapter 4 identified social media platforms as potential stages for health misinformation, conspiracy narratives and so-called disinformed social

movements. The identified narratives resulted in protest movements of vaccination sceptics that posed risks to public health and may result in long-lasting groups or movements pertaining to fragmenting conspiracy narratives. Believing these narratives can result in a decrease in trust in government, society and democratic institutions. The third paper in Chapter 5 argued that using constitutional metaphors may have negative implications for the legitimacy of democratic institutions. The study implies that these metaphors may contribute to democratic fragmentation, for instance, by reducing citizens' trust in democratic institutions and increasing openness to belief in conspiracy narratives. All three studies' findings imply that social media platforms act as accelerators of democratic fragmentation by providing a potential stage for extreme actors and conspiracy narratives, actively using constitutional metaphors, and copying democratic institutions like the investigated supreme court metaphor for the Meta oversight board.

With respect to RQ2 (*"In the arena of political campaigns and elections, to what extent do social media platforms fragment political election campaigns?"*), Chapter 3 illustrated the potential risks of social media platforms in the arena of political election campaigns. Social media platforms provide digital political campaign spheres can enable extreme actors to influence mainstream parties and politics. For instance, partisan online behaviour influences online campaigns and candidates' visibility. Moreover, extreme actors like the far-right party AfD use strategies such as hashjacking to polarise online discourses and increase the visibility of their divisive messages (Darius and Stephany, 2019). The study finds that the observed sample of the German political sphere on Twitter has significantly changed from the election in 2017 compared to 2021. Whilst the far-right party AfD and its partisans are located in an isolated cluster in 2017, they are part of a larger right-wing cluster in the final week of the 2021 Federal election campaign. This is most likely the result of right-wing partisans also retweeting individuals from other parties and, in particular, politicians from the right-wing within the CDU like Friedrich Merz (who later became party leader) and Hans-Georg Maaßen (against whom there is a current party exclusion procedure by the CDU for repeated antisemitic posts on social media). Correspondingly, Twitter networks may reveal ideological shifts of parties and individual politicians, but the evidence remains indicative. However, the study reveals online political communities that shape due to retweeting behaviour by individuals and may reflect individuals' ideological positions on topics or candidates and parties. The study also underlines what methods platform

companies and advertisers are using to infer attitudes and traits of individuals and, thus, also emphasises the importance of regulation of targeting practices beyond user-focused transparency (Dobber et al., 2023). Concerning the fragmentation hypothesis, the study signifies how separate groups of partisans shape online political discourses based on ideological behaviour and selective reinforcement like retweeting and following even on relatively open social platforms such as Twitter.

Regarding RQ3 ("*With regards to protest movements, what factors can result in fragmenting protest movements on social media platforms?*"), the study presented in Chapter 4 focused on the so-called dark side of civic participation and protest movements by examining the presence of conspiracy narratives during lockdown protests in the UK. Groups of users and highly active network hubs shared hundreds of messages with conspiracy narratives related to vaccinations and the telecommunications technology 5G during the Covid-19 pandemic on UK Twitter. In the study, the main factors for fragmenting protest movements were anti-science narratives and the dissemination of harmful content concerning public health during the Covid-19 pandemic. My co-author and I made sense of an increasing overlap of the groups as disinformed social movements since protests were also carried out on the streets and fulfilled several criteria for social movements. These criteria offer forms of identification, opposition and totality for individuals forming communities via their online behaviour on Twitter, Telegram groups, and street protests. Despite content moderation efforts by Twitter during the pandemic, we found a high presence of harmful content, of which we gave some examples in the study. The study's findings imply that mis- and disinformation on social media platforms can contribute to the fragmentation of society. Especially with regards to conspiracy narratives, they may contribute to violent protests or the organisation of citizens into groups on Telegram or Signal that function very much like echo chambers where the administrators of the groups are in control of information and access to the group. Whilst social media platforms have increased their efforts to delete harmful content like health misinformation, messenger services remain largely unregulated. Nevertheless, provide a stage for divisive conspiracy narratives which contributes to democratic fragmentation.

Chapter 5 presented a study exploring the relationship between social platforms and democratic institutions in investigating RQ4 ("*Considering democratic institutions, how might*

social platforms contribute to a fragmentation of democratic institutions?"). The study illustrates that social platforms' governance initiatives, like Meta's oversight board, may negatively affect democratic institutions. Online news and academic debates can contribute to democratic fragmentation and, when uncritically using constitutional metaphors, erode the legitimacy of democratic institutions. In the case of the supreme court metaphor used to describe the Meta oversight board, the presented study showed the persistence of the use in online news articles, even long after the companies officials refrained from using the metaphor. The investigation of the use of constitutional metaphors in the US news discourse on Meta's oversight board demonstrates that researchers, policymakers and the broad public must understand and be aware of discourse dynamics as they unfold since the use of terms and narratives may significantly influence the perception of problems and imaginaries of the future (Mager and Katzenbach, 2021). These imaginations include debating and designing policy approaches to regulating social media platforms. If governments seek to approach risks of democratic fragmentation, they need to be more active in defining guidelines for content moderation decisions, risk reporting and mechanisms for appeals via legislation. Otherwise, companies will continue to build new organisations similar to democratic institutions, which may further decrease trust in democratic institutions and increase the fragmentation of established democracies.

6.2 Discussion of the limitations and ethical considerations

The presented studies face limitations that the dissertation's chapters discuss in more detail. Regarding the broader research question, the studies focused on Twitter and the governance of Facebook or its mother company Meta as some of the largest social media platforms that have been increasingly moderated in recent years. Messengers like Telegram and Signal, however, may pose direct risks in terms of radicalisation and spread of disinformation because they allow closed groups that function as echo chambers and are often unmoderated by the companies providing the services (Garimella and Eckles, 2017; Gursky and Woolley, 2021). As a theoretical limitation, it is difficult to disentangle political fragmentation caused by social platforms from increasing fragmentation as a result of technology-influenced surveillance capitalism (Zuboff, 2015) or socio-cultural phenomena related to social platforms that increase the individual desirability of singularity or individualisation (Reckwitz, 2020). Thus, the author focuses on three specific arenas and

carefully interprets the results of the studies concerning the fragmentation of democracy on a macro-level.

Additionally, social media platforms have a democratising effect that has been discussed with respect to civic participation. They may give people a voice and especially in countries without or with low media freedom, the possibility to inform themselves and organise protests. Not without reason, Twitter, Facebook and other social media platforms are prohibited in many authoritarian regimes and can only be accessed when using a virtual private network (VPN). Thus, the dissertation focuses on established democracies where freedom of the press and the right to demonstrate are protected. However, these crucial rights are also under attack in established democracies like the United States, the United Kingdom, and Germany. They must remain protected by governments, civil society organisations, and social media platforms. The democratic implications need to be considered in the design of regulatory initiatives, which the following section discusses after briefly considering ethical issues.

Regarding ethical considerations, the investigation of large-scale online discourse also comes with the responsibility to protect the individual privacy of accounts. Whilst all tweets on Twitter are public, many users might need to be made aware that they are the subject of research (Zimmer, 2010; Tinati et al., 2014). Thus, the studies only show user names of public figures and professional accounts, anonymised messages and content examples, and considered ethical aspects during the research process. The following section briefly discusses the relevance of the findings for current governance and regulatory initiatives concerning social media platforms.

6.3. Discussion of regulatory initiatives and Internet governance

For democratic governments, directly regulating social media platforms poses an elemental policy puzzle. In the view of many, implementing content control or a strong state-based regulator risks the separation of powers and freedom of press principles. However, the inactivity in regulating or actively governing the social media ecosystems has resulted in election interferences by internal and external actors, fuelling conflicts and potential negative impacts on billions of global adopters' psychological and physical health. The reference to direct consequences is relatively vague because the platforms are often a

factor or the stage for processes that may otherwise also happen but would likely be less accelerated and on a smaller scale. Additionally, research and knowledge production has been slowed down by limited data access and changing access regimes, e.g., API access for researchers (Tromble, 2021).

To address these two issues, 1) the policy puzzle of social platform regulation and 2) the lack of scientific knowledge on the social implications of social platforms, current regulatory approaches to social media governance take a multi-stakeholder approach, including civil society actors such as NGOs and researchers. Many policymakers understand the regulation of content as establishing a "ministry of truth"⁵⁶ and often need more knowledge, information, and data access to understand algorithmic content moderation systems, not speaking of how to regulate them efficiently. Consequently, some argue that accountability should not stop with regulating procedures but be applied to the overall business operations of social platforms, particularly their advertisement business, based on surveillance-like insights into individual preferences (Zuboff, 2015; Griffin, 2022). These calls for regulation of advertisement-based business models resonate with a report⁵⁷ that tested the moderation of political advertisements on major social media platforms. In most places, Meta and TikTok failed to stop paid advertisements that contained electoral disinformation. YouTube performed better in deleting or not allowing disinforming ads in the US elections. However, in other countries, such as during the Brazilian elections, Meta, TikTok, and YouTube failed to detect electoral disinformation in paid advertisements. This study underlines fundamental accountability problems in the advertisement business that may contribute to and even accelerate political conflicts and democratic fragmentation.

Regarding the history of Internet governance, Hoffmann (2005) describes several phases moving from a technical regime to a phase of self-governance and then to a phase of governance based on multi-stakeholder processes. The recent debate on social media governance reflects this broader question of the state's role in regulating media platforms (Stockmann, 2022). While finding effective legislation is crucial, democratic states must also

⁵⁶ Reference to the Ministry of Truth in George Orwell's novel "1984"

⁵⁷ Global Witness Briefing October 2022, "TikTok And Facebook Fail to Detect Election Disinformation in the US, While YouTube Succeeds", https://cybersecurityfordemocracy.cdn.prismic.io/cybersecurityfordemocracy/39e0ef2e-2818-4210-92fc-61922140e8f9_Election+disinformation+on+social+media+in+the+midterms+-+Global+Witness_C4D_Oct22.pdf, accessed December 6, 2023.

be careful not to take a road to illiberal democracy, especially regarding information control, content removal and data access by state actors (Flonk et al., 2020; Flonk, 2021). Thus, EU legislation like the Digital Services Act (DSA) takes a multi-stakeholder and procedural approach. The DSA establishes oversight over social platforms, harmonises risk and transparency reporting responsibilities and requires clear procedures in content moderation. For the success of the DSA, it is important how national Digital Service Coordinators and potential auditing structures will be designed and equipped as described in detail by two recent policy briefs (Jaursch, 2022; Darius et al., 2023). Moreover, an intermediary body will participate in the research data access regime as the European Digital Media Observatory recommended.⁵⁸ Whilst the DSA will improve the understanding of social media platform risks and social implications, the legislation only regulates targeted advertisement for minors. However, a better empirical understanding of the risks of recommender systems and targeted advertisements can inform future regulation of social media platforms' advertisement-based business models.

Further research in the social and political sciences should provide an empirical basis for policy decisions and continue investigating online discourse and potential efforts by foreign state actors and coordinated groups to spread disinformation and interference in democratic elections. It is essential to be aware that existing privately-governed social media platforms have not been designed to foster democratic discourse but to maximise user engagement. Thus, the fragmentation of democracies may continue without more explicit rules and regulations of business models and recommender systems informed by non-governmental organisations and academic research.

6.4. Outlook and Conclusions

This dissertation illustrated that social platforms contribute to democratic fragmentation in three arenas, 1) political campaigns, 2) protest formation and 3) democratic institutions. The wide use of social media in politics and the increased power of platforms as information

⁵⁸ European Digital Media Observatory Working Group on Platform-to-Researcher Data, <https://edmoprod.wpengine.com/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf>, accessed December 12, 2022.

gatekeepers makes social platform architecture and governance crucial factors for the quality and legitimacy of democracy. Elon Musk recently acquired Twitter, laying off hundreds of employees. This rapid restructuring results in doubts about the further operability of the company, in particular within the trust and safety team responsible for content moderation and user harm prevention. This recent development underlines the need for independent academic work on developments in the sector, as well as active monitoring and oversight by researchers and state agencies. Moreover, technological innovations like the Metaverse and large-language models like Chat-GPT and other applications of generative artificial intelligence pose new challenges to the governance of social media platforms and online content. Only if researchers adapt their research methods to new technologies and receive extended research data access can we better understand the social implications of social media platforms and other communication technologies to protect democratic societies from further fragmentation.

Bibliography

Baranowski P, et al. (2022) Patterns of Negative Campaigning during the 2019 European Election: Political Parties' Facebook Posts and Users' Sharing Behaviour across Twelve Countries, *Journal of Information Technology & Politics*, 0(0), pp. 1–18. Available at: <https://doi.org/10.1080/19331681.2022.2115598>.

Bennett WL, Iyengar S (2008) A New Era of Minimal Effects? The Changing Foundations of Political Communication, *Journal of Communication*, 58(4), pp. 707–731. Available at: <https://doi.org/10.1111/j.1460-2466.2008.00410.x>.

Bennett WL, Segerberg, A (2013) *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge: Cambridge University Press. Available at: <https://doi.org/10.1017/CBO9781139198752>.

Bimber B (1998) The Internet and Political Transformation: Populism, Community, and Accelerated Pluralism, *Polity*, 31(1), pp. 133–160. Available at: <https://doi.org/10.2307/3235370>.

Bode L, et al. (2015) Candidate Networks, Citizen Clusters, and Political Expression: Strategic Hashtag Use in the 2010 Midterms, *The ANNALS of the American Academy of Political and Social Science*, 659(1), pp. 149–165. Available at: <https://doi.org/10.1177/0002716214563923>.

Bruns A, Burgess JE (2011) The use of Twitter hashtags in the formation of ad hoc publics, in *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*.

Chadwick A (2017) *The hybrid media system: Politics and power*. Oxford University Press.

Chadwick A, Dennis J, Smith AP (2015) Politics in the age of hybrid media: Power, systems, and media logics, in *The Routledge companion to social media and politics*. Routledge, pp. 7–22.

Colleoni E, Rozza A, Arvidsson A (2014) Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter, *Journal of Communication*, 64(2), pp. 317–332. Available at: <https://doi.org/10.1111/jcom.12084>.

Conover M, et al. (2011) Predicting the Political Alignment of Twitter Users, in, pp. 192–199. Available at: <https://doi.org/10.1109/PASSAT/SocialCom.2011.34>.

Conover MD, et al. (2012) Partisan asymmetries in online political activity, *EPJ Data Science*, 1(1), pp. 1–19. Available at: <https://doi.org/10.1140/epjds6>.

Copeland L, Römmele A (2014) Beyond the base? Political parties, citizen activists, and digital media use in the 2009 German federal election campaign, *Journal of Information Technology & Politics*, 11(2), pp. 169–185.

Couldry N, Hepp A (2017) *The mediated construction of reality*. John Wiley & Sons.

Cowls J, et al. (2020) Freedom of Expression in the Digital Public Sphere – Strategies for bridging information and accountability gaps in algorithmic content moderation. Available at: <https://doi.org/10.5281/zenodo.4292408>.

Dahlgren P (2005) *The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation*. *Political Communication* 22,2(2): 147–62. <https://doi.org/10.1080/10584600590933160>.

Darius P, Stephany F (2019) “Hashjacking” the Debate: Polarisation Strategies of Germany’s Political Far-Right on Twitter, in I. Weber et al. (eds) *Social Informatics*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 298–308. Available at: https://doi.org/10.1007/978-3-030-34971-4_21.

Darius P, Stockmann D, Bryson J, Cingolani L, Griffin R, Hammerschmid G, Kupi M, et al. (2023), *Implementing Data Access of the Digital Services Act: Collaboration of European Digital Service Coordinators and Researchers in Building Strong Oversight over Social Media Platforms*. <https://doi.org/10.48462/opus4-4947>.

DiMaggio, P. et al. (2001) Social implications of the Internet, *Annual review of sociology*, 27(1), pp. 307–336.

Dobber T, et al. (2023) Shielding citizens? Understanding the impact of political advertisement transparency information, *New Media & Society*, p. 14614448231157640. Available at: <https://doi.org/10.1177/14614448231157640>.

Ferree MM, Gamson WA (2002) Four Models of the Public Sphere in Modern Democracies, p. 37.

Fletcher R, et al. (2019) Polarisation and the news media in Europe: a literature review of the effect of news use on polarisation across Europe. Available at: [http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634413/EPRS_STU\(2019\)6344_13_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634413/EPRS_STU(2019)6344_13_EN.pdf) (Accessed: 19 May 2020).

Flonk D, Jachtenfuchs M, Obendiek AS (2020) Authority Conflicts in Internet Governance: Liberals vs. Sovereignists?. *Global Constitutionalism* 9(2), pp. 364–86.

Flonk D (2021) Emerging Illiberal Norms: Russia and China as Promoters of Internet Content Control. *International Affairs* 97(6), pp. 1925–44.

Garimella K, Eckles D (2017) Image based Misinformation on WhatsApp. Available at: https://gvrkiran.github.io/content/whatsapp_image_misinformation.pdf.

Gibson R, Römmele A, Williamson A (2014) Chasing the Digital Wave: International Perspectives on the Growth of Online Campaigning, *Journal of Information Technology & Politics*, 11(2), pp. 123–129. Available at: <https://doi.org/10.1080/19331681.2014.903064>.

Gibson RK (2015) Party change, social media and the rise of “citizen-initiated” campaigning, *Party Politics*, 21(2), pp. 183–197. Available at: <https://doi.org/10.1177/1354068812472575>.

Gil De Zúñiga H, Diehl T (2019) News finds me perception and democracy: Effects on political knowledge, political interest, and voting, *New media & society*, 21(6), pp. 1253–1271.

Gillespie T (2017) Governance of and by platforms, *SAGE handbook of social media*, pp. 254–278.

Gillespie T (2018) Platforms are not intermediaries, *Georgetown Law Technology Review*, 2(2), pp. 198–216.

Gorwa R (2021) Elections, Institutions, and the Regulatory Politics of Platform Governance: The Case of the German NetzDG. *Telecommunications Policy, Norm entrepreneurship in Internet Governance*, 45(6): 102145. <https://doi.org/10.1016/j.telpol.2021.102145>.

Grimmelmann J (2015) The Virtues of Moderation. preprint. LawArXiv. Available at: <https://doi.org/10.31228/osf.io/qwxf5>.

Guess AM, et al. (2021) The consequences of online partisan media, *Proceedings of the National Academy of Sciences*, 118(14), p. e2013464118. Available at: <https://doi.org/10.1073/pnas.2013464118>.

Gursky J, Woolley S (2021) Countering disinformation and protecting democratic communication on encrypted messaging applications. *Foreign Policy at Brookings*. Available at: https://www.brookings.edu/wp-content/uploads/2021/06/FP_20210611_encryption_gursky_woolley.pdf

Habermas J (1991) *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*. MIT Press.

Habermas J (2022) Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere, *Theory, Culture & Society*, 39(4), pp. 145–171. Available at: <https://doi.org/10.1177/02632764221112341>.

Hadgu AT, Garimella K and Weber I (2013) Political hashtag hijacking in the US, in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 55–56.

Hofmann J (2005) *Internet Governance: A Regulative Idea in Flux*. SSRN Scholarly Paper. Rochester, NY, September 1, 2005. <https://papers.ssrn.com/abstract=2327121>.

Jaurisch J (2022) *Strengthening EU Proposals on Deceptive Platform Design*.

Klinger U, et al. (2022) From the fringes into mainstream politics: intermediary networks and movement-party coordination of a global anti-immigration campaign in Germany, *Information, Communication & Society*, pp. 1–18. Available at: <https://doi.org/10.1080/1369118X.2022.2050415>.

Klinger U, Koc-Michalska K, Russmann U (2022) Are Campaigns Getting Uglier, and Who Is to Blame? Negativity, Dramatization and Populism on Facebook in the 2014 and 2019 EP Election Campaigns, *Political Communication*, 0(0), pp. 1–20. Available at: <https://doi.org/10.1080/10584609.2022.2133198>.

Klinger U, Svensson J (2015) The emergence of network media logic in political communication: A theoretical approach', *New Media & Society*, 17(8), pp. 1241–1257. Available at: <https://doi.org/10.1177/1461444814522952>.

Klonick K (2020) The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression, *The Yale Law Journal*, p. 82.

Knüpfer C, Hoffmann M, Voskresenskii V (2020) Hijacking MeToo : transnational dynamics and networked frame contestation on the far right in the case of the '120 decibels' campaign, *Information, Communication & Society*, pp. 1–19. Available at: <https://doi.org/10.1080/1369118X.2020.1822904>.

Lazer DMJ, et al. (2020) Computational social science: Obstacles and opportunities, *Science*, 369(6507), pp. 1060–1062. Available at: <https://doi.org/10.1126/science.aaz8170>.

Lorenz-Spreen P, et al. (2019) Accelerating dynamics of collective attention, *Nature Communications*, 10(1), p. 1759. Available at: <https://doi.org/10.1038/s41467-019-09311-w>.

Lorenz-Spreen P, et al. (2023) A systematic review of worldwide causal and correlational evidence on digital media and democracy, *Nature Human Behaviour*, 7(1), pp. 74–101. Available at: <https://doi.org/10.1038/s41562-022-01460-1>.

Mager A, Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified, *New media & society*. SAGE Publications Sage UK: London, England.

Mancini P (2013) Media Fragmentation, Party System, and Democracy, *The International Journal of Press/Politics*, 18(1), pp. 43–60. Available at: <https://doi.org/10.1177/1940161212458200>.

Pfetsch B (2018) Dissonant and Disconnected Public Spheres as Challenge for Political Communication Research, *Javnost - The Public*, 25(1–2), pp. 59–65. Available at: <https://doi.org/10.1080/13183222.2018.1423942>.

Reckwitz A (2020) The Society of Singularities. In *The Society of Singularities*, 141–54. De Gruyter. <https://doi.org/10.1515/9783110669398-009>.

Scharkow M, et al. (2020) How social network sites and other online intermediaries increase exposure to news, *Proceedings of the National Academy of Sciences*, 117(6), pp. 2761–2763. Available at: <https://doi.org/10.1073/pnas.1918279117>.

Schulz W (2018) Regulating Intermediaries to Protect Privacy Online – the Case of the German NetzDG, p. 15. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3216572 .

Smythe DW (1981) On the audience commodity and its work. *Media and cultural studies: Keyworks*, 230, 256.

Srnicek N (2017) *Platform Capitalism*. John Wiley & Sons.

Sternisko A, Cichocka A, Van Bavel JJ (2020) The dark side of social movements: social identity, non-conformity, and the lure of conspiracy theories, *Current Opinion in Psychology*, 35, pp. 1–6. Available at: <https://doi.org/10.1016/j.copsy.2020.02.007>.

Stockmann D (2023) Tech companies and the public interest: the role of the state in governing social media platforms, *Information, Communication & Society*, 26(1), pp. 1–15. Available at: <https://doi.org/10.1080/1369118X.2022.2032796>.

Strömbäck J (2008) Four Phases of Mediatization: An Analysis of the Mediatization of Politics, *The International Journal of Press/Politics*, 13(3), pp. 228–246. Available at: <https://doi.org/10.1177/1940161208319097>.

Theocharis Y (2015) The Conceptualization of Digitally Networked Participation, *Social Media + Society*, 1(2), p. 205630511561014. Available at: <https://doi.org/10.1177/2056305115610140>.

Theocharis Y, et al. (2022) Platform affordances and political participation: how social media reshape political engagement, *West European Politics*, 0(0), pp. 1–24. Available at: <https://doi.org/10.1080/01402382.2022.2087410>.

Theocharis Y, Jungherr A (2021) Computational social science and the study of political communication, *Political Communication*, 38(1–2), pp. 1–22.

Tinati R, et al. (2014) Big data: Methodological challenges and approaches for sociological analysis, *Sociology*, 48(4), pp. 663–681.

Touraine A (1995) *Beyond Social Movements?*, in S.M. Lyman (ed.) *Social Movements: Critiques, Concepts, Case-Studies*. London: Palgrave Macmillan UK (Main Trends of the Modern World), pp. 371–393. Available at: https://doi.org/10.1007/978-1-349-23747-0_16.

Tromble R (2021) 'Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age', *Social Media + Society*, 7(1), p. 205630512198892. Available at: <https://doi.org/10.1177/2056305121988929>.

Vaccari C, et al. (2021) *Outside the Bubble: Social Media and Political Participation in Western Democracies*. Oxford, New York: Oxford University Press (Oxford Studies in Digital Politics).

Yablokov I (2015) Conspiracy theories as a Russian public diplomacy tool: The case of Russia Today (RT), *Politics*, 35(3–4), pp. 301–315.

Zimmer M (2010) "But the data is already public": on the ethics of research in Facebook, *Ethics and information technology*, 12, pp. 313–325.

Zuboff S (2019) Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum*, 28(1), 10–29. <https://doi.org/10.1177/1095796018819461>.

Appendix

RT network	Nodes	Edges	Modularity	C1 (in %)	C2 (in %)	C3 (in %)
2021						
#BTW21	91,789	225,925	0.429	69.08	15.65	10.56
#AfD	15,572	47,947	0.385	53.95	42.44	0.2
#CSU	5,594	7,065	0.565	52.52	24.53	8.4
#CDU	18,975	34,004	0.347	72.96	21.59	0.11
#FDP	8,845	11,890	0.572	55.47	18.89	15.05
#GRÜNE	7,116	8,819	0.56	50.87	38.08	0.7
#LINKE	4,292	5,290	0.578	48.9	29.64	13.68
#SPD	10,608	15,782	0.525	52.75	37.44	0.39
#Baerbock	13,509	21,230	0.528	60.97	31.57	3.09
#Laschet	25,991	51,893	0.179	84.96	11.67	0.05
#Scholz	11,792	16,683	0.453	67.97	27.33	0.11
2017						
#BTW17	72,745	168,239	0.254	86.03	9.69	0.17
#AfD	27,713	68,781	0.497	75.10	20.57	0.10
#CSU	2,084	2,262	0.766	25.77	24.86	13.77
#CDU	4,989	7,175	0.630	36.10	33.81	8.84
#FDP	4,160	5,325	0.742	27.00	23.65	21.39
#GRÜNE	2,570	2,976	0.784	27.70	11.71	10.82
#LINKE	2,224	2,952	0.410	57.55	18.26	7.19
#SPD	6,578	9,310	0.717	28.15	25.37	15.72

Table A1: Summary of collected retweet networks

Screenname	Retweeters	Number of Retweets	Community Membership	Partisan
heuteshow	6556	9731	Red	No
tagesschau	5249	7164	Red	No
zeitonline	4527	5100	Red	No
extra3	3190	4299	Red	No
swagenknecht	1398	3010	Red	Yes
fraukepetry	1401	2821	Blue	Yes
dielinke	1033	2731	Red	Yes
spiegelonline	2071	2574	Red	No
zdfheute	2029	2565	Red	No
die_gruenen	1221	2554	Red	Yes
marspet	1731	2257	Red	No
europeselects	1200	1927	Blue	No
kiser_let	1358	1740	Red	No
raindiercks	1636	1714	Red	No
fdp	866	1610	Red	Yes
deintherapeut	1559	1592	Red	No
zdf	1297	1540	Red	No
wahlrecht_de	904	1533	Blue	No
piratenpartei	385	1385	Red	Yes
br24	950	1322	Red	No
wahlleiter_bund	1114	1278	Red	No
matthiasmeisner	702	1140	Red	No
fackjugoehnte	1077	1078	Red	No
lawyerberlin	506	949	Blue	Yes
rbb24	748	880	Blue	No
krk979	450	785	Blue	Yes
mundaufmachen	421	783	Blue	Yes
poggenburgandre	466	774	Blue	Yes
niggi	609	669	Red	No

Table A2: Top 30 most retweeted accounts in #BTW17

Screenname	Number of Retweeters	Number of Retweets	Community Membership	Partisan
m_ziesmann	2436	4161	Blue	No
heuteshow	2929	3636	Red	No
hagen	2992	3163	Red	No
watch_union	1583	2949	Red	Yes
wahlen_de	1428	2658	Red	No
die_gruenen	1765	2656	Red	Yes
jumpsteady	2146	2532	Red	No
afd	1163	2522	Blue	Yes
volksverpetzer	2083	2515	Red	Yes
europeselects	1346	2468	Gray	No
wahlrecht_de	1400	2291	Red	No
dielinke	1226	2167	Red	Yes
tagesschau	1599	2120	Red	No
bild	1197	1963	Blue	Yes
compact	1335	1868	Red	No
theeconomist	1350	1806	Gray	No
abaerbock	1787	1787	Red	Yes
dwnews	1317	1753	Gray	No
afd_muenster	764	1669	Blue	Yes
der_postillon	1542	1605	Red	No
stephanschmidt	1540	1542	Red	Yes
piratenpartei	440	1475	Red	Yes
konstantinnotz	1305	1446	Red	Yes
alice_weidel	1256	1430	Blue	Yes
iblali	1221	1347	Red	No
kaffeecup	1239	1319	Red	No
lgbeutin	910	1244	Red	Yes
hartes_geld	852	1201	Red	Yes
raykanders	1059	1186	Red	No
drwaumiau	1109	1156	Red	No

Table A3: Top 30 most retweeted accounts in #BTW21

Cluster Membership	2017	2021
@afd	Blue	Blue
@CDU	Red	Blue
@CSU	Red	Blue
@fdp	Red	Blue
@Die_Gruenen	Red	Red
@dieLINKE	Red	Red
@spdde	Red	Red

Table A4: Cluster membership of official federal-level party accounts in 2017 and 2021

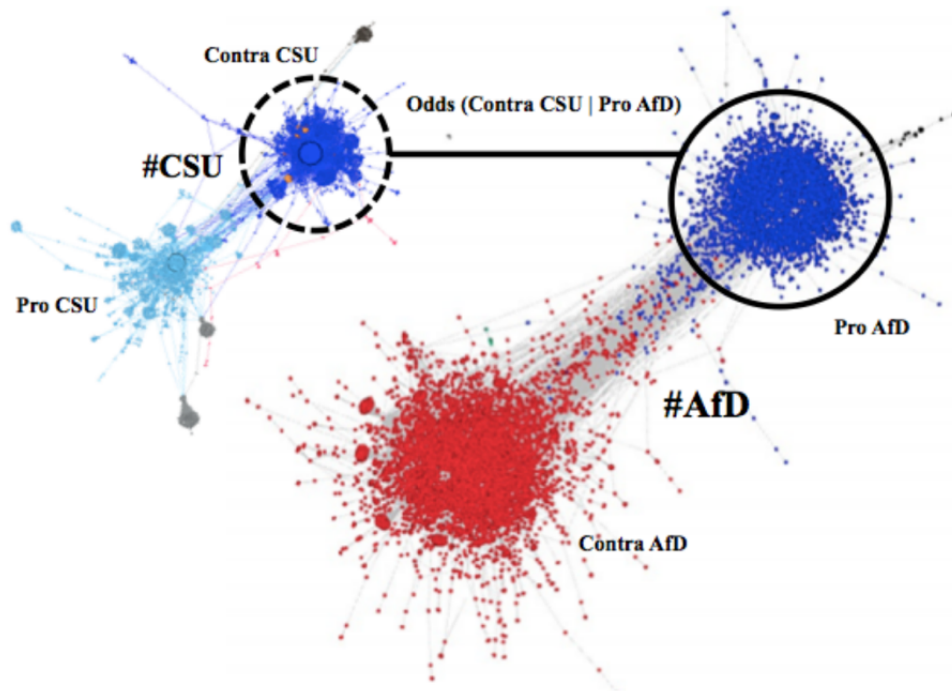


Figure A 1: Illustration of 'hashjacking' in Darius and Stephany (2019)

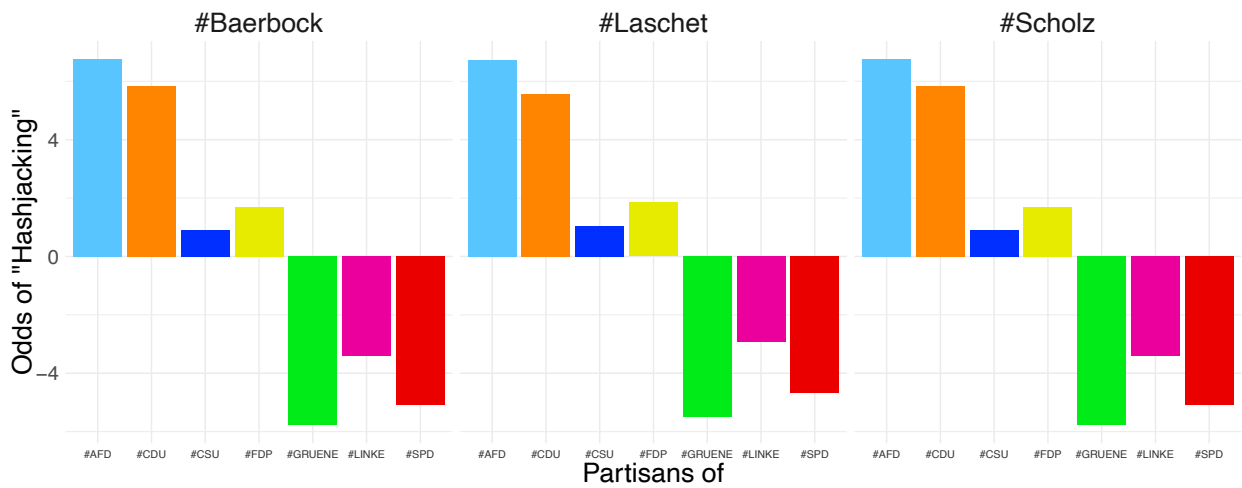


Figure A 2: Log-likelihoods of the co-occurrence of partisans in the two major network clusters



Figure A 3: Example post 1

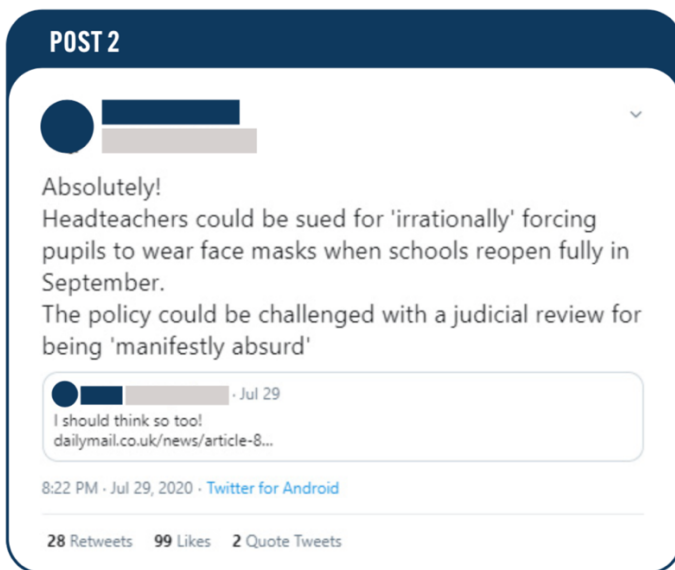


Figure A 4: Example post 2



Figure A 5: Example post 3

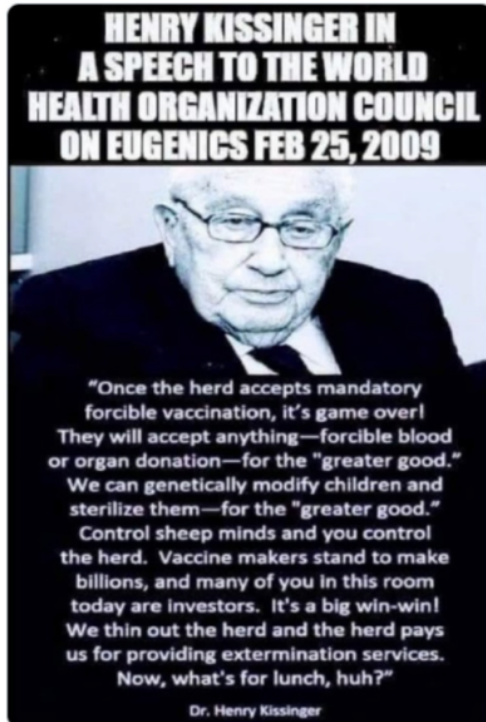


Figure A 6: Example post 4

POST 5

The #WHO #BillGates #Soros #MegaCorps World CONTROL + #Eugenics through #Vaccination + #5G is where #COVID19 #Convirus is conveniently #LOCKDOWN-ing the world.

There are proven cures (eg #Chloroquine) to #CV19. Use them!
#NO2CVaccine!
RECLAIM our freedom May1
#MayDayNormalDay



5:37 PM · Apr 11, 2020 · Twitter Web App

126 Retweets 129 Likes 16 Quote Tweets

Figure A 7: Example post 5

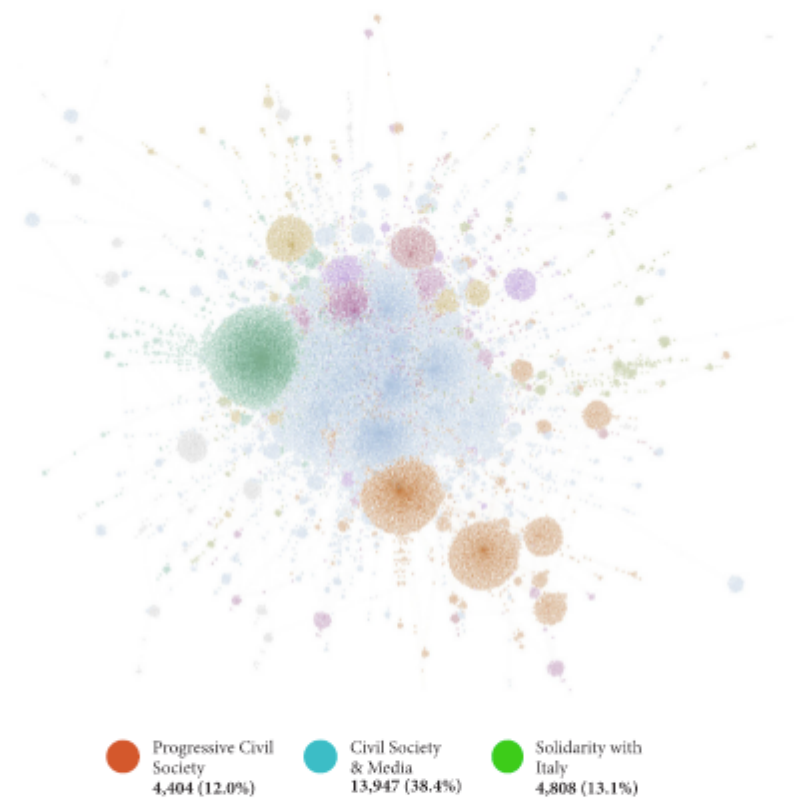


Figure A 8: #Lockdown RT networks with cluster names. Top: T₁ (January 1, 2020 - March 22, 2020). Bottom: T₂ (March 23, 2020 - June 10, 2020).

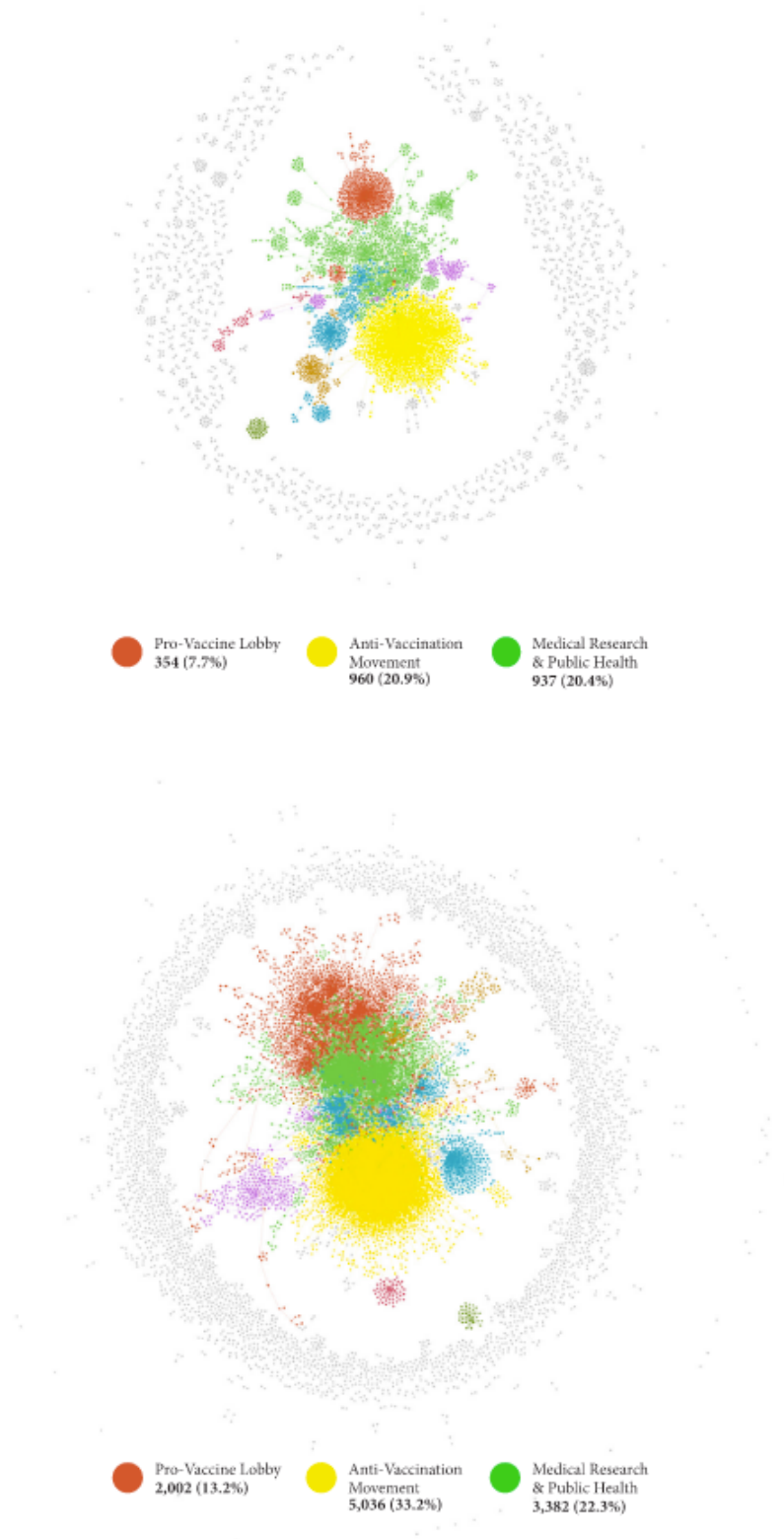


Figure A 9: #Vaccines RT networks with cluster names. Top: T1 (January 1, 2020 - March 22, 2020). Bottom: T2 (March 23, 2020 - June 10, 2020).

List of papers

This cumulative dissertation project is comprised of three individual studies and an envelope (Chapter 1, 2, and 6). The studies have been published and can be cited using the following references:

- Chapter 3: Darius, Philipp. "Who Polarizes Twitter? Ideological Polarisation, Partisan Groups and Strategic Networked Campaigning on Twitter during the 2017 and 2021 German Federal Elections 'Bundestagswahlen.'" *Social Network Analysis and Mining* 12, no. 1 (December 2022): 151. <https://doi.org/10.1007/s13278-022-00958-w>.
- Chapter 4: Darius, Philipp, and Michael Urquhart. "Disinformed Social Movements: A Large-Scale Mapping of Conspiracy Narratives as Online Harms during the COVID-19 Pandemic." *Online Social Networks and Media* 26 (November 2021): 100174. <https://doi.org/10.1016/j.osnem.2021.100174>.
- Chapter 5: Cows, Josh, Philipp Darius, Dominiquo Santistevan, and Moritz Schramm. "Constitutional Metaphors: Facebook's 'Supreme Court' and the Legitimation of Platform Governance." *New Media & Society*, April 5, 2022, 14614448221085560. <https://doi.org/10.1177/14614448221085559>.