Fakultät für Informatik
Technische Universität München

# Multiplatform Analysis of Political Communication on Social Media

**Juan Carlos Medina Serrano**

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

**Vorsitzender:**
Prof. Dr. Uwe Baumgarten

**Prüfende der Dissertation:**
1. Prof. Dr. Simon Hegelich
2. Prof. Dr. Jürgen Pfeffer

Die Dissertation wurde am 05.07.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 12.11.2021 angenommen.

*Dedicated to my parents*

# Acknowledgements

First, I would like to thank Simon Hegelich for supporting me and giving me advice during the last four years. I am grateful for the opportunity I had to be part of his team and for letting me explore any research idea. Simon made my Ph.D. feel like a fun experience and never like a stressful job.

Second, I would like to thank Jürgen Pfeffer for accepting being my co-advisor, for his advice, and for reviewing this thesis.

I would not have been able to complete this thesis without the help of my co-authors, Morteza Shahrezaye, and Orestis Papakyriakopoulos. Thank you Morry for convincing me into pursuing a Ph.D. and later for helping me in my projects. Thank you Orestis for the great collaborations and projects we had together. Everyone in the political data science team had a contribution to this thesis and I am grateful to them and for the fun evenings that we enjoyed together.

I am grateful to have amazing parents who have supported me all the time. Thank you for always being there for me. This thesis is dedicated to you.

Finally, I would like to thank my friends and my girlfriend Aline for their support and help.

# Abstract

Social media has transformed political communication. Nowadays, political discourse occurs mostly in public and private digital spaces. Originally, social media was regarded as a tool that gave everyone a voice, especially in oppressive systems. However, it has proved to also have dangerous side-effects as it potentiates negative sentiments ingrained in society, such as polarization, extremist rhetoric, and racism. During the second decade of the twenty first century, social media companies have been in the news limelight, seen as important stakeholders that oversee and indirectly control the flow of communication. A new area of research has emerged, which studies the effect of algorithms and design on political communication. However, social media remains a complex ecosystem, and specific platform analyses do not suffice to explain the dynamics of politics-laden digital spaces. Multiplatform analysis consists of evaluating different platforms based on a common theme, actor, or political event. Enough data makes possible the performance of a cross-platform analysis that additionally investigates the diffusion of information between platforms.

This thesis presents a framework for understanding three of the main stakeholders in political discourse on social media: political actors, partisan users, and bad-natured agents. The last encompasses different types of entities that make the political conversation toxic, irrespective of their degree of automation. Throughout a series of case studies, this work tries to understand these stakeholders and their interactions on different social media platforms. This thesis underlines the importance of monitoring social media to gain insights into these interactions. In the case of political actors, this thesis investigates the rise of the populist right-wing German political party, the Alternative für Deutschland (AfD). Additionally, it studies the online political advertising of the AfD and the other main German political parties, in the months leading up to the 2019 European election. Regarding partisan users, this work presents the first analysis of political communication on TikTok. The analysis also focuses on the interactions between Democratic and Republican users in the United States and on the new distinct feature called *the duet*. The final case study concentrates on bad-natured agents on YouTube, in the context of misinformation about the COVID-19 pandemic. With the help of state-of-the-art natural language processing (NLP) methods, the case study shows how to detect conspiratorial videos based on the user comments. To conclude, this thesis presents the challenges and future research needed to study political communication on new platforms and how to address the social problems that have manifested in the offline world.

# Zusammenfassung

Soziale Netzwerke haben die politische Kommunikation verändert. Heutzutage findet der politische Diskurs vor allem in öffentlichen und privaten digitalen Räumen statt. Ursprünglich wurden Soziale Netzwerke als ein Instrument angesehen, das jedem Bürger, besonders in repressiven Systemen, eine Stimme verleiht. Es hat sich jedoch gezeigt, dass es auch gefährliche Nebenwirkungen gibt, da in der Gesellschaft verankerte negative Phänomene wie Polarisierung und extremistische Rhetorik potenziert werden können. Social-Media-Unternehmen stehen im Fokus der Nachrichten, da sie als wichtige Akteure angesehen werden, die den Kommunikationsfluss überwachen und indirekt kontrollieren. Es ist ein neuer Forschungsbereich entstanden, der die Auswirkungen von Algorithmen und Plattformdesign auf die politische Kommunikation untersucht. Soziale Netzwerke sind jedoch nach wie vor ein komplexes Ökosystem und spezifische Plattformanalysen reichen nicht aus, um die Dynamik politischer digitaler Räume zu erklären. Eine Multiplattform-Analyse besteht darin, verschiedene Plattformen anhand eines gemeinsamen Themas, Akteurs oder politischen Ereignisses auszuwerten. Wenn genügend Daten zur Verfügung stehen, ist es möglich, eine "Cross-Plattform Analyse" durchzuführen, die darüber hinaus die Diffusion von Informationen zwischen Plattformen untersucht.

Diese Arbeit stellt einen Rahmen vor, um drei der Hauptakteure des politischen Diskurses in Sozialen Netzwerken zu untersuchen: politische Akteure, "partisan"-Nutzer und "bad-natured agents". Letztere umfassen verschiedene Arten von Entitäten, die die politische Konversation unabhängig von ihrem Automatisierungsgrad kontaminieren. Anhand einer Reihe von Fallstudien versucht diese Arbeit, diese Akteure und ihre Interaktionen auf verschiedenen Plattformen zu untersuchen. Dabei wird die Bedeutung des Monitorings Sozialer Netzwerke hervorgehoben, um Einblicke in deren Interaktionen zu erlangen. Im Bezug auf die politischen Akteure untersucht diese Arbeit den Aufstieg der rechtspopulistischen Partei Alternative für Deutschland (AfD). Darüber hinaus untersucht sie die politische Online-Werbung der AfD und der anderen großen deutschen Parteien in den Monaten vor der Europawahl 2019. Im Hinblick auf "partisan"-Nutzer präsentiert diese Arbeit die erste Analyse der politischen Kommunikation auf TikTok. Die Analyse konzentriert sich auch auf die Interaktionen zwischen demokratischen und republikanischen Nutzern in den Vereinigten Staaten und speziell auf das Feature namens "duet". Die letzte Fallstudie untersucht "bad-natured agents" auf YouTube im Kontext von Falschmeldungen über die COVID-19-Pandemie. Mit Hilfe von "Natural Language Processing" (NLP) zeigt die Fallstudie, wie man konspirative Videos anhand der Nutzerkommentare erkennen kann. Abschließend werden in dieser Arbeit die Herausforderungen und der zukünftige Forschungsbedarf für die Untersuchung politischer Kommunikation auf neuen Plattformen und die Bewältigung der sozialen Probleme, die sich in der Offline-Welt manifestiert haben, dargestellt.

# Peer-Reviewed Publications

This thesis is a compilation of five first-author peer-reviewed papers published in international conferences. Table 1 shows the bibliographic references. This thesis provides a framework to understand how to study political communication on social media. It specifically focuses on the political landscape of Germany and the United States.

| Reference |
| --- |
| **[1]** J. C. M. Serrano, M. Shahrezaye, O. Papakyriakopoulos, and S. Hegelich<br>**The Rise of Germany's AfD: A Social Media Analysis**<br>In *Proceedings of the 10th International Conference on Social Media and Society*, pages 214–223, 2019. |
| **[2]** J. C. M. Serrano, O. Papakyriakopoulos, M. Shahrezaye, and S. Hegelich<br>**The Political Dashboard: A Tool for Online Political Transparency**<br>In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 983–985, 2020 |
| **[3]** J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich.<br>**Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections**<br>In *International Conference on Social Media and Society*, pages 111–121, 2020 |
| **[4]** J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich<br>**Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok**<br>In *12th ACM Conference on Web Science*, pages 257–266, 2020 |
| **[5]** J. C. M. Serrano, O. Papakyriakopoulos, and S. Hegelich<br>**NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube**<br>In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, 2020. |

**Table 1:** Peer reviewed first author publications.

Additionally, I contributed to the following peer-reviewed publications:

- [6] O. Papakyriakopoulos, M. Shahrezaye, A. Thieltges, J. C. M. Serrano, and S. Hegelich. **Social Media und Microtargeting in Deutschland**. *Informatik Spektrum*, 40(4):327–335, 2017.

- [7] O. Papakyriakopoulos, S. Hegelich, M. Shahrezaye, and J. C. M. Serrano. **Social media and microtargeting: Political Data Processing and the Consequences for Germany**. *Big Data & Society*, 5(2):2053951718811844, 2018.

- [8] M. Shahrezaye, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. **Estimating the Political Orientation of Twitter Users in Homophilic Networks**. In *2019 AAAI Spring Symposium Series*, 2019.

- [9] O. Papakyriakopoulos, M. Shahrezaye, J. C. M. Serrano, and S. Hegelich. **Distorting Political Communication: The Effect Of Hyperactive Users In Online Social Networks**. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 157–164. IEEE, 2019.

- [10] M. Shahrezaye, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. **Measuring the Ease of Communication in Bipartite Social Endorsement Networks: A Proxy to Study the Dynamics of Political Polarization**. In *Proceedings of the 10th International Conference on Social Media and Society*, pages 158–165. ACM, 2019.

- [11] O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. **Political Communication on Social Media: A Tale of Hyperactive Users and Bias in Recommender Systems**. *Online Social Networks and Media*, 15:100058, 2020.

- [12] O. Papakyriakopoulos, S. Hegelich, J. C. M. Serrano, and F. Marco. **Bias in Word Embeddings**. *FAT* '20*, page 446–457, New York, NY, USA, 2020. Association for Computing Machinery

- [13] O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. **The spread of COVID-19 conspiracy theories on social media and the effect of content moderation**. *Harvard Kennedy School Misinformation Review*, 2020.

# Contents

**Acknowledgements** **v**

**Abstract** **vii**

**Zusammenfassung** **ix**

**Peer-Reviewed Publications** **xi**

**Contents** **xiii**

**1 Introduction** **1**
1.1 A Reality Check . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
1.2 Political Communication on Social Media . . . . . . . . . . . . . . . . . . 3
1.2.1 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
1.2.2 Main Framework . . . . . . . . . . . . . . . . . . . . . . . . . . 5
1.3 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
1.4 Selected Topics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
1.4.1 Microtargeting . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
1.4.2 Social Bots . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
1.4.3 Fake News Detection . . . . . . . . . . . . . . . . . . . . . . . . 11
1.4.4 Cross-Platform Social Media Analysis . . . . . . . . . . . . . . . . 12
1.5 Thesis Structure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
1.6 Theoretical Contributions . . . . . . . . . . . . . . . . . . . . . . . . . 16

**2 Methods** **17**
2.1 The Different Paradigms of Data Science . . . . . . . . . . . . . . . . . . 17
2.1.1 Frequentist Statistics . . . . . . . . . . . . . . . . . . . . . . . . 18
2.1.2 Machine Learning . . . . . . . . . . . . . . . . . . . . . . . . . . 21
2.1.3 Bayesian Statistics . . . . . . . . . . . . . . . . . . . . . . . . . 24
2.1.4 Bayesian Machine Learning . . . . . . . . . . . . . . . . . . . . . 27
2.1.5 Deep Learning . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
2.2 Methodological Contributions . . . . . . . . . . . . . . . . . . . . . . . 34

**3 Social Media Monitoring** **37**
3.1 The Political Dashboard: A Tool for Online Political Transparency . . . . 37

**4 Political Actors** **41**
4.1 The Rise of Germany's AfD: A Social Media Analysis . . . . . . . . . . . 41
4.2 Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections . . . . . . . . . . 53

**5 Partisan Users** **65**
5.1 Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok . . . 65

**6 Misinformation and Conspiracy Theories** **77**
6.1 NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube . . . 77

**7 Discussion** **85**
7.1 Main Framework (Revisited) . . . 85
7.2 Implications & Challenges . . . 88
7.3 Future work . . . 89

**Bibliography** **93**

**A Misinformation in the 2017 German Election** **109**

**B Coordinated and Suspended Accounts on Twitter in the run-ups to General Elections** **121**

# 1 Introduction

*The real problem with humanity is the following: We have paleolithic emotions, medieval institutions, and god-like technology.*

**- Edward O. Wilson [14]**

Social media is now an important part of many people's lives [15]. It defines how people communicate, interact, and form groups in the digital world. This has repercussions for how society evolves over time [16]. One of the main aspects that these new media have transformed is political communication. Even though social media platforms were not designed to foster political conversations, they have become central spaces for political exchange, communication, and campaigning. Citizens discuss political issues, externalize their political preferences, and take part in online groups for political activism [17, 18]. At the time that social media started to become ubiquitous in daily life, it provided hope for a more diverse, open, and democratic political discourse [19, 20]. Especially, as it provided a new space for every politician to share his or her message [21], and could help to raise voices that authoritarian regimes have repressed [22, 23]. The hope depended on social media providing a space for information [24], connection [25], civic responsibility [26], and diversity [27].

Nowadays, politics plays a large role in social media. Political parties have developed new methods and tools for externalizing party attitudes and evaluating the reactions of potential voters. They deploy digital political campaigns in order to influence public opinion, especially in the form of personalized advertising [28]. Political parties train data-intensive models used for decision-making, using large databases that contain demographic and personal information [29, 30]. However, negative events have shadowed the benefits that social media brings to political discourse. Social media can foster negative and toxic content that damages political communication. In turn, this makes social media spaces a fertile ground for various actors to exploit. Notable examples are automated, fake, and militant accounts that spread misinformation as part of disinformation campaigns targeting vulnerable groups [31, 32]. Detecting the different types of manipulation techniques on social media is a considerable challenge, given that they evolve over time with new strategies to fool the audience, the politicians, and maybe even the researchers.

The users that produce political content on social media are not the only actors that influence political communication. The algorithms that decide what to present to users directly impact it. They can even lead to a reality construction for the users [33]. Through these automated decision-makers, social media platforms play a role in forming users' opinions. The knowledge that users obtain from these channels then can be transformed into action in both the online and offline worlds. The closeness of algorithms and politics has led to the emergence of data politics [34], which deals with the functionality of algorithms [35], the biases they may introduce [36, 37], and, most importantly, how do they influence users and social groups [38, 39]. Researchers have attempted to investigate whether social media helps to the democratization of society or has a net negative political impact [40, 41, 42]. This question relates to studies on the detrimental aspects of using social media, which expose how the data business model behind platforms does not always align with ensuring civic political discourse [43].

This dissertation presents several case studies of political communication on social media. It focuses on the online political landscape of Germany and the United States. It presents a framework that identifies the different stakeholders of political communication on online platforms. The cases show the importance of monitoring social media to understand political communication. Finally, this work calls for the use of statistical algorithms to defend the discourse from bad-natured agents, but warns of the oversimplification of content moderation.

## 1.1 A Reality Check

At the time of this writing, Joe Biden has been officially inaugurated as the 46th president of the United States. The era of Donald Trump comes to its end (for now). In the last three years since I started my dissertation, politics and online digital platforms have lived through many ups and downs. I feel the need to perform a quick reality check of what were the events that occurred in this period, as they strongly influenced my research. This also gives to you, the reader, the political context in which the included publications were written. This period starts with the Cambridge Analytica data scandal in 2017, where the personal data of millions of Facebook users were acquired without their consent. Some assume that the data helped in electing Donald Trump [44]. Although the veracity of this assumption is contested, this event changed how the public and the government regard social media. It is no longer all about connecting people for the social good [45]. It is also a helpful tool for extremist voices. In my first publication [1], I studied how the Alternative für Deutschland, Germany's far-right populist party successfully employed social media to spread its message.

As a response to the Cambridge Analytica scandal, Facebook and the other major tech companies pledged more transparency. In 2019, they introduced ad libraries that reported the advertisements with political content. These libraries provided previously undisclosed data regarding online political campaigns. I performed the first analysis of these libraries in the context of the European election in 2019 [3]. Apart from transparency, the social media companies promised to increase their content moderation of misinformation. After the 2016 U.S. election, the topic of disinformation on social media became a relevant topic of research. However, the real consequences of disinformation campaigns for the offline world were not evident. This all changed with the COVID-19 pandemic that not only changed the world but it also created a *misinfodemic* [46]. In this light, I studied how conspiracy theories spread on YouTube, and I developed an algorithm to detect videos with conspiratorial content [5].

The digital world changes fast, and social media evolves at a similar pace. In 2019, TikTok, a new video-sharing social media platform, arose to popularity. It became the most downloaded app in 2020. As on Facebook and Twitter, politics play an important role on TikTok. I performed the first analysis of political communication on this platform [4]. In ten years, the social media ecosystem may be completely different. Thus, research that is relevant now may not seem so in the near future. Nevertheless, I hope this work can help future researchers in the study of political communication on social media.

## 1.2 Political Communication on Social Media

Communication between the government and the people is central to any political system [47]. The political discourse permeates the civic discourse in a society [48] and is an essential part of a democracy [49]. Denton and Woodward [50] characterize political communication in terms of the intentions of its senders to influence the political environment. Therefore, its content and purpose characterize political communication. The elements of political communication reside in the actors that undertake it: citizens, the media, and political organizations [51]. Social media has allowed for a large part of political communication to move to digital spaces. These new digital public squares allow users to interact and exchange ideas. Social media platforms also create online groups that share interests or ideologies. Facebook, Twitter, TikTok, YouTube[1], and other platforms have transformed how political communication takes place. In this subsection, I present first related work and then the main framework of this thesis.

### 1.2.1 Related Work

There is vast research on the topic of political communication and social media. In this subsection, I only consider general and relevant studies. This omits countless research papers on specific political events in countries around the world (e.g., general elections, debates, social issues, policy-related topics). Such a compilation is outside the scope of this thesis. Coelho et al. [52] provide a literature review of the importance of public and political communication through social media. Others focus on the disruption by social media of political communication [53, 54].

The first general topic on social media and political communication is the effect of the new media on democracy. Iosifidis and Wheeler [55] tried to find out if social media is an aid to democratic representation or contributes to a greater destabilization of modern politics. They conclude that engagement with social media in preserving political consensus is highly questionable, and it has exposed the fissures in modern democracies. The work of Van Aelst et al. [56] supports this, focusing on the main concerns that social media present as a challenge to democracy: a high-choice media environment, declining quality, and diversity of news; increasing media concentration, increasing fragmentation and polarization, and increasing inequality in political knowledge.

The second general topic focuses on the usage of social media by politicians. Stier [57] investigated how politicians use different social media platforms for political communication. Similarly, Hegelich et al. [58] investigated how politicians use social media to interact with the public. Varol et al. [59] studied how promoted political campaigns dynamically develop on social media. Xi et al. [60] tried to understand how images on social media convey the political ideology of politicians. With text, images and videos, social media gives politicians the ability to control the online political landscape. For example, through his political messages, Donald Trump successfully diverted the media from topics that he considered threatening [61].

The third general topic focuses on user behavior in politically-laden circumstances. Himelboim et al. [62] find that interpersonal informational trust is positively associated with the perception of online activities as political participation. Kruikemeer et al. [63]

[1] Although YouTube is not a social media platform per se, it is often considered under the same umbrella. YouTube videos are shared on social media platforms, they have a comment section, and users can follow channels.

show that highly interactive and personalized online communication increases citizens' political involvement. This effect differs across age groups, and it can directly influence people's inclination to participate politically [64]. Studies in political participation show that users are likely to argue over wide ideological divides and increasingly likely to engage with those who differ from them [65]. This suggests that users have an interest in engaging in prolonged political discussions. Barnidge [66] studied the exposure to political disagreement in social media versus face-to-face and anonymous online settings. The political news diet that users have implies different political alignments [67]. In his seminal paper, Barbera [68] demonstrates an approach to determining ideological positions based on the social network of users.

Social media gives researchers the possibility of performing large studies of behavior [69] and classical political-science research on large datasets [70]. Facebook performed one of the most famous examples of large-scale experiments on voter mobilization, with 61 million users [71]. Researchers should be careful when interpreting social media results, as the sample is often not representative of the general population [72]. However, social media can complement survey data as an indicator of changes in public opinion [73] or as signals in unpolled topics [74]. Stieglitz and Dang-Xuan [75] propose a framework for social media analytics in a political context, from data tracking and data analysis to analytical methods that gather insights into political discussions. Additionally, network analysis enables depicting political communication as graphs, where users are nodes, and their connections represent communication patterns. Pfeffer [76] presents a thorough review of political networks visualizations.

The thesis by Kalsnes [77] is the most similar to the present dissertation; she tries to understand the logic of social media in the presence of political communication. However, the similarities only apply to the general setting; the study cases and focal points differ substantially. Both works complement each other by focusing on the impact of social media and how disruptive technology translates into real-world events. Whereas Kalsnes focuses on the new mechanisms for attention, visibility, and popularity on social media platforms, this thesis focuses on a multiplatform framework.

This thesis's main framework is grounded on previous theories of political communication. The basics were developed by Harold Lasswell in his studies on propaganda effects [78]. He proposed a simple communication framework that starts with the source, then the message, followed by the channel, and ends in the receiver. For much of human history, the flow of communication was mostly a linear, top-down process from leaders to people [47]. Democratization of the political process allowed political involvement from other actors in the community. Technology has always been at the center of the changes in political communication, transforming the paths of interactions. From the press, the radio, the television, the Internet, and most recently social media, theories of political communication have been adopted to understand the complex processes the occur between actors. Dahlgren [79] presents the role of the Internet as a destabilization of political communication systems, where the destabilization can present positive effects as it extends and pluralizes the public sphere. Shah et al. [80] present a model that identifies two types of actors; those that encourage civic engagement and those that erode institutional legitimacy, foster distrust, and partisan divergence. Political communication models, irrespective of the technological channels, include opinion leaders as a central part of the ecosystem [81]. These leaders, including government officials, opposition, and pressure groups, use communication channels to convey their ideas and exert influence. The following subsection presents a general framework that puts social networks on center stage.

### 1.2.2 Main Framework

I introduce a framework that considers the main stakeholders who define how political communication occurs on social media. It is platform-independent and can also function for future social media applications. Figure 1.1 presents a visual guide to this framework (including emojis to keep up with the *zeitgeist* of the social media era). The main subjects in the framework appear as circles. **Political actors** can be politicians, governmental organizations, political parties, or institutions, such as unions. They represent the decision-making actors in a government, the opposition to the government, or a system that supports a group's interest. The **news media** include organizations and the journalists working for them. They play the role of gatekeepers of information, even though this role is thwarted in the social media era [82]. In the framework, I divide users into two categories: partisan users and passive users. **Partisan users** actively create political content, either by means of original posts, commenting on others' posts, or directly communicating with political actors. Partisan does not necessarily imply that the content supports one specific political actor, but that the users have a political ideology or ideologies that they share on social media. On the other hand, **passive users** do not actively create political content. However, they are still interested in political communication and interact with political content through reactions or shares. The normal definition of passive users refers to any user who follows and consumes content on a social network, without interacting with it. In the proposed framework, passive users do interact with political content and are interested in politics. For this reason, they are an important part of this political communication framework.

The fifth type of actor is part of the framework: the **bad-natured agents**. They encompass different types of entities that make the political conversation toxic. Toxicity includes false information, hate speech, racism, and sexism [83]. The accounts may be automated (trolls, bots, coordinated accounts) [84], fake personas controlled by real users [85], or real users. I decided not to separate them, as there is no consensus on the real taxonomy of fake accounts. It is hard to define the boundaries of these entities, as they are most commonly a mixture of automated sources and real persons. It is more important to measure these agents at the level of what they share in the network and the implications they may have for the political environment, irrespective of the type of agent. Bad-natured agents distort reality and undermine the principles of respectful argumentation. Especially difficult is categorizing users, political actors, or news media as completely bad-natured agents, as they may sometimes behave in a toxic manner and sometimes not. This means that the concept of bad-natured agents is fluid and depends on the conversation, topic, and social media "place" (forum, post, page, or similar venue). Extensive studies on bad-natured agents employ classifiers that divide accounts into two classes (social bots or not, coordinated accounts or not, trolls or not) by assigning strict thresholds that divide them [86, 87, 88]. This is acceptable for a computer science approach but does not depict the social reality that is almost never black or white. Rauchfleisch and Kaiser [89] discuss the problem of false positives in the study of bot detection. These facts are highlighted with the dashed lines in Figure 1.1 and the hexagon instead of the circle for the bad-natured agents.

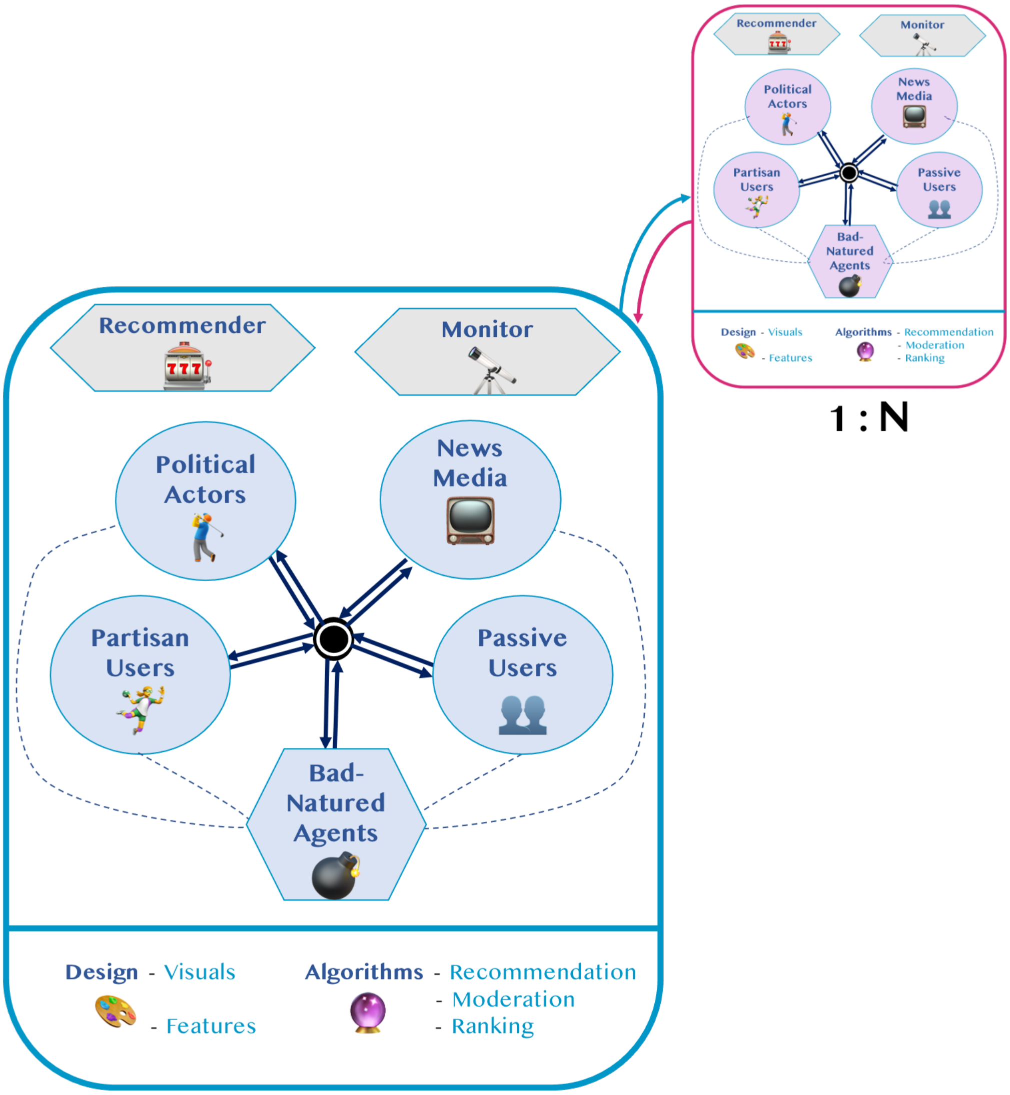


**Figure 1.1:** Framework of stakeholders in the political communication taking place on a social media platform.

Two other important agents have a high impact on the political communication that occurs on social media: the recommenders and the monitors. Whereas the previously discussed actors also apply to offline settings, these two are specific to the Internet era. Additionally, they are a combination of programmed algorithms and heuristics that each social media company introduces. The **recommender** is in charge of deciding which political content to present to a user, based on user interactions, preferences, and optimized algorithms. The algorithms contain parameters that engineers of the recommender systems can tweak. For this reason, recommenders are not only pieces of software, but also a product of people's decisions. In the extreme case, editorial decisions without the use of algorithms determine what to show to the user. For example, Facebook uses human intervention at almost every stage of its trending news headlines, similar to a traditional media organization [90]. Months before the U.S. election in 2020, Facebook even changed its recommendation system to lift news from authoritative outlets, over hyperpartisan sources, and rolled it back at the end of the year [91]. Apart from the engineers and decision-makers at social media companies, monetary offerings can influence the recommenders to push specific political views through **political advertising**. On the other hand, the **monitor** is the agent who oversees the political communication and can directly or indirectly affect it—directly, by influencing the recommender agents (e.g., if users are engaging more, decide to increase the political content; if polarized politics affect users, decide to decrease the amount of polarized news, probably to the detriment of user engagement); indirectly through advertising platforms that show user trends (e.g., users that like a specific type of music are right-leaning) and can be employed for microtargeting techniques (Section 1.3.1).

The arrows determine the flow of political communication between the stakeholders. The center dot in black is the abstract representation of political communication itself. On social media, the flow of communication follows a different broadcasting model than the classical one, where the news media had the main agenda-setting effect [92]. As Figure 1.1 shows, the communication path is two-sided for all actors. This is evident for political actors and partisan users who are creators of political content and have a high number of two-sided interactions on the platform. For news media accounts, they report politically relevant stories but also decide what to report, depending on what is popular. For example, trending topics can become agenda-setting mechanisms [93]. Passive users mainly receive political content. However, they also interact with it, by sharing it with their network or by simply reacting to it. Even passive users who consume political content impact the political communication flow. The more political content, the more it is expected to be recommended to them. At the same time, similar users will receive similar political content. Thus, passive users contribute to the flow of communication, albeit with a lower degree of influence. Social media provides the channels for different interests and opinions to be expressed, heard and counterposed [11]. These elements constitute the very essence of political communication.

Users perceive social media platforms as different entities and decide where to share their experiences, ideas, or other content. The distinction comes primarily from the **design** of each platform, what the user sees, and how the user interacts with it. Therefore, design affects the user experience on the platform and the way political communication can take place on it. I divide design into two main subcategories: visuals and features. **Visuals** relate to the user interface (UI), such as how to use screen space, the interaction mechanisms, the colors, and the fonts, specifically, the adaptation to different displays: mobile, web, and tablet [94]. A successful visual design is the first step toward having satisfied users who stay engaged with the platform [95]. For example, the color of a platform can directly influence how much time a user spends on it [96], or how much users interact with it [97]. The second step is implementing **features** that allow the users to interact with other users. These are related to the user experience (UX) and define how information flows in a social network. From Facebook shares and Twitter retweets to Twitter trending topics and TikTok duets, each feature affects political communication and provides measurements to identify relevance on the network.

Not only the actors in the network but also by the **algorithms** that decide what information is shown to a specific actor determine the flow of political communication. I select three main types of algorithms that influence political communication:

- **Recommendation:** This is the principal component of a social media platform, with the responsibility of either giving the user choices of what to consume next or automatically populating a content feed (naturally, directly connected to the recommender actor described before). The algorithms are based on the users' actions and the general dynamics of the information flow. Usually, they are trained to solve an optimization problem, set specifically to each social media platform, but most probably set to try to maximize user engagement [98]. The effects of recommendation systems have been studied before [11] and do not play a central part in this thesis.

- **Ranking:** Similar to the recommendations, these algorithms make decisions about what to show to a user. However, they differ in that ranking will also affect content that is not a recommendation to the user. For example, responses to a user post are not recommendations, but direct contact from other users. The responses can be ordered simply by time of posting or by a more complex ranking mechanism that decides which is more relevant. Ranking mechanisms play a larger role for content that receives a high number of interactions. Most of the users to whom this content appears will contextualize it, depending on the top responses the ranking algorithm selects.

- **Moderation:** Social media platforms depend on moderation algorithms to keep undesired content away from the platform or to advise its users of inappropriate content. The algorithms may be simple, from removing content containing swear words to more complex ones that detect hate speech or label false claims. What is moderated will directly influence political communication. An example is the moderation of Donald Trump's tweets after the 2020 U.S. election [99]. Gillespie [100] states that moderation shapes social media platforms as tools, as institutions, and as cultural phenomena, resounding more in the context of political content that may affect public opinion.

Figure 1.1 shows on the top right a second social media platform. Although its design, functionality, and algorithms may work differently, the political communication framework presented also holds for any of the existing social media platforms. The connections between them correspond to the information that flows between them. Nowadays, the possibility of sharing content between some platforms is a built-in feature. In case this functionality does not exist, users also share screenshots, or re-upload videos from other platforms. Additionally, a user can share a link found in one platform on another one. All these possibilities make the social media ecosystem complex, and political communication dynamic. Political content flows from fringe social media platforms to the dominant platforms and the other way around. Given that users, political actors, and even bad-natured agents can have profiles on more than one social media platform, defining and understanding how political communication takes place on social media is a complex task. For this reason, political communication must be studied simultaneously on multiple platforms.

## 1.3 Motivation

The first motivation behind the presented framework relies on identifying the actors that influence political communication online and their interactions. Given that social media play a significant role in the learning of political information within the modern media environment [101], it is important to identify how the structure has changed in relation to traditional mass media. According to Klinger [102], social media platforms operate with a distinctly different logic from that of traditional media, though overlapping with it. By providing a framework, this thesis tries to fill in the gap between research on classical political communication and countless analysis of political events on social media.

A different approach would be to focus on the content and treat the actors as users of a platform without differentiating their political roles. This would be correct, as every user has the same possibilities of online interaction and have the same tools at their disposal (apart from blue ticks that signify the account's authenticity of a person of public interest). However, this approach would fail to identify the intent that different actors have. Although the discussed topics may be similar, the user that creates a post on social media will have a varied impact depending on the role they represent. Moreover, avoiding an actor-centric approach would neglect the communication mechanics between different platforms. Political actors and campaign strategies depend on exploiting the various social media to reach different demographics and interest groups. The second motivation of this thesis is to show the importance of a multi-platform analysis. Studying political actors throughout the social media ecosystem allows understanding the strategies and communication channels that reach citizens.

This thesis presents to the reader a collection of papers that navigate the different political actors present in digital channels. The case studies undertake the analysis of different social media platforms, actors, and political events with a particular focus on US and German politics. This work represents an addition to the research of social media and politics. Its uniqueness relies on presenting a unifying framework of political communication that can be useful to study current and future digital platforms.

## 1.4 Selected Topics

This section presents a selection of topics from social media analysis that provides a better theoretical background for the papers I present in the next chapters. I selected them in terms of relevance, complexity, and impact on social media research. Important omissions include **political polarization**, **hyperactive users**, and **bias** in social media analysis. These topics are extensively discussed in two theses to which I contributed: [103] and [104] (Section 1.4).

### 1.4.1 Microtargeting

With the coming of the Big-Data era, increasing number of fields of human behavior can be categorized, quantified, and aggregated, especially through social media where large amounts of information are obtained from users, most of whom do not know the power that sharing their data gives to third parties. One of the most familiar usages of this data was in the U.S elections, by political parties who used microtargeting to influence voters [105].

Microtargeting is a strategic process in which detailed information is collected and processed to obtain results that will allow influencing voters. It tries to find out political preferences and form a precise description of each voter, so it is a personalized method. Microtargeting includes not only relevant data from the voter, such as names, addresses, and gender, but also abstract characteristics, e.g., social interactions, cultural interests and sociodemographic factors [7]. The term was first used by political consultant Alexander P. Gage.

The usage of microtargeting was first identified in the 2000 U.S. presidential elections, when the Republican party tried to use it to influence voters. Nevertheless, the full effects of microtargeting where seen in 2008, with Obama's campaign [106]. Since then, the use of microtargeting has been an essential part of U.S. elections. U.S. laws allow companies to store users' data and pass it to third parties for analysis. This is not the case in Germany, where the laws are very sensitive regarding data protection. But even in Germany, it is possible to use microtargeting in the elections [6]. Inferences can be made by tracking users who have interacted with different public pages. One example is the case of users who have interacted with more than one of the political-party pages. They are presumably the ones to convince to vote for one of those parties since they can appear as indecisive users interested in politics.

One of the main elements of microtargeting is *profiling*, which is the generation of user profiles generated by computer analysis of data. The company Cambridge Analytica used this method heavily to influence past political events. According to [107], microtargeting was mixed with psychological operations, normally used in military strategies on the civilian population. The study of microtargeting as a phenomenon has increased in the last few years. Endres [108] and Kruikemeier et al. [109] studied the impact of microtargeting on the creation of knowledge in individuals. Bodó et al. [110] discussed the abilities of microtargeting to influence individuals and its limits. Schipper et al. [111] investigated its efficiency in comparison to other information channels, with the help of simulations. Zarouali et al. [112] showed that users are more strongly persuaded by political ads that match their own personality traits. This result elevates the relevance of personalization in political campaigns.

### 1.4.2 Social Bots

Social bots are automated accounts that try to appear as real users, often with the specific goal of manipulating public opinion [113]. Social bots do not act alone and are often part of bot networks of thousands [114]. They amplify articles from low-credibility sources in the early spreading stages before an article goes viral [115]. Investigations of their activity during different political events, such as Brexit [116], the U.S. 2016 elections [117], a 2017 German state election [118], and many more, have occurred. Oxford's computational propaganda research project monitors the use of algorithms, automation and computational propaganda in politics across the world[2].

The main problem in tackling misinformation through automated accounts is that detecting social bots is a hard job. Their behavior varies and evolves over time to avoid detection. The developed algorithms to detect them are far from perfect and have varying accuracy, depending on the data available. The literature shows two main approaches to dealing with social bots: machine learning approaches [119] and heuristic approaches [120]. The Botometer tool [121] appears extensively in the literature, used for detecting bot accounts on Twitter. It is based on a machine learning model, trained on hundreds of thousands of English tweets. Recently, more advanced tools from deep learning have been applied in bot detection [122].

[2]https://comprop.oii.ox.ac.uk/

The direct effect of social bots on users has not yet been quantified. However, simulations show that they can have a profound impact on content popularity [123]. Even with the complete support of platforms like Twitter and Facebook, understanding the reach of social bots is a challenging task. For example, one of the most documented events of organized manipulation was the Internet Research Agency's (IRA) attempt to influence the 2016 U.S. election. Twitter provided the U.S. Congress with the data on the accounts involved. Since then, researchers have analyzed the data thoroughly [124, 125, 126, 127]. Although the results are helpful in understanding the IRA strategy, no real quantification of the effect of the misinformation campaign on the U.S. election exists.

### 1.4.3 Fake News Detection

Fake news is a loose term, used increasingly in the literature and in the media. Lazer et al. [128] define it as "fabricated information that mimics news media content in form but without the editorial norms and process to ensure the credibility of the information." There are different categories of fake news, for which different classification frameworks have been developed. For example, Rubin et al. [129] present three types of deceptive news: serious fabrications, large-scale hoaxes, and humorous fakes. A major problem with fake news is that it can deceive and transform the perceived reality of targeted groups. Moreover, it can spread online faster than the truth [130].

The spread of fake news has been in the limelight of traditional media since the aftermath of the 2016 U.S. election. According to several reports, fake news did not play a significant role in the 2016 U.S. election [131], nor in the 2017 German election [132]. However, they are part of misinformation campaigns whose goal is to create a different perspective on world events and to polarize the discourse. The techniques used often incite aggressive behavior that goes hand-in-hand with sensationalist reporting.

The task of detecting fake news is as loose as the definition of the concept. Most of the time, it revolves around a classification problem between "fake" or "not". However, most fake news presents some true information and misleads the reader only through some controversial claims. So, the task can also correspond to finding veracity on the claim level. Since users tend to only read the news headline, some fake news simply uses a fake headline that does not correspond to the content of the article. For this reason, some research studies means of detecting whether the title corresponds to the news article's text [133].

Automatic classification methods should be able to tackle disinformation on the World Wide Web [134]. As with any classifier, the features used to categorize an item define the accuracy of the model. Shu et al. [135] distinguish between news-content features that include linguistic and visual content from and social-context features based on user and publisher interactions. As with the bot detection problem, machine learning approaches are used to identify fake content.

Increasingly, deep learning techniques like CNNs, LSTMs, and Transformers (Section 2.1.5) are also being used for fake news detection [136, 137, 138, 139]. Most present work combines network analysis with deep learning approaches [140]. Zhou and Zafarani [141] present a complete overview of research on fake news.

### 1.4.4 Cross-Platform Social Media Analysis

The analysis of political communication on social media comprises three categories:

- **Single Platform Analysis**: The analysis centers on one social media platform and treats it as a single entity.
- **Multiplatform Analysis**: The analysis centers on more than one social media platform. It quantifies the interactions between actors for each platform separately. It treats social media platforms as separate entities.
- **Cross-Platform Analysis**: The analysis compares more than one social media platform and the communication units between them. The social media ecosystem allows users to share entries from one platform to another. In this way, content flows throughout the whole ecosystem.

These three categories are integrated with the framework presented in the last section. However, the title of this thesis refers to multiplatform analysis. The reason for this is that most of the research I present here considers either multiplatform or single platform analysis. I preferred to keep this thesis in the mulitplatform category as I do not focus on understanding the flow of information throughout social media channels (With the exception of [1], where I introduce a framework to compare interactions between platforms). Nevertheless, I have contributed to research regarding cross-platform analysis. In [13], we try to explain how content moderation on one platform affects the virality of content on other platforms. The methodology comes from previous research by Zannettou et al. [142]. They analyzed how mainstream and alternative news flows between Twitter, Reddit, and 4chan. Spangher et al. [125] analyzed the aforementioned IRA Russian strategy on Twitter, Facebook, and Bing. They focused on finding similarities between platforms, to identify which platform had taken a larger role in the disinformation campaign.

Cross-platform analysis enables understanding how moderation policies on some platforms affect misinformation on others. Information or discussion boards that start on 4chan or Reddit can move to Facebook and Twitter and then back to other fringe social media platforms. Ribeiro et al. [143] analyzed two communities that Reddit banned and that subsequently migrated to their own websites. New social media platforms emerged after content moderation practices arose on Twitter and Facebook—for example, Parlor, and Gab [144]. The social media ecosystem is changing rapidly, and without a thorough understanding of the flow of political content between platforms, there is no way to completely understand the ecosystem and act accordingly, to prevent the spread of misinformation.

## 1.5 Thesis Structure

This work is a collection of five peer-reviewed papers that try to explain the different aspects of political communication on social media. The main research question that this thesis seeks to answer is:

> **Research Question** *How do different actors (defined in Section 1.2) express their political ideas on multiple social media platforms?*

The approach to answering this question stems from investigating relevant case studies. In this case, relevance is confined to political and social events that happened between 2017 and 2020. The social media ecosystem transforms itself often, and new platforms start to overshadow old ones. With this thesis, I try to explore aspects of political communication that did not exist a couple of years ago and may not exist in the near future. For this reason, my goal is not to provide a generalized theory of political communication on social media. Doing so would require an ever-evolving framework that copes with a disruptive ecosystem. My approach is more practical as it provides selected examples of **how to perform multiplatform social media analysis**. These approaches are still useful even for new social media platforms in the future. The main platforms in this study are Facebook, Twitter, Instagram, YouTube, and TikTok.

For a complete contextual background, I feel obliged to present two other theses that are closely connected to this work. With the permission of both authors, I assert they can be considered the first two volumes of a trilogy. The three works complement each other, by portraying different aspects of online political ecosystems:

- Shahrezaye [103] presents the infrastructure to gather social media data also used in this thesis. He introduces a framework to continuously store the raw data on **scalable distributed databases**. Additionally, he studies polarization in social media and how to detect political orientations on networks of friends and followers. He approaches political communication from the big data side, by providing methodologies to detect political patterns in large-scale datasets.

- Papkyriakopoulos [104] presents a generalized theory of **political machines**. He defines a political machine as an extension of Wiener's social machine theory [145]. He provides an overview of machine learning and natural language processing techniques for the evaluation of political interactions. His main focus is on hyperactive users, recommendation systems, and bias in machine learning models. His approach to political communication is more theoretical in nature; he focuses on interpreting the human-computer-social media interaction as a complete cybernetic system.

As an extension of these two theses, my work provides a practical approach to studying political communication on multiple social media platforms. The focus of this work relies on five case studies to explain the dynamic transformation of political communication. The thesis is divided into the following four sections:

### A: Monitoring Political Communication

The first step to correctly understand how political communication occurs on social media is through data collection and monitoring. For researchers, accessing the data is not always possible. Twitter only provides a percentage of the total tweets at a given moment in time, and this sample has proved to be biased [146]. Facebook severely restricted access to platform data via its Application Programming Interface (API), in the aftermath of the Cambridge Analytica controversy [147]. Other social media, such as TikTok, do not even have official APIs. This makes monitoring political communication difficult. However, it is important to try to understand as much as possible of what happens on the platforms. In *The Political Dashboard: A Tool for Online Political Transparency*, I present a dashboard that monitors the German online political landscape. It shows live analysis of Twitter, Facebook, and online news media. The dashboard also covered analysis for the 2020 U.S. election. Dashboards are useful tools for researchers, journalists, and politicians to get an understanding of the current environment. They present a more unbiased view than navigating the social media platform itself, as such an experience is highly personalized.

### B: Political Actors

Political actors use social media as a major channel of communication to present their ideas to the public in a more direct manner, without such gatekeepers as journalists or news media. Some politicians are avid social media producers (e.g., Donald Trump), others stay mostly quiet (e.g., Angela Merkel). Political parties as entities also have their presence online to persuade and mobilize voters. As Section 1.2 discusses, there are two types of communication from political actors to users: organic content, and sponsored content. The spread of organic content depends on the structure of the follower network and the users that spread their content. On the other hand, the reach of sponsored content depends on the parties' financial investment. In the 2016 campaign, Hillary Clinton spent 38 million dollars on Facebook ads, and Donald Trump spent 44 million dollars [148]. By comparison, in 2020, Joe Biden spent 103 million dollars, and Donald Trump spent 85 million dollars on this platform [149]. Political content is booming on social media, from which political actors try to profit. At the same time, social media platforms should keep political actors accountable for the content they share. Researchers can help with auditing whether the platforms are following the norms and laws that regulate online political communication.

In *The Rise of Germany's AfD: A Social Media Analysis*, I investigate the extent to which Germany's far right-wing party, the Alternative für Deutschland, employed social media to attract attention in its first five years of existence. I also analyze their content and their reach on four social media platforms. At the same time, I compare their success with the rest of the German political parties, considering both federal and regional accounts. In this study, I only consider the organic content that the political parties created. By contrast, in *Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections*, I focus on studying the political ads that generated German political parties generated during the period leading up to the 2019 European election. They comprise ads on Facebook, Instagram, Google, and YouTube, the main platforms for political advertising. Other platforms, including Twitter and TikTok, prohibited political ads [150]. I could investigate sponsored content only because in 2019, Facebook and Google created ad archives that included the political ads that ran on their respective platforms. Although these archives do not provide the complete information on the advertisers' targeting schemes, they represent a step towards accountability and transparency.

### C: Partisan Users

Social media allows citizens to express their political opinions and share their ideologies. Some social media users are highly active in political discussions, whereas others are rather passive users who perceive social media as a source of information. Journalists and politicians are no longer the only sources of political content. Partisan users belong to the active user category. They create content that supports their party or their political ideas. Popular partisan users become *influencers* on a social media platform, therefore, play an important role in the political communication. Exchanges between partisan users of different political parties tend to permeate the political fabric of social media.

TikTok, a new social media platform, has taken the spotlight in terms of creating new forms of political communication. Its user base increased 75% in 2020 and is predicted to soon reach more than 1 billion monthly active users [151]. Based on short videos propelled by music trends, TikTok's algorithm rewards creativity more than popularity [152], which lowers the entry bar for new users with good ideas. In *Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok*, I study how users create and interact with political content on this platform. This study centers around U.S. politics, given the high amount of political content on the platform leading up to the U.S. 2020 election [153]. I investigate the patterns of communication and interaction between Republican and Democratic users. On TikTok, users are the central part of the political communication. Although some politicians are already on this platform [154], they stay at the outskirts of the conversation. This user-driven political ecosystem allows users to find new, creative paths to persuading others of their political ideologies.

## D: Bad-Natured Agents

The final section of the thesis deals with bad-natured actors, who behave against the code of conduct of a platform, share misinformation, or constantly spread toxic comments. As previously discussed in Section 1.2, these actors can be automated or not, political actors or partisan users. The distinction is not straight-forward, especially if users share ideas or links to stories that they believe are the truth because they trust a source (cable news, newspaper, blog, or other) that provides false information. Research on bad-natured agents increased after the 2016 U.S. election, but the consequences of these actors in the real world were hard to measure. However, the 2020 coronavirus pandemic originated an environment of misinformation on social media, a *misinfodemic* [155] that has had spillover effects on real events. Conspiracy theories, such as coronavirus being a product of 5G towers or the QAnon conspiracy [156], are at the center of social media discourse. In *NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube*, I propose an algorithm to detect YouTube videos containing conspiracy theories related to COVID-19. Instead of analyzing the video itself, the methodology looks at the comment section. It classifies comments as conspiratorial or not, using state-of-the-art natural language processing techniques. Videos with a high number of conspiratorial comments tend to include conspiracy theories in their content. In this way, I exploit the published content from bad-natured agents as input to an early-detection mechanism for identifying malicious videos.

As a complement to the papers in this thesis, I append two other first-author works that I wrote during work on the thesis, related to bad-natured agents. In Appendix A, I present excerpts from *Social Media Report: The 2017 German Federal Elections* [132], a published (non-peer-reviewed) report that analyzed social media activity in the months leading to the 2017 German election. The excerpts discuss the effects of social bots and trolls on Twitter. I explore the activity of the Russian trolls detected during the 2016 U.S. election [157] as they were also active during the months before the German election. In Appendix B, I include the unpublished article *Coordinated and Suspended Accounts* *on Twitter in the run-ups to General Elections.* This study investigates coordinated behavior and suspended accounts on Twitter during the fourteen days leading up to the general elections in the United Kingdom, Mexico, Germany, and Greece. I manually labeled the partisanship from a sample of both coordinated and suspended accounts to understand their political intent. In this way, I show how Twitter responded to coordinated accounts in different elections, and which political party these accounts supported.

## 1.6 Theoretical Contributions

Here I state the theoretical contributions of my thesis. Contributions measure the impact of the research and the key takeaways. The methodological counterpart appears at the end of the next chapter.

- I propose a multiplatform approach that unifies the different features of social media channels. This allows researchers to compare the reach of political accounts on multiple social media. The approach reflects four categories: party engagement, user engagement, user support, and message dissemination.
- I show the superior social media popularity of the AfD in comparison to the other German political parties.
- I demonstrate that as part of its social media strategy, the AfD avoided discussion of its economic proposals and instead focused on pushing its anti-immigration agenda to gain popularity.
- I provide the first analysis of political advertising in Germany and show that the German political parties are still not deploying large-scale microtargeting in their ad campaigns.
- I illustrate the shortcomings of Facebook's and Google's ad archives, and discuss the challenges for enhancing transparency in the online advertising ecosystem.
- I perform the first analysis of political communication on TikTok. I show that politics is an important part of the TikTok ecosystem in the United States. I find that the duet feature creates new forms of communication between partisan users, which I describe as a *communication tree.*
- I show that Democratic users on TikTok engaged significantly more in cross-partisan discussions, whereas Republican users preferred to duet with users who professed their same ideology to boost their message.

# 2 Methods

In this chapter, I discuss the methodological techniques that I employed for this thesis. I do not include an exhaustive list of statistical methods that can be applied to analyze online political communication (For a thorough presentation of methods, please refer to one of these books [158, 159, 160, 161]). I group the methods according to the different schools of statistical inference and statistical learning. I provide a short explanation of the different paradigms and the algorithms I employed from each of them. Although most statistical techniques tend to be labeled as artificial intelligence (AI) [162], it is crucial to understand their difference and how the interpretation of the results depends on the paradigm that the method belongs to. At the end of the chapter, I present the methodological contributions of this thesis.

## 2.1 The Different Paradigms of Data Science

Data science is a "concept to unify statistics, data analysis, and their related methods" to "understand and analyze actual phenomena" with data [163]. It combines classic statistical methods with computational methods from artificial intelligence and computer science fields. Its potential relies on the fact that data science can be applied to any field of research. However, expert knowledge in a given field is necessary for data science to be useful. Without this knowledge, the algorithms can provide good accuracy to a problem, but the interpretation would be missing. In politics, sociology, and other social sciences, a better term for applied data science is **computational social science** [164]. Social science research has always been based on the classic statistical methods to understand social phenomena. By including computational approaches, computational social science employs more advanced statistical methods that heavily rely on computational power. Given the plethora of methods available for statistical modeling, it is important to

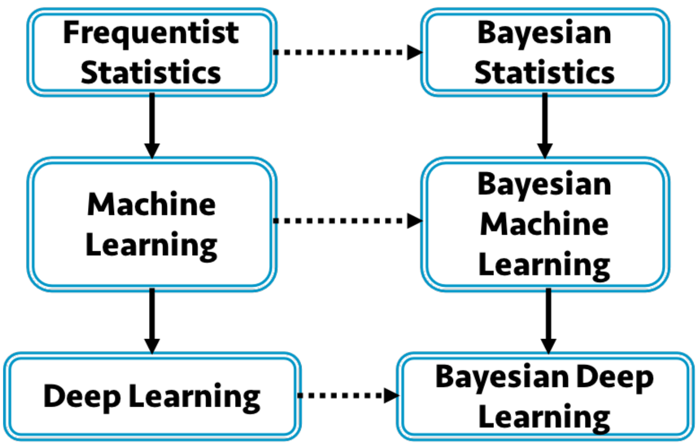


**Figure 2.1:** Different paradigms of data science and their connections.

understand the different paradigms of data science (or similarly computational social science) to decide which methodology to use to answer a research question. Figure 2.1 show a categorization of the different paradigms. On the top, the two schools of statistical inference appear: frequentist and Bayesian statistics. The dotted line shows that frequentist statistics can be expanded to a Bayesian setting. The next level is machine learning, which focuses more on learning from data and predicting. However, it needs a solid base on the statistical methods for inference. The deep learning paradigm appears at the bottom of the figure. Although it is a sub-field of machine learning, deep learning is considered separately given its popularity and the framework based on neural networks. The following subsections will provide a more detailed description of each of these paradigms. I exclude Bayesian deep learning as I did not employ it in my research. Its methods are still more on the experimental side and have not been applied yet to the social sciences. It may take more relevance in the future, given its potential for explainability and parameter uncertainty.

Table 2.1 shows which paradigms I employed for the published papers included in this thesis. Four papers include algorithms belonging to more than one paradigm. This is often the case in research as each method provides different perspectives to solve the research question. In the following subsections, I also show the methods that I employed in the research papers. For a detailed list of methods that each paper used, please refer to the paper itself.

| **Paper Short Title** | **FS** | **BS** | **ML** | **BML** | **DL** |
|---|---|---|---|---|---|
| The Rise of Germany's AfD | ■ | | | ■ | |
| The Political Dashboard | | | ■ | ■ | |
| Exploring Political Ad Libraries | ■ | | | | |
| Dancing to the Partisan Beat | ■ | | ■ | ■ | ■ |
| NLP-based Feature Extraction | ■ | ■ | ■ | | ■ |

**Table 2.1:** Published papers and the paradigms employed for the analysis (Abbreviated with their respective capitals)

Note: The following sections discuss different probability distributions. An explanation of each distribution is outside the scope of this thesis. Please refer to any classic statistics book to get a better understanding of them.

### 2.1.1 Frequentist Statistics

Frequentist statistics or classical statistics has been a staple of scientific research. The term frequentist derives from the fact that it relies on the concept of a sampling distribution, a distribution that an estimator has when applied to multiple data sets sampled from the true but unknown distribution. The frequency of the events determines the statistical parameters and their standard error. All the parameters are viewed as fixed and the data as random.

The frequentist school uses conditional distributions of data given specific hypotheses. A hypothesis plays the center role in designing a rigorous statistical analysis. First, a null hypothesis ($\mathbf{H_0}$) is defined, which presupposes the absence of a specific property or relationship. Then, a statistical test tries to quantify if the null hypothesis holds or not. For this, a test statistic is calculated by comparing a proposed probability distribution and the sample distribution. The selection of the test statistic and the probability distribution highly depends on the problem at hand and is selected if the dataset follows specific mathematical assumptions. If the test statistic exceeds a specific value (or equivalently, the **p-value** is smaller than a predefined threshold like 0.5), the null hypothesis is rejected and the alternative hypothesis ($\mathbf{H_1}$) holds. If this is not the case, the conclusion is that the data does not provide sufficient evidence to reject the null hypothesis but does not imply that it holds.

For statistical inference, the main principle of classic statistics is the **maximum likelihood estimation** (MLE). It allows estimating the parameters in a statistical model, which would make the observed data the most probable. The likelihood is defined as:

$$L(\theta|x) = f(x|\theta)$$

where $\theta$ represents the parameter or parameters and $f$ is the selected likelihood function to model the data. For independent and identically distributed random variables, $f$ will be the product of univariate density functions. It is important to notice that likelihoods are not normalized as probabilities.

Most of the statistical methods were conceived before the computational intensive era. Therefore, they use mainly mathematical approaches with assumptions to consider to calculate significance. However, there are also computational intensive methods in frequentist statistics. For example, the **bootstrap** is used to approximate the sampling distribution. It works by selecting a random sample of the data with replacement and calculating parameters. The average of all the samples is selected as the estimated parameter.

The statistical tests I present in the following subsections are still used extensively in scientific research. They are part of the non-parametric approaches employed when the observed data does not follow a normal distribution. This is the case for most of the data originating on social networks. Most complex social media interactions follow a log-normal distribution [165, 166]. Non-parametric tests mostly work on the principle of **ranking**. The lowest value in a dataset has rank 1, the next rank 2, and so on. If observations have the same value, they become tied ranks.

### Kolmogorov-Smirnov Test

This test is useful to identify if two probability distributions are statistically similar. It can measure the goodness of fit between the distribution of a data sample to a reference distribution. The test compares the empirical distribution function of the sample to the cumulative distribution function of the reference. The Kolmogorov-Smirnov test takes the longest distance between these two distributions as the test statistic and compares it to the Kolmogorov distribution. It is also possible to use this test to compare two sample distributions to identify if they originate from the same distribution. In this case, the null hypothesis is rejected at a level $\alpha$ if

$$D_{n,m} > \sqrt{-ln(\frac{\alpha}{2}) * \frac{1 + \frac{m}{n}}{2m}}$$

where n and m are the size of the two datasets. Figure 2.2 shows the two possible use cases for the Kolmogorov-Smirnov test.

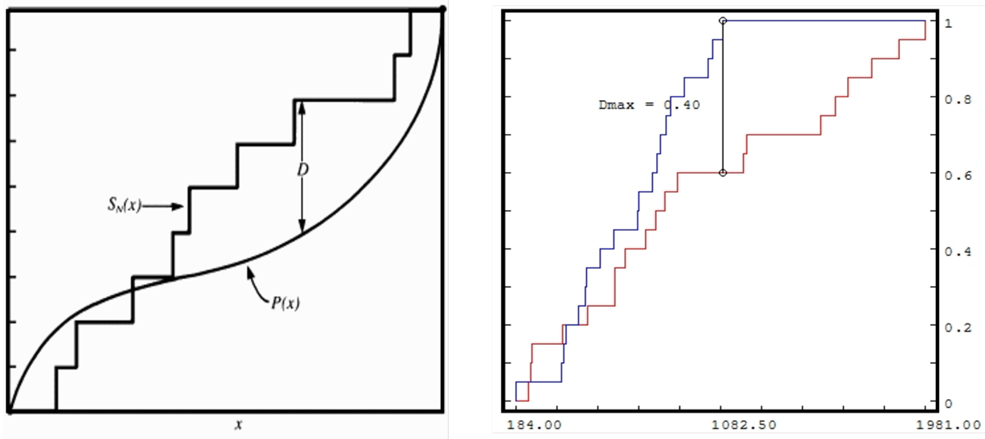


**Figure 2.2:** Left: Comparing the empirical distribution function of a sample distribution and a cumulative distribution function of a reference distribution. Right: Comparing two empirical distribution functions. Both figures show the maximum distance between distributions.[167].

### Wilcoxon Rank-Sum Test

This test is the non-parametric equivalent to the independent t-test, which compares the means between two independent groups (control and treatment). The core idea is that the groups are different if we rank the data from both groups together and then find if one of the groups has significantly higher ranks than the other. In this way, it compares the **medians** of the groups instead of the means. The test performs a simple **z-test** between the summation of the ranks of a group and the mean rank of the two groups.

This test is also known in the literature as the Mann-Whitney U test.

### Kruskal-Wallis Test

This test is the non-parametric version of the one-way ANOVA test, which compares the means between more than two independent groups and finds if they are statistically different. Similar to the Wilcoxon rank-sum test, the values of all groups are combined and ranked. It then calculates a test statistic with the ranking of the different groups. Finally, it uses the chi-squared distribution to compare with the test statistic. The null hypothesis is that all the groups have the same median. Rejection does not imply which of the groups are different. To uncover this, we perform pairwise Wilcoxon rank-sum tests between each of the groups. We add a correction for each of the pairwise tests to consider that we apply multiple tests to the same data. The Bonferroni correction, for example, takes the division of the predefined significance level $\alpha$ by the number of comparisons as the new criteria to reject or not the null hypothesis. We perform the pairwise tests only after the rejection of the null hypothesis from the Kruskal-Wallis tests.

### Chi-Squared Goodness of Fit Test

This test is a non-parametric test for categorical variables with counts. This means different categories where each category has several occurrences in the dataset. It compares the different proportions in the categories with a null hypothesis that entails that the proportions (counts in a category divided by total counts) follow a predetermined pattern. For example, it can test whether the proportion of all categories is the same. The test statistic for $C$ categories is:

$$\chi^2 = \sum_{c=1}^{C} \frac{(O_c - E_c)^2}{E_c}$$

where $O_c$ and $E_c$ are the observed and expected counts in category $c$. As the name implies, the test uses the chi-squared distribution with $C - 1$ degrees of freedom to compare with the test statistic. The null hypothesis can test for a specific pattern of proportions other than equality between categories.

### 2.1.2 Machine Learning

Machine learning is a subfield of computer science based on many of the fundamentals of classic statistics. Instead of focusing on inference, machine learning relies more on predictive algorithms that learn from the data. Accuracy, optimization, and data mining approaches play a larger role than interpretability and closely following statistical assumptions. In other words, it is a more data-driven discipline, and without considering the Bayesian paradigm, it is sometimes better labeled as statistical learning. Whereas inference allows us to test hypotheses, learning is interested in making predictions from future data.

There are three main types of learning: supervised learning (uses labeled data), unsupervised learning (uses unlabeled data), and reinforcement learning (consists of an agent with actions, states, and rewards). Supervised learning further subdivides into two main categories: regression, where the target to predict is a real-valued label, and classification, where the target is discrete, often belonging to a set of categories. All model-based learning algorithms try to minimize the objective known as the cost function, which consists of a term that averages differences between the real target labels and the predicted ones. There are different ways of representing these differences according to the specific method. The differences are commonly referred to as loss functions. In the following subsections, I present three supervised classification models that I used in this thesis.

#### Logistic Regression

Even though it has the term regression in its name, logistic regression is a classification method. It takes its name from the fact that its mathematical formulation is similar to linear regression. For two classes, it tries to calculate the probability of a data sample x belonging to class 1: $P(y = 1|x)$. This is called binary classification. The **sigmoid** function is applied to x to model this probability:

$$P(y = 1|x) = \sigma_{\mathbf{w},b}(\mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}}$$

The sigmoid transforms any input to be in the $[0, 1]$ interval to interpret it as a probability. The argument of the exponential function is equivalent to the linear regression function, where the weights multiply with x together with a bias term. The weights and the bias are the parameters that logistic regression should learn. A threshold must be defined to classify the data. For example, all values below 0.5 belong to class 0 and higher to 0.5 to class 1. Figure 2.3 shows the difference between linear and logistic regression. The importance of using the sigmoid function should become more evident with the image.

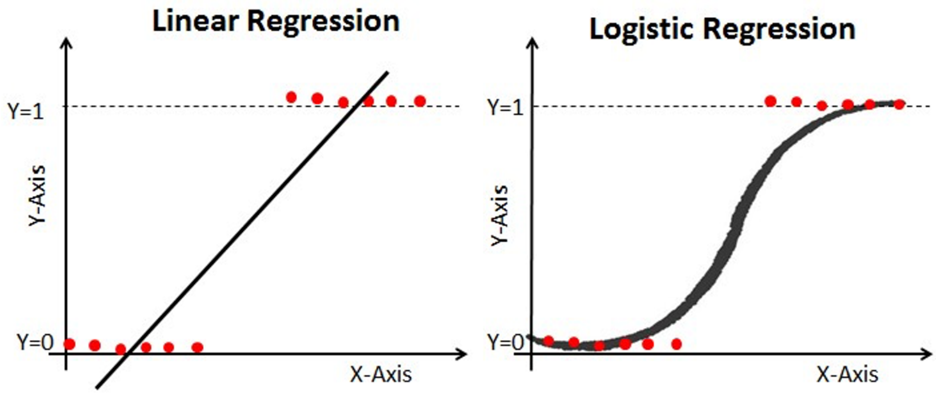


**Figure 2.3:** Linear regression vs. logistic regression [168].

The optimization criterion in logistic regression is the maximum likelihood, which tries to maximize the model's likelihood. As is the case of most machine learning algorithms, the logarithm of the likelihood is easier to work with:

$$LogL_{\mathbf{w},\mathbf{b}} = \sum_{i=1}^{N} y_i ln(\sigma_{\mathbf{w},b}(x_i)) + (1-y_i) ln(1-\sigma_{\mathbf{w},b}(x_i))$$

There is no closed analytical solution to optimizing this quantity. A numerical optimization that is commonly employed to solve it is **gradient descent** [169].

It is straightforward to extend this algorithm to more than two categories (multiclass classification). Instead of the sigmoid, the **softmax** function is used:

$$P(y=c|\mathbf{x},\mathbf{W}) = softmax_{\mathbf{W}}(\mathbf{x}) = \frac{exp(\mathbf{w}_c^T \mathbf{x})}{\sum_{c'=1}^{C} exp(\mathbf{w}_{c'}^T \mathbf{x})}$$

where $\mathbf{w}_c$ is the c'th column of $\mathbf{W}$, and $c$ represents a class. The bias term still exists, but is included in the $w$ vector for notation clarity. The log-likelihood is simply generalized as:

$$LogL_{\mathbf{W}} = \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} ln(softmax_{\mathbf{w}_c}(x_i))$$

and is known as the cross entropy loss.

### Support Vector Machine

A support vector machine (SVM) [170] tries to find a good decision boundary between data from different classes. In a high-dimensional space, the boundary consists of a hyperplane that separates positive examples from negative ones. A good decision boundary should have the largest margin between classes. The margin is defined as the distance between the closest examples of the two classes to the decision boundary. The equation of the hyperplane is simply the linear regression formulation again:

$$\mathbf{w}^T\mathbf{x} - b = 0$$

The predicted label depends on the sign of the right-hand side of the equation. SVM requires that the positive label has the numeric value of +1, and the negative label has the value of −1. The constraint for a correct classification is:

$$y_i(\mathbf{w}^T\mathbf{x} + b) \geq 1$$

The one is selected to make the margin to be of size $\frac{2}{|\mathbf{w}|}$ (the denominator is the norm of $w$). This is a result of a geometrical interpretation of the distance between a point a line. The data points that lie in the margin are called support vectors. The graphical illustration of this appears in Figure 2.4.

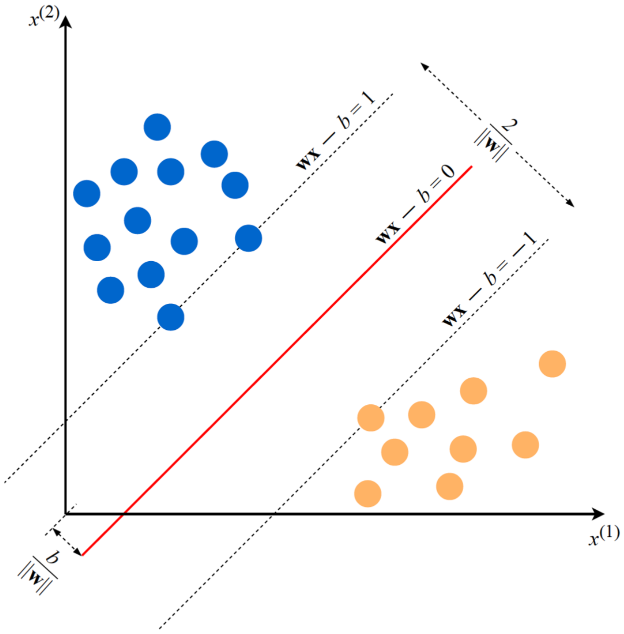


**Figure 2.4:** A SVM model for two-dimensional feature vectors [171].

The problem consists of optimizing the norm of $w$ subject to the previously defined constraint to find the maximum margin. This is a constrained convex optimization problem, which requires knowledge in Lagrangian multipliers and algorithms such as the sequential minimal optimization [172]. This implementation of the SVM is better known as linear-SVM. The method can extend to non-linear boundaries by incorporating **kernels**. Kernels are outstanding mathematical calculations that allow to calculating inner products in high-dimensional (even infinite) dimensions without explicitly transforming the data.

### Random Forests

Random forests is an ensemble algorithm that consists of simpler classifiers called **decision trees**. A decision tree is an acyclic graph with branches that represents decisions.

Each branch node compares a feature of $\mathbf{x}$ to a given threshold. If the value is lower than the threshold, a left branch follows, otherwise a right branch. The tree can be followed from top to bottom until reaching a leaf node to make a prediction. The predicted probability $P(y = c|x \ in \ leaf \ node) = \hat{\pi}_c$ is easily calculated by dividing the number of training samples that fall into a leaf node with class c by the total number of samples in the given node. A decision tree considers for each graph node which feature and threshold maximize the reduction of the cost function. The cost function consists of either the entropy (H) or the Gini impurity (G) of the two lead nodes, weighted by the number of samples in each node:

$$C = \frac{|S_{left}|}{|S|} H(S_{left}) + \frac{|S_{right}|}{|S|} H(S_{right})$$

where $S$ is the set of training data and $H$ can be replaced by $G$. Entropy and Gini impurity are defined as:

$$H(\hat{\pi}_c) = \sum_{c=1}^{C} \hat{\pi}_c ln(\hat{\pi}_c)$$

$$G(\hat{\pi}_c) = \sum_{c=1}^{C} \hat{\pi}_c (1 - \hat{\pi}_c)$$

Additionally, the algorithm needs a set of stopping conditions. The most logical one is if a leaf node contains only samples from one class. Decision trees are non-parametric models, as the number of parameters are not fixed at the beginning of training. To avoid the overfitting of training data, a pruning mechanism can remove branches that don't contribute significantly to the error reduction.

A decision tree model has high interpretability but has less predictive power than other machine learning classifiers as they are high variance estimators. An ensemble of trees reduces the variance and creates a more stable model. A random forest model combines the prediction of a large number of decision trees via bagging. Bagging (short for bootstrap aggregation) is a sampling mechanism that takes several random samples *with replacement* from the training data and trains a tree for each one. Moreover, each decision tree considers only a random sample of the input features at split time. The trees will be different from each other due to both random sampling approaches. After training N number of trees, the average prediction is the ensemble model prediction. Figure 2.5 compares a single decision tree versus an ensemble of trees. The random forests algorithm also works for regression tasks by considering a cost function similar to the mean squared loss.

### 2.1.3 Bayesian Statistics

Bayesian statistics has a different philosophy to interpret data than frequentist statistics. It considers data as being fixed and the parameters as being variable. In this way, it tries to identify the probability distribution of the parameters and not only point estimates. Bayesian statistics rely on the Bayes theorem to achieve this:

$$P(\theta|X) = \frac{P(X|\theta) * P(\theta)}{P(X)}$$

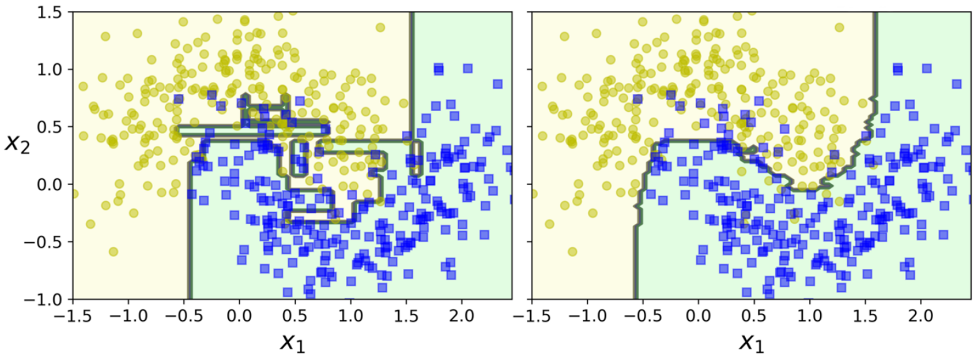


**Figure 2.5:** A single decision tree (left) vs. a random forest model (right) [173].

where $X$ represents the data, and $\theta$ the parameters of the model. The term $P(X|\theta)$ corresponds to the likelihood from frequentist statistics. The main difference relies on the $P(\theta)$, which is referred to as the prior distribution. The prior convey some previous knowledge of the model and is one of the tricky parts of Bayesian statistics. With a large number of data, the prior will not play a large role as with small datasets. If the prior is the uniform distribution (no previous knowledge of the model), the Bayesian setting transforms into the frequentist setting. The distribution of the parameters is referred to as posterior distribution. The confidence intervals can be obtained directly from this distribution and do convey the uncertainty of the parameters. In contrast, the confidence intervals in frequentist statistics can convey the same but often do not, as they are based on the data and not on any parameter distribution [174].

Bayesian statistics do not rely on the classic hypothesis testing with null hypotheses, alternative hypotheses, and p-values. Although a Bayesian approach allows p-values, called Bayes factors, they do not play a decisive role. The prior can be interpreted as the probability that the hypothesis is true before the data is observed. The likelihood is the evidence about the hypothesis given the data, and the denominator is the total probability of the data taking into account all possible hypotheses.

Bayesian statistics is more computationally intensive than frequentist statistics because the denominator of the Bayes formula $P(X)$ has to be approximated for most of the cases. There are few analytical Bayesian solutions (for example, Bayesian linear regression) given that this probability needs to integrate over many parameters. To approximate the posterior distributions, Bayesian statistics rely on sampling techniques, mainly Markov chain Monte Carlo (MCMC). A thorough analysis of Bayesian statistics can be found in Gelman's book [175].

### Bayesian Logistic Regression

The goal of this method is the same as for logistic regression explained in the machine learning section. However, the Bayesian setting is slightly different:

$$\theta = sigmoid(\mathbf{w}^T\mathbf{x} + b)$$

$$y \sim Bernoulli(\theta)$$

$$b \sim Normal(\mu_b, \sigma_b)$$

$$w_i \sim Normal(\mu_{w(i)}, \sigma_{w(i)})$$

The sigmoid still defines the main structure of the logistic regression. The prediction is modeled as a Bernoulli trial. In the non-Bayesian setting, this interpretation is similar. However, what changes is that the parameters are modeled as normal distributions. The prior distributions of the model relate to the mean and standard deviations of the normal distributions. The prior for the mean is modeled as a normal distribution, and the prior for the standard deviation is modeled as a Half-Cauchy distribution. The extension to multiclass classification changes the sigmoid to a softmax function and the Bernoulli to a Categorical distribution. Figure 2.6 shows this Bayesian model represented as a Kruschke diagram.

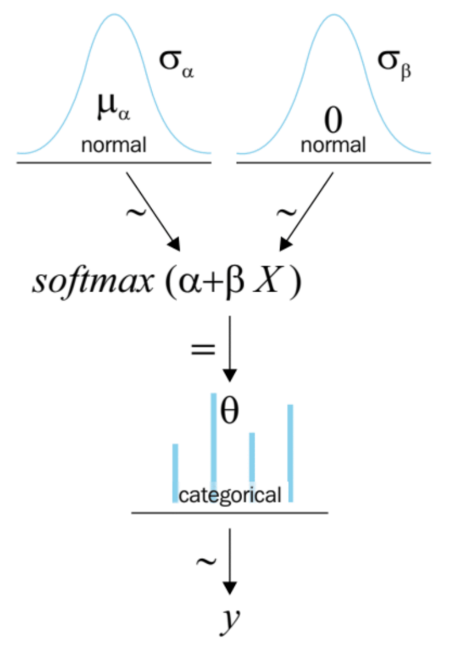


**Figure 2.6:** Kruschke diagram of multiclass logistic regression [176]. Small notation changes: $w$ to $\beta$ and $b$ to $\alpha$

An attentive reader would ask why did I include logistic regression in the machine learning section and not in the frequentist statistics section, or similarly, why did I not include Bayesian regression in the Bayesian machine learning part (next section)? This illustrates the difficulty in separating methods in the different paradigms of data science. Especially the most popular ones that exist in all paradigms, such as linear and logistic regression. Each paradigm treats the same method in different manners, especially the interpretation part. The logistic regression in machine learning is mostly used for learning, whereas the logistic regression in classic statistics is used for inference, where the weights of the model play the centric role, not the accuracy. In this thesis, I employed Bayesian logistic regression for inference and interpretability and not to calculate posterior predictive distributions. Posterior predictive distributions consist of integrating over all parameters to make a prediction and have a larger focus in Bayesian machine learning.

### 2.1.4 Bayesian Machine Learning

Similar to Bayesian statistics, Bayesian machine learning uses Bayes rules and prior probabilities to calculate posterior probabilities of the variables under consideration. The main set of models that comprise Bayesian machine learning are the **Probabilistic Graphical Models** (PGMs). They encode complex joint multivariate probability distributions using graphs. A node in the graph corresponds to a random variable, and the edges correspond to conditional independence relationships between variables. The graph structure allows solving inference (by computing marginal probabilities) and learning tasks (by estimating parameters of the probability functions) easier than by considering the full joint distribution probability of all variables. There are two main flavors of PGMs: directional graphs, better known as Bayesian Networks, and undirectional graphs, called Markov Random Fields. Figure 2.7 shows an example of a Bayesian Network where the random variables are the nodes and the joint probability distribution is factorized to be composed of simpler conditional probabilities. Observed variables are depicted as colored nodes, and non-observable variables that (may) have some influence on the observed variables are depicted as uncolored nodes. These variables are referred to as **latent** variables.

It is important to understand which variables are independent of each other to work with graphical models. With simple diagrams this is trivial, but with highly complex networks, more strategic algorithms are needed, for example, d-separation [177]. In Figure 2.7, the probability of having congestion and the probability of muscle-pain are independent if we know that the person has flu or not. If we do not know this fact, both probabilities have a dependency.

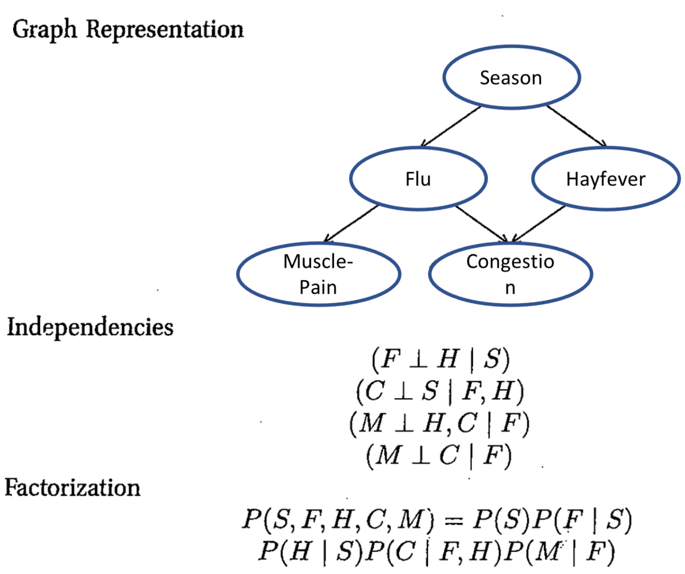


**Figure 2.7:** Schematic of a Bayesian Network, with the independencies and a factorized joint probability [178].

Some graphical models can be solved analytically but require long mathematical calculations. Other models with an intractable analytical solution require variational inference or sampling strategies to be solved. Koller et al. [178] presents a complete treatment on probabilistic graphical models. One of the most popular Bayesian machine learning algorithms in computational social science is latent Dirichlet allocation.

### Latent Dirichlet Allocation

The goal of this method is to assign a distribution of K predefined topics to the $M$ documents in a text corpus $T$. Each document $D$ consists of a collection of N words. A word $w$ is represented as a one-hot vector of size V, the vocabulary size. Latent Dirichlet allocation (LDA) [179] is a generative probabilistic model that assigns high probability both to the documents and to other similar documents. The assumptions of the model are that documents with similar topics will have similar words, that documents are probability distributions over latent topics, and that topics are probability distributions over words. The generative process is as follows:

1. Choose $\theta \sim Dir(\alpha)$
2. For a predefined number of N words $w_n$:
   a) Choose a topic $z_n \sim Multinomial(\theta)$
   b) Choose a word $w_n|z_n \sim Multinomial(\beta)$

   where $\theta$ is a K-1 Dirichlet random variable, and $\alpha$ is the parameter that controls its distribution, working as a prior. The Dirichlet can be seen as a set of probabilities that sum up to one, and $\alpha$ as a concentration parameter of the probability mass. With a value much less than 1, the mass is highly concentrated in few components, whereas with a value greater than 1, the mass will be dispersed almost equally among all the components. The $\beta$ parameter represents the prior on the per-topic word distribution. The model can be better appreciated as a graphical model (see Figure 2.8). The rectangles are *plates* that represent 1 to N words and 1 to M documents. Only the words are observable and this is represented with a filled node.

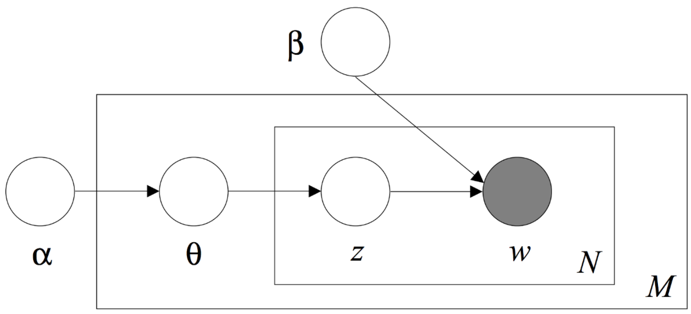


**Figure 2.8:** Plate notation of the LDA model [179]

The factorized joint distribution is expressed as:

$$p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

In other words, LDA is a method that uses a bag of words approach with a three-level hierarchical clustering model. An iterative approach similar to k-means clustering is applied to solve it. This iterative process is based on variational inference to determine the posterior distribution of the words and topics. The most common solution approach is through collapsed Gibbs sampling [180]. There are several methods to optimize the number of K topics similar to the methods that identify the optimal number of clusters in clustering algorithms.

### 2.1.5 Deep Learning

Deep learning is a branch of machine learning that uses neural networks as its core component. Neural networks are universal function approximators as they can learn an approximation of any function *f()*. They are composed of nested functions applied to a given input. Figure 2.9 shows a schematic of a deep neural network, in this case, a feed-forward neural network (FNN). Each inner layer represents a linear function $Wx$, where $W$ is a matrix that multiplies with the previous input $x$ (Similarly $\sum w_i * x_i$ in non-matrix notation). The edges between nodes in layers represent the weights in the $W$ matrix. Additional to the linear transformation, each layer can include a non-linear function. This non-linearity (also known as activation) is almost always applied since it helps the network learn complex non-linear functions. The most famous non-linear function is the ReLU, which converts every negative number to zero and leaves the positive values untouched. Even though it is a simple non-linearity, it has proven to be successful. A three-layer neural network with input $x$ and ReLU activation would be equal to the function:

$$f = W_3 * max(0, W_2 * max(0, W_1 * x))$$

The input layer consists of nodes that represent the features (numerical) input data. A simple neural network with only one hidden layer and a sigmoid (or softmax) activation function is equivalent to the linear regression equation for binary classification (or multiple classification). It represents a sum of the features multiplied by the weights and then the activation function. For a binary classification problem, one output node is used to represent the probability of belonging to class one. Multiple classification needs N output nodes. The function will be identical to a linear regression model in the absence of an activation function. Feature selection in deep learning plays a smaller role than in machine learning. The common practice is to let the model overfit the data and then select the point in time before the overfit started (early stopping). The interpretability factor is not inherent in deep neural networks. This has the highest contrast with frequentist statistics which requires assumptions to be met to accept a model to be valid. In deep learning, having more features is better, as accuracy is commonly the final goal.

Similar to most machine learning algorithms, a loss function has to be defined to train the neural network. In the case of classification tasks, most of the time the cross-entropy loss is used. The method to train a neural network is called backpropagation. It consists of calculating the function values (forward pass) and then calculate gradients of each neuron with respect to the weights from the deepest layer and backpropagate the gradients to the upper layers (backward pass). The weights are updated according to the gradient descent mechanism that tries to find the lowest loss by subtracting the gradient times a learning rate parameter. Given that the number of parameters is high and the input data normally consists of thousands of training examples, it is not advisable to calculate the gradients with respect to all the input data. Stochastic gradient descent

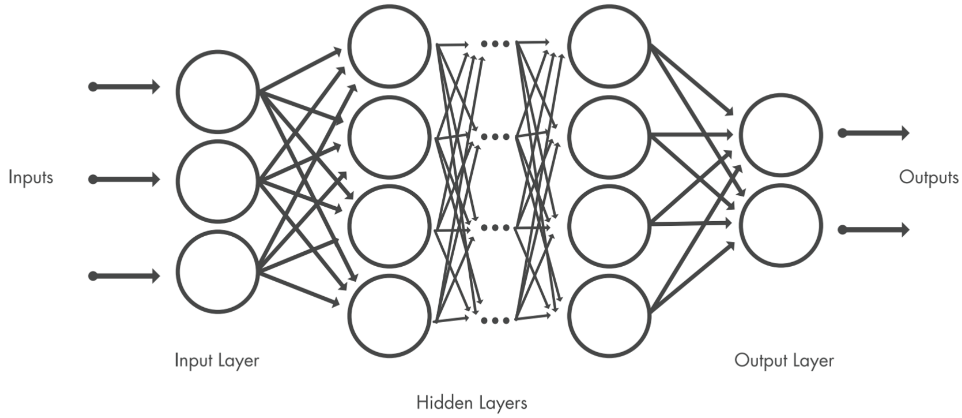


**Figure 2.9:** Schematic of a deep feed forward neural network (FNN) [181].

allows to take non-repeating mini-batches of the data and update the gradients only with the gradients of the mini-batch. A complete pass of weight updates with the complete data constitutes one epoch. For thorough explanations on deep learning, please refer to Goodfellow et al. [182]. In the following subsections, I describe state of the art neural architectures for visual and textual data.

### Convolutional Neural Networks

The most popular neural architecture for dealing with visual data (images or video) is the convolutional neural network (CNN). CNNs rely on discrete convolutions between images and filters. A 1-dimensional convolution is described as:

$$(f * g)[n] = \sum_{m=-M}^{M} f[n-m]g[m]$$

where f is the input and g is the filter. This convolution is an element-wise multiplication of the filter at continuous positions of the input. A 2-d convolution would require a 2-d filter that covers the different parts of the input. Figure 2.10 shows an image of binary pixels and a filter (numbers in red) of size 3x3 applied to the image's first nine elements. The element-wise multiplication is then summed, and a new element (in the case of the image a 4) is extracted. The filter would then move to the right and later downwards to cover the uncovered pixels. In the end, the 5x5 image reduces to a 3x3 feature.

A deep CNN architecture consists of several convolutional layers. A convolutional layer consists of many filters of the same size that allow extracting many features simultaneously. Although the filters decrease the dimensionality of the original image, concatenating many filters increases the depth dimension. Figure 2.11 shows a typical deep CNN architecture for classification. On the left, the image consists of a 3xNxN matrix where each element is a pixel value between 0 and 255. The number three is related to the three visual channels (red, blue, green). The first convolutional layer transforms the three channels into a K dimensional block (with K number of filters) each of a reduced size depending on the filter size. The filters are not fixed during training, they are learned through backpropagation. The goal is that each filter learns to extract meaningful features. After the convolution, a RELU activation is applied. Before passing to the

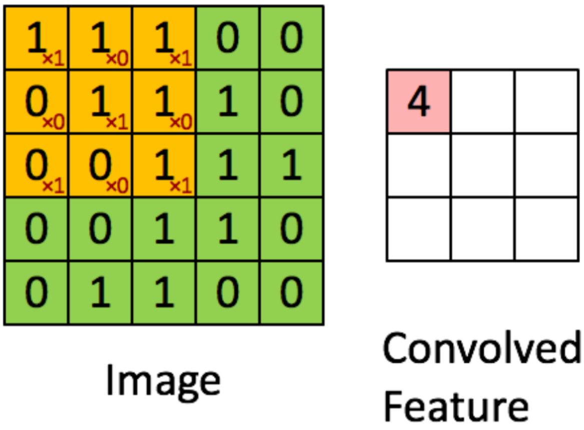


**Figure 2.10:** A dummy image, a filter, and the result of the convolution between them [183].

next convolutional block, there is often a **pooling** layer. This layer is similar to a filter given that it visits continuous parts of the image. However, its purpose is to calculate either the maximum or the average of a given patch. In this way, it is not a convolution but a simple reduction mechanism. Pooling is useful for extracting dominant features and to reduce the spatial size of the convolved feature.

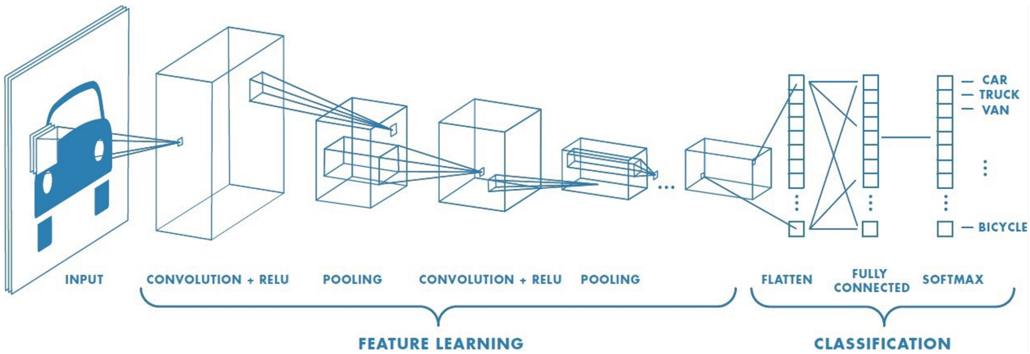


**Figure 2.11:** A typical deep CNN architecture for classification [181].

The stack of convolutional blocks represents the feature learning part of the architecture. At the other end of the architecture, there is a fully connected (also called dense) layer that takes the flattened output (transformed from N-d to 1-d) and uses a softmax function to transform every output into a probability between zero and one. The softmax function acts as a simple logistic regression that selects the category with the highest probability. Even in complex architectures, logistic regression is a staple of classification tasks.

The major progress for CNNs came with **transfer learning**. This technique fine-tunes pre-trained large models to work for another application with substantially fewer data. Transfer learning works by exchanging the last fully connected layer of a CNN architecture and connecting it to a dense layer suited for a different task. All the parameters from previous layers can be fine-tuned. However, it is also possible to freeze some layers (commonly the lowest ones) and only propagate the gradient to the unfrozen layers. Transfer learning represented a schism in the deep learning world. It has allowed developers to create highly accurate applications with less data than usual and tailored tasks with pre-trained knowledge.

### Transformers

Transformers are a family of neural architectures based on the Transformer model presented by Vaswani et al. [184]. Since 2018, they have proven superior to recurrent neural network (RNN) architectures in processing textual data. Figure 2.12 shows the main blocks that constitute the Transformer. It consists of two main blocks: the Encoder and the Decoder. The Encoder block (right) takes as input the word embeddings of a sentence (or sentences) summed with a positional encoding.

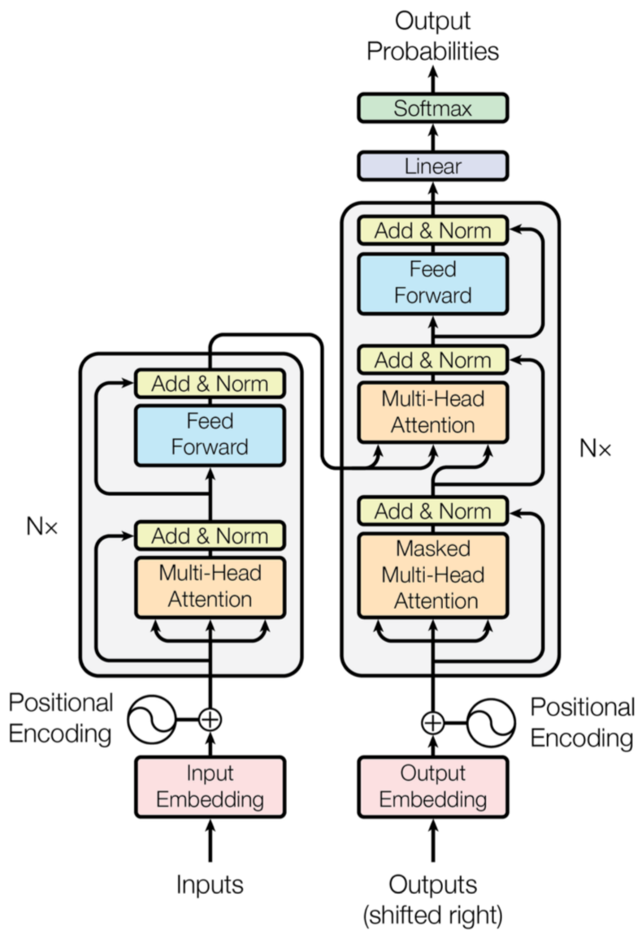


**Figure 2.12:** Transformer architecture [184].

- **Word Embeddings** are numerical vectors that represent a word (or a character) [12]. The mapping from string to an n-dimensional vector can be learned during the training process of the neural architecture. Words with similar meanings should have similar vectors, and directions in the word vector space should have a semantic meaning.

- **Positional Encodings** They are fixed vectors of the same size as the word embeddings. Their purpose is to tell the model the position of each word in the sentence(s). For example, they can be simply ordinal numbers from one to the number of words in the input. However, the size of the input will vary with different examples and this makes this simple encoding prone to errors. The original paper selects complex sine and cosine functions to create the encoding. This allows the model to understand that the text is sequential and that structures often repeat themselves (e.g., noun, verb, noun, punctuation, noun, verb, noun, punctuation).

The Encoder block consists of a multi-head attention layer and a feed-forward neural network. The multi-head attention creates inner products between linear transformations of the input embeddings to calculate which words in a sentence relate to each other. It creates many self-attention blocks, each learning different syntactic and semantic features of a sentence. Attention is the mechanism that drives the Transformer to understand the relationship between words. For a thorough explanation of the multi-head attention refer to this tutorial[1]. The feed-forward neural network consists of a dense layer of higher dimensionality than the input embedding and a second dense layer that returns the vector to the original embedding. Both layers have a normalization layer on top with the input embedding being summed to the output embedding. This direct connection between input and output relies on residual connections that allow training very deep neural architectures.

While the Encoder encodes the important information of the input sentence, the Decoder tries to decode this information by transforming it into a new representation. In the original Transformer, the Decoder learns to translate a sentence from one language to another. The input to the Decoder at training time is the translated sentence. The Decoder block also has a multi-head self-attention layer as its first component. However, for each word, the attention can only occur between itself and the preceding words. This is needed, given that at inference time, there is only the possibility to generate a translated word one step at a time. However, during training, the model will always know the correct translation (called teacher forcing) and will apply all calculations in parallel. After the first self-attention block, there is another multi-head attention layer. However, this is not a self-attention block but an attention block between the Encoder's output and the Decoder's input. Here the Decoder can use the complete information from the original sentence.

Both the Encoder and Decoder can consist of one to N number of layers. The input of the first Encoder layer is the word embeddings with positional encodings. The subsequent layers take as input the output of the previous layer. Each Decoder block obtains the information from the last Encoder layer. The last layer of the architecture consists of a Dense layer with a softmax activation function. The classification task consists of selecting the word from the output language vocabulary that represents the translated word.

The original Transformer architecture was applied mainly to the task of machine translation. The breakthrough occurred after taking the same building blocks and apply them to other NLP tasks. **BERT** (Bidirectional Encoder Representations from Transformers) [185] takes the Encoder part to train a language model. Language modeling is the task of predicting the next word given a sequence of previous words. A good language model

[1] http://jalammar.github.io/illustrated-transformer/

takes a text corpus and is good at understanding common sequences of words and helpful in the language generation task. Normal language models are autoregressive models. They take the previous input (the past) and predict the next words (the future) and incur a loss if the predicted word is not the original from the text corpus. BERT is a bidirectional model given that it changes the training task to allow using the past and future words to predict a word at a specific position. For this, it trains the Encoder with the masked language model, which consists in *masking* 15% of the words in the text corpus and try to predict them, similar to a classification task. Figure 2.13 shows this task with the masked word embeddings and a simple positional encoding. The power of BERT and similar methods rely on transfer learning. First, the language model is trained on an extremely large dataset. Then, BERT is fine-tuned by exchanging the last layer of the architecture to train on a specific task, commonly on a significantly smaller dataset.

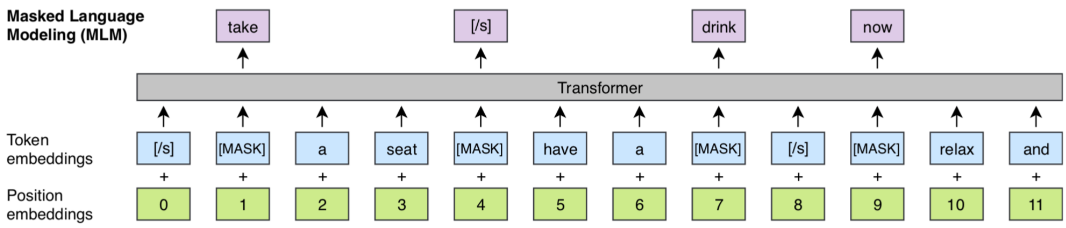


**Figure 2.13:** BERT's masked language modeling task [184].

A variety of BERT-based models have emerged in the last years. For example, RoBERTa [88] takes the same architecture but trains the language model longer, with more data, more parameters, and longer sequences. The newest language models continuously scale the number of trainable parameters and computations. Additionally, another family of Transformer models takes the Decoder block to train a language model. They are autoregressive models given that the masked self-attention layer of the Decoder does not allow to learn from future words. The most famous model with this architecture is GPT [186]. As of the time of writing, Transformers are starting to become ubiquitous for every NLP task. They are even applied to computer vision tasks and may prove superior to the CNN architectures in the near future [187].

## 2.2 Methodological Contributions

- I propose a simple method to compare the social media content from a political account and the manifesto, or official message, of a party with the help of LDA.
- I introduce a method to measure the success of boosted Facebook posts even without the real ads' click-through rates. I compare the engagement between organic posts without any sponsorship and boosted organic posts using Kruskal-Wallis tests. I propose that significant differences between the two groups be considered as a proxy of campaign success.

- I present a framework to explore multimedia content, which includes audio recognition, text extraction from images, and deep learning techniques to identify the age and gender of users on Tiktok videos. With this framework, I analyzed the different levels of communication made possible by the platform design.

- I developed a multi-label classifier based on transfer learning and state-of-the-art language models that can detect conspiratorial comments. I employ this classifier to calculate the percentage of conspiracy comments on a YouTube video and select it as a feature to detect COVID-19 misinformation videos. The accuracy of the model reaches 89.4%. The high accuracy also applies by considering only the first hundred comments of a video. Thus, this method would work as an early-detection mechanism as it requires significantly less computation than a classifier of the video itself.

# 3 Social Media Monitoring

## 3.1 The Political Dashboard: A Tool for Online Political Transparency


#### Authors

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Morteza Shahrezaye, Simon Hegelich



#### In

Proceedings of the International AAAI Conference on Web and Social Media. Association for the Advancement of Artificial Intelligence (AAAI), volume 14, pages 983–985, (2020).
URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7371



#### Abstract

Contemporary political communication is a multi- and cross-platform process. Because of its complexity, new tools are necessary to monitor and understand it. We present a system that ingests, stores, and processes political data from Twitter, Facebook, and online news articles. We visualize the data in the form of a freely accessible online dashboard. *The political dashboard* (https://political-dashboard.com/) aims to provide online political transparency and assist researchers, journalists, and the general public in understanding the German online political landscape.


#### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# The Political Dashboard: A Tool for Online Political Transparency

**Juan Carlos Medina Serrano, Orestis Papakyriakopoulos**
**Morteza Shahrezaye, Simon Hegelich**
Technical University of Munich
{juan.medina, orestis.p, morteza.shahrezaye}@tum.de, simon.hegelich@hfp.tum.de

### Abstract

Contemporary political communication is a multi- and cross-platform process. Because of its complexity, new tools are necessary to monitor and understand it. We present a system that ingests, stores, and processes political data from Twitter, Facebook, and online news articles. We visualize the data in the form of a freely accessible online dashboard. *The political dashboard* (https://political-dashboard.com/) aims to provide online political transparency and assist researchers, journalists, and the general public in understanding the German online political landscape.

## Introduction

The web and the datafication of society have transformed political communication. Not only does news consumption increasingly take place online, but individuals and politicians also use online social networks as platforms for political exchange. Under this framework, political campaigns have developed new campaigning techniques, such as political microtargeting (Hersh 2015; Papakyriakopoulos et al. 2018), while traditional gatekeeping has been replaced by complex processes of news media production and consumption (King, Schneer, and White 2017). This new form of political communication occurs in a political space that spans over multiple platforms. On the one hand, it is interconnected but on the other, it is difficult to monitor and analyze. To that end, we developed *the political dashboard*, a tool that monitors digital media outlets, Facebook, and Twitter, with the aim to provide an overview of online political activities in Germany. The dashboard contributes to filtering and understanding of political information, providing multi- and cross-platform transparency.

## Data Collection

To monitor politically relevant data, we continuously collect data from different online sources. Our system consists of an array of Raspberry Pi devices that either connect to application programming interfaces (APIs) or employ crawling mechanisms to retrieve data. The collection procedure differs for each data source:

- **Twitter:** We collect tweets with the help of the Twitter Streaming API[1]. It allows us to retrieve data by providing a list of hashtags and users. We carefully select 239 relevant hashtags and 13,633 users, including accounts from political parties, politicians, media portals, journalists, bloggers, and other important political actors. We collect their tweets, mentions and retweets. For the hashtag list, we selected four types: German political parties, politicians, political topics, and media sites from all political orientations. We made an effort to avoid generating bias toward a specific political ideology through the data by carefully selecting a balanced list of hashtags and users that represented the complete German political spectrum. We are further aware that in the case of hashtags, Twitter only provides a sample of the complete tweets, which can make the data biased. However, we hope that by collecting a significant number of tweets, these insights are representative of political activity on Twitter.
- **Facebook:** We include two data sources from Facebook. First, we collect the posts from 102 public political pages; these correspond to the main page of the seven German political parties in Parliament and their regional pages from the 16 German states. We use the Crowdtangle[2] service to obtain the posts. The data do not include any personal data—neither the users who interacted with the posts nor their comments. Secondly, we collect political ads that target users in Germany on this platform. To do this, we connect to the Facebook ad archive API[3]. The archive has historical ads and active ads. We constantly update our database by collecting only the current, active ads.
- **News Outlets:** To retrieve online news media articles, we use RSS feeds of the news media websites and the Python package BeautifulSoup. We select 40 online German media sources from the top sites of online traffic in Germany (Alexa). We include media from all different political orientations and only collect the news articles that appear on the political sections of each news outlet.



[1]https://developer.twitter.com/en
[2]https://www.crowdtangle.com
[3]https://www.facebook.com/ads/library/api

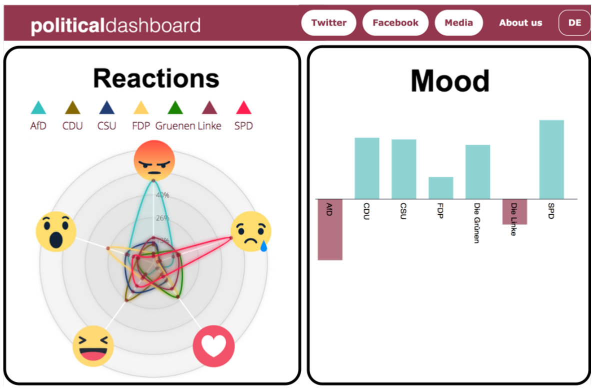


Figure 1: A screenshot of the political dashboard's Facebook page on January 10, 2020.

After retrieval, the Raspberry Pi devices send the data to a distributed Elasticsearch database. The data is then processed with Python scripts using open source libraries. Finally, the servers send the analysis results to an application web server where live plots are created for the dashboard. The system implementation uses batch processing for the ingested data[4]. We designed the system to be able to adapt to the changing political discourse. Adding and deleting entries in MySQL tables (e.g., Twitter hashtags or users) will have a direct effect on the collection procedures.

### Privacy Concerns

The constant and vast collection of data can raise privacy concerns. We neither display nor share the content of the collected tweets, Facebook posts, and news articles on the dashboard. We only provide aggregate information and analytical results. For the news articles, the full text is under strict data protection and should not be reproduced without the consent of the news media; therefore, we do not recreate the content in any way. Moreover, we do not provide individual data on the dashboard to ensure user privacy.

## Description of the Dashboard

The dashboard gives a live overview of online German political trends—for example, partisanship activities, the popularity of content, and issue saliency. Although we originally conceived it as a system for internal research purposes (Serrano et al. 2018), we decided to create a front-end public tool with live analyses. Users can navigate between three web pages corresponding to each of the data sources. In the Twitter main page, we present the top hashtags and media URLs of the last 24 hours. They differ from Twitter's trending topics as we only focus on politically relevant tweets. We explicitly distinguish between biased hashtags, which are the hashtags we have pre-selected, and unbiased hashtags, which co-occur with the biased hashtags and were used by the users we follow or the users that interact with them.

[4] e.g., 1,834,953 tweets, 64 Facebook posts, 2,380 active ads and 214 news articles in one hour on January 10, 2020

Moreover, we show for each German political party the top hashtags used by *partisan* users. We define partisan users as those who have retweeted a political party account more than five times. A spider plot shows to which percentage the followers of each party are using the general top hashtags.

For the dashboard's Facebook page, we first show the number of posts published by the political parties in the last seven days. We then display the accumulated number of likes and shares per party. We plot the rest of the user reactions together in a spider plot as they are often in the same order of magnitude. We also apply a sentiment analysis algorithm to assign a mood score to each party (Figure 1 depicts a screenshot of the reactions and mood plot.) A second part of the Facebook page concentrates on advertising. It shows the advertisers that have more active ads and the advertisers whose ads generate the most user impressions. A map of Germany shows the percentage of the extent to which the seven political parties are targeting each state.

The third page focuses on online news. With the help of topic modeling, we process the texts and show the top seven topics, each represented by eight most important nouns per topic. The page also shows the top news articles shared on Facebook as a proxy of general online interest. We further categorize the news outlets according to their political orientation and use a spider plot to compare the proportion of articles that each media group publishes on the top topics.

The current implementation has two limitations. First, the design process focused on users interested in current online political activities; we did not design the system to allow retrieving historical data. However, we can add this functionality in the future as we store all results on our servers. Second, we are aware that the data could be biased and not replicate all online interactions. However, we made extensive efforts to minimize the bias and collect data from all political orientations. After monitoring the dashboard constantly for one year, we are confident that the results are reliable and helpful to understand the online political landscape in Germany.

## Related Work and Impact

Few other websites collect online political interactions for public display. The WhatsApp monitor collects the most shared audiovisual content in WhatsApp public groups from Brazil, India, and Indonesia (Melo et al. 2019). It is part of the "Fake Elections" project, which has also developed a website that shows the number of likes and user demographics of politician's Facebook pages in Brazil (DCC 2018). Google and Facebook each provide search libraries to find political ads that were active on their platforms (Google 2019; Facebook 2019). The political dashboard stands out as it is a live monitor that shows processed analyses from three different sources. Our dashboard has already been an explicit information source to researchers, journalists, political candidates, and PR agencies. During the 2019 European elections it was used by the German Media Authorities. Since its creation, the dashboard has contributed to information extraction in multiple areas; for instance, in understanding the diffusion of right-wing and xenophobic content on social media platforms, as well as to understand partisan bias in media outlets.

## References


DCC. 2018. Political Audience: Demographics of Brazilian politicians and parties on Facebook. *Eleiçoes Sem Fake*. Retrieved January 10, 2020 from http://www.audiencia-dos-politicos.dcc.ufmg.br/.

Facebook. 2019. Facebook ad libraries. Retrieved January 10, 2020 from https://www.facebook.com/ads/library/.

Google. 2019. Political advertising on google. Retrieved January 10, 2020 from https://transparencyreport.google.com/political-ads/home.

Hersh, E. D. 2015. *Hacking the electorate: How campaigns perceive voters*. Munich, Germany: Cambridge University Press.

King, G.; Schneer, B.; and White, A. 2017. How the news media activate public expression and influence national agendas. *Science* 358(6364):776–780.

Melo, P.; Messias, J.; Resende, G.; Garimella, K.; Almeida, J.; and Benevenuto, F. 2019. WhatsApp monitor: A fact-checking system for WhatsApp. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 676–677.

Papakyriakopoulos, O.; Hegelich, S.; Shahrezaye, M.; and Serrano, J. C. M. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society* 5(2).

Serrano, J. C. M.; Hegelich, S.; Shahrezaye, M.; and Papakyriakopoulos, O. 2018. *Social Media Report: The 2017 German Federal Elections*. Munich, Germany: TUM University Press.

# 4 Political Actors

## 4.1 The Rise of Germany's AfD: A Social Media Analysis

### Authors

Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos , Simon Hegelich

### In

Proceedings of the 10th International Conference on Social Media and Society. Association for Computing Machinery, New York, NY, USA, pages 214–223 (2019)
DOI: https://doi.org/10.1145/3328529.3328562

### Abstract

In 2017, a far-right party entered the German parliament for the first time in over half a century. The Alternative für Deutschland (AfD) became the third largest party in the government. Its campaign focused on Euroscepticism and a nativist stance against immigration. The AfD used all available social media channels to spread this message. This paper seeks to understand the AfD's social media strategy over the last years on the full gamut of social media platforms and to verify the effectiveness of the party's online messaging strategy. For this purpose, we collected data related to Germany's main political parties from Facebook, Twitter, YouTube, and Instagram. This data was subjected to a unified multi-platform analysis, which relies on four measures: party engagement, user engagement, message spread, and acceptance. This analysis proves the AfD's superior online popularity relative to the rest of Germany's political parties. The evidence also indicates that automated accounts contributed to this online superiority. Finally, we demonstrate that as part of its social media strategy, the AfD avoided discussion of its economic proposals and instead focused on pushing its anti-immigration agenda to gain popularity.

### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# The Rise of Germany's AfD: A Social Media Analysis

Juan Carlos Medina Serrano
Bavarian School of Public Policy
Munich, Bavaria
juan.medina@tum.de

Morteza Shahrezaye
Bavarian School of Public Policy
Munich, Bavaria
morteza.shahrezaye@tum.de

Orestis Papakyriakopoulos
Bavarian School of Public Policy
Munich, Bavaria
orestis.papakyriakopoulos@tum.de

Simon Hegelich
Bavarian School of Public Policy
Munich, Bavaria
simon.hegelich@hfp.tum.de

## ABSTRACT

In 2017, a far-right party entered the German parliament for the first time in over half a century. The Alternative für Deutschland (AfD) became the third largest party in the government. Its campaign focused on Euroscepticism and a nativist stance against immigration. The AfD used all available social media channels to spread this message. This paper seeks to understand the AfD's social media strategy over the last years on the full gamut of social media platforms and to verify the effectiveness of the party's online messaging strategy. For this purpose, we collected data related to Germany's main political parties from Facebook, Twitter, YouTube, and Instagram. This data was subjected to a unified multi-platform analysis, which relies on four measures: party engagement, user engagement, message spread, and acceptance. This analysis proves the AfD's superior online popularity relative to the rest of Germany's political parties. The evidence also indicates that automated accounts contributed to this online superiority. Finally, we demonstrate that as part of its social media strategy, the AfD avoided discussion of its economic proposals and instead focused on pushing its anti-immigration agenda to gain popularity.

## CCS CONCEPTS

**• Networks → Social media networks**; **• Human-centered computing → Social network analysis**;

## KEYWORDS

political campaigns, social media, AfD, multi-platform, Twitter, Facebook, Instagram, YouTube

**ACM Reference format:**
Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos, and Simon Hegelich. 2019. The Rise of Germany's AfD: A Social Media Analysis. In *Proceedings of International Conference on Social Media and Society, Toronto, ON, Canada, July 19–21, 2019 (SMSociety '19),* 10 pages.
https://doi.org/10.1145/3328529.3328562


*SMSociety '19, July 19–21, 2019, Toronto, ON, Canada*

ACM ISBN 978-1-4503-6651-9/19/07…$15.00
https://doi.org/10.1145/3328529.3328562

## 1 INTRODUCTION

The rise of the Alternative für Deutschland (AfD) represented a schism in German politics. The AfD originally emerged as an anti-Euro party and then gradually adopted the language of right-wing populism [54]. The AfD has advocated for anti-Euro, anti-immigration and anti-refugee policies, and has been outspoken about other previously taboo topics in German politics. Its anti-establishment rhetoric parallels that of other EU right-wing populist parties, such as the National Front in France, the Party for Freedom in Netherlands and the Lega Nord in Italy. The recent surge in far-right voting in Europe calls for careful study of the roots of this new political trend.

The AfD was founded as a eurosceptic party by a group of university professors and former politicians in February 2013. Their proposals were centered on economic liberalism, ordoliberalism and free market ideas. Though the AfD was originally a single-issue party, it soon found support from right-wing groups and started shifting toward an anti-immigration ideology. Before the AfD, right-wing populist parties had achieved only limited electoral success in Germany. The AfD overcame this burden by distancing itself from previous right-wing ideologies and presenting itself as a party with economic expertise and scientific authority [22]. Moreover, it formed a stable nation-wide organization [4]. In the 2013 federal elections, the AfD missed the 5% threshold for entering parliament by only 0.3%. Nevertheless, by the next year, the party won seven seats in the European Parliament and later entered three state parliaments.

This rapid increase in electoral success would not have been possible without the AfD's wide base of supporters. At its beginnings, the AfD's constituents were mostly well educated, high-income citizens [5]. Following 2014, the party enjoyed a great surge of support from low-income citizens [42]. According to the latest study of the issue, the 2017 election report from the Infratest dimap Institute [10], the largest demographic group that voted for the AfD was East German men. Kim [30] investigated why lower socio-economic groups like blue-collar workers and the unemployed would support a party that advocates radical market-oriented policies, which would not benefit them. Kim argues that the AfD strategically avoided discussion of the party's economic proposals to prevent divisions among its supporters.

The AfD's base of support tripled in recent years, from 5% in 2015 to 15% in 2018. This inflection in the opinion polls started in September 2015, at the beginning of the *refugee crisis*. The AfD's popularity in polling grew over one year, and then the polls remained stable until several months before the 2017 federal elections. The party's support increased further after the elections, indicating approval of the party's work in parliament.

The rise of the AfD cannot be understood fully without taking its online presence into consideration. This study aims to describe the AfD's social media strategy and compare its effectiveness with that of other German parties by examining various platforms over multiple years. First, we give an overview of political parties' activity on social media and then we focus on the AfD's user engagement strategy. We propose a unified multi-platform analysis for the evaluation of all the different social media channels. This analysis covers the four social media platforms that are used regularly by German political parties to examine how political messages spread across them. To further understand the AfD's strategy, we compare its social media discourse with their campaign proposals. Moreover, we show that the spread of the AfD's messages was boosted by automated accounts on at least one of the platforms.

# 2 THEORETICAL BACKGROUND

## 2.1 Political Parties' Activity on Social Media

The use of social media by political parties has transformed political communication. Since the World Wide Web provides a plurality of possibilities for communicating with the electorate [43], political parties have developed new methods and tools for externalizing party attitudes and evaluating the reactions of potential voters [59]. In social media, candidates and parties have official pages and accounts, through which they make political statements and declare their positions on salient issues. These messages are diffused further on the platforms by journalists and other users [64]. Voters can then respond to political actors, providing them with rapid and granular feedback; the directness of this political dialog was not possible when only traditional media was available [44]. This form of political interactivity has been proven to benefit political actors, making them more favorable to the electorate [63]. Additionally, in accord with existing trends in data collection [34], political parties are using social media to collect, monitor, and analyze voter reactions to political messaging [56]. These analyses help the parties to design, correct, and strategically adapt campaign activities. Finally, political parties use social media as spaces for political microtargeting [45], sending personalized messages to users to encourage support. Social media has become so critical to political campaigning that social media has become 'environmental'[58]: parties and candidates cannot avoid or neglect its use, as social media platforms are now a cornerstone of political communication.

Given the constant flow of information on social media, political parties and candidates no longer present a static or complete overview of their views and positions to the users. They tend to comment and respond on topics that were made prominent by exogenous events, like economic crises or natural catastrophes, or they adapt rapidly to the topics in the agenda set by mass-mediated public debates and news-media platforms [25]. In addition, they choose to address topics about which they hold strong and influential positions and that are of concern to voters [68], while avoiding other topics that might decrease their popularity. Finally, the parties concentrate on issues and strategies that are tailored to the audience on each social media platform [62]. This behavior means that the image of the political parties that is presented to the electorate often deviates from the official party positions that are expressed in political manifestos [25].

Although the above behaviors and strategies are characteristic of all parties, the opportunities available on social media are of greater importance for outsider parties. As they have limited access to traditional mass media, they use these new communication channels to overcome disadvantages in communication and to contact potential voters [31]. Therefore, outsider parties emphasize their online presence and intensify their interactions with users to achieve effective communications [27]. Due to their limited resources, they are even more selective about the topics that they express opinions about and they deploy strategies that promote only their strongest arguments [68]. For example, left-wing populist parties tend to concentrate on economic issues on social media, while right-wing populist parties tend to focus on issues that resonate with xenophobic voters [14].

## 2.2 The AfD's Social Media Strategy

As an outsider party, social media has been an important communication channel for the AfD since its foundation, because social media platforms provided a space to influence public opinion outside of the traditional media. In recent years, the AfD has been effective on social media as reported on media channels and in previous research: Arzheimer [1] analyzed Facebook posts from 2013 and 2014 and found that the AfD used more populist rhetoric on Facebook than it did on other communication channels; Schelter et al. [52] evaluated the Facebook posts of six political parties in Germany in 2014 and 2015 and reported that social media was a major factor in the success of the AfD; and both Hegelich [23] and Medina Serrano et al. [36] studied social media campaigns in the months leading up to the 2017 German federal election.

From the literature, we deduce three main factors that help explain the AfD's effectiveness on social media:

- **Alternative media:** The AfD relies on social media platforms to spread its message. The party leaders have blamed traditional media for presenting them in a negative light and obscuring their intentions. Using social media as an alternative ecosystem, the AfD reached a part of the German population that felt disenchanted with conventional communication channels. Indeed, a study from the Otto-Brenner-Stiftung [9] confirmed that followers of the AfD place less trust in the German media. Hence, they prefer to obtain information from social media platforms. Furthermore, a 2018 Pew Research Center report [46] ascertained that people with populist preferences in Germany tend to have less trust in the media. The right-wing political party has taken advantage of this fact by employing a strong social media campaign.
- **High online activity:** The AfD's strategy is to make use of social media as much as possible and get its content to go viral. To achieve this, the AfD regularly asks its supporters to share content. Furthermore, it uses a provocative tone, which together with its critical position on political correctness [40], encourages users to engage and reply with positive

or negative comments. Another factor that stimulates the high user response is the negative and aggressive tone in the AfD's anti-establishment and anti-immigrant stances, following the work of Fan et al. [15] that showed that hate spreads faster on social media.

- **Online Manipulation:** The right-wing party was not alone in spreading its message, as pro-AfD social bots were active on Facebook [3] and Twitter [41]. Social bots [16] are automated fake accounts that are fashioned to look like real users and whose purpose is to viralize topics and manipulate trends. Even though Neudert et al. [41] found low levels of automation in the time leading up to the 2017 German federal election, they did find that social bots in their sample were working in favor of the AfD. Although it is not possible to track the origin of these bots, two online communities —Infokrieg and Reconquista Germania— had the explicit goal of trolling social media in support of the AfD [12]. While the effects of online manipulation attempts on public opinion are difficult to quantify, these automated accounts likely amplified the online reach of the AfD's message.

These three points are in line with the social media activity of other populist parties in Europe [11, 35, 53]. Social media has given populist political actors greater freedom to articulate their ideology and spread their message [14]. Social media acts as a sort of people's voice in populist movements [19] facilitating the reinforcement of the anti-establishment ideology that is common to populist parties [37]. Furthermore, social media has no gatekeepers to fact-check the information, which gives populists a fertile space to spread their rhetoric [29].

### 2.3 Multi-Platform Schema

In order to understand and evaluate the AfD's strategies on social media, a long-term multi-platform analysis is needed. Even though research on digital campaigns over the last two decades has been extensive, few studies have focused on how candidates and parties use multiple social media channels [50, 57]. The social media environment itself has become more complex, and the fast pace of technology and digitalization calls for researchers to adapt their methods to cross-platform research. Indeed, researching media platforms separately ignores the reality of today's contemporary media experience [7]. Already in 1968, when the term media ecology was first coined, researchers realized that media should not be considered in isolation [48]. However, researchers tend to conveniently study the types of social media that can be accessed, without taking into consideration the political relevance of specific aspects of diverse media platforms [28].

Rains and Brunner [49] found that between 1997 and 2013 more than two-thirds of studies on social media were limited to a single platform, most often Facebook. However, research on Twitter has become more predominant in the last few years [26], given the easy availability of the data in comparison to other platforms. Moreover, Facebook decided to restrict the access to its API, which is likely to impede future research on this platform.

On the other hand, there exist only a handful of studies about political campaigns that focus on data from YouTube [33, 66] or Instagram [17, 39]. However, according to a 2018 Pew Research Center report [47], these two platforms have become the most popular among younger citizens, which may make them more relevant to politics in the near future. Additionally, politicians are not bound to use only the thoroughly-researched platforms. Germany's chancellor, Angela Merkel, for instance, has an Instagram account but not one on Twitter.

Bossetta [8] presents a framework for comparing different platforms and how their idiosyncratic features affect political campaign strategies on social media. He exhibits differences between platforms in terms of their network structures, functionalities, and algorithms. Our multi-platform analysis does not focus on the platforms' structure but instead deals with online interactions between political parties and users. We hope that this focus will prove helpful in advancing understanding of the overall social media strategy of political parties.

## 3 DATA AND METHODS

We propose a multi-platform approach that unifies the different features of each social media channel. The analysis is based on the following four categories:

- **Party engagement:** Quantifies the activity of a political party on a social media channel. It constitutes the main source of interaction with the users.
- **User engagement:** Represents the amount of users' online interaction with the political party. Interactions may be in the form of a direct response to the social media content or a message sent to the party's account. User engagement does not imply direct support for a party, given that it may have a positive or a negative tone.
- **User support:** Determines the level of users' acceptance of and alignment with the party's social media content.
- **Message dissemination:** Quantifies the success of the party in spreading its message across a social media channel, which is one of the main purposes of an online political campaign.

These measures give a general overview of the fields in which a political party is performing better than others. The specific measures used on each platform and their association with one of the four categories are listed in Table 6. Each of these is explained thoroughly in the next section.

In order to apply this multi-platform approach, we collected data from Facebook, Twitter, YouTube, and Instagram using the application programming interfaces (APIs) of each platform. We also collected the AfD's 2017 manifesto as a reference for the party's official ideology and proposals. From the social media channels, we collected data for the AfD and for the other six main political parties in Germany: CDU, Germany's main conservative party; CSU, the sister party of the CDU in Bavaria; Bündnis90/Die Grünen, the green party in Germany; FDP, a neo-liberal party; SPD, Germany's social-democratic party; and Die Linke, the radical left party. This allowed us to compare and measure the AfD's effectiveness on social media against the effectiveness of the other parties' online activity.

For Facebook, we retrieved posts made by the political parties in the period from January 2015 to May 2018. These posts amounted to a total of 12,912 posts. The data included all comments and reactions to the posts and their respective comments. The number of posts is smaller in comparison to those collected by Arzheimer [1] or Schelter et al. [52]. This is due to changes in the Facebook API. Previously, the API allowed access to drafts, status changes, and post modifications. Our data only includes the final posts written by each party, which is helpful, since we are not interested in considering every modification made by the page administrators.

For Twitter, we collected the tweets from political parties' Twitter accounts over the one-year period, starting in July 2017. We also included tweets from users that mentioned or retweeted the political parties. This dataset includes 1,961,318 tweets. We also collected tweets that included the name of one or more of the political parties. Overall, we gathered 30,437,991 tweets.

We obtained the tweets from the parties using Twitter's Search API, which allows access to the last 3,200 tweets from an account. We gathered the rest of the tweets with an automated procedure that continuously accesses data from Twitter's Streaming API. In contrast to the Search API, the Streaming API allows the retrieval of tweets in near real-time based on certain criteria like hashtags, keywords or geolocations. The limitation of this API is that it only provides a sample of the complete tweets. The enterprise version of the API, called Firehose, makes it possible to query the entire Twitter history, but its cost is prohibitive. Morstatter et al. [38] analyzed the differences between Firehose and the Streaming API and showed that samples gathered from the public API are biased.

For YouTube, we used its Data API to collect metadata from videos published between October 2016 and May 2018. We focused on the videos published on the YouTube channels belonging to the political parties. The AfD has two channels, namely *AfD Kompakt* and *AfD-Fraktion Bundestag*, while each of the other parties only has one. The former was created in October 2016 and the latter in December 2017, two months after this party entered the parliament. From their two channels, we also extracted the videos' subtitles using the Linux package *youtube-dl*. Some of the videos do not include dialogue, but only have a written message. We manually transcribed these videos for further analysis.

For Instagram, we collected 4,155 Instagram posts from the political parties' accounts before the capabilities of Instagram's public API were diminished. The period of time matches the period we used for collecting Facebook data. For Twitter, the period of time is shorter, given that historical data cannot be collected with Twitter's public API. For YouTube, we only collected data beginning with the creation of AfD's first YouTube channel.

In order to find insights in the collected data, we applied the following methods:

*Exploratory Data Analysis.* We first performed simple qualitative and quantitative analyses on the data. We gathered and summarized the interactions for each platform. For Facebook and Twitter, we also included a time series of how the parties interacted over time. For the four different measures in the multi-platform schema, we performed Kolmogorov-Smirnov tests to determine the goodness of fit to the log-normal distribution and used Vuong tests [67] to compare with similar distributions. The results provide a better intuition about how the different social media measures behave.

*Bot Detection.* A growing body of literature deals with social bots and their influence on politics [60]. Most of these studies analyze the percentage of bot activity in a given narrative. For example, Neudert et al. [41] analyzed the users who tweeted hashtags related to German political parties a month before the 2017 elections. They found that tweets with AfD-related hashtags showed the highest percentage of users that behaved as if they were automated. However, even though hashtags are helpful to characterize the Twitter conversation, their use does not directly indicate support for the party since they can be attached to both positive and negative messages. In contrast, we concentrate on users who spread the political parties' messages directly and look for bot behavior. We can perform such an analysis only on Twitter, since information about who is sharing which political content on the other platforms is not available.

We categorized the users who had retweeted the parties' original contents with the help of Botometer[1], which is a public bot detection framework created by the University of Indiana. The framework implements a machine learning algorithm that has been trained on tens of thousands of labeled examples [65]. For a given user, it returns a score from 0 to 1, which determines the probability that the user is a bot. We selected 0.5 as the threshold to classify bots. Additionally, we only used Botometer's language independent features for the classification, given that the other features can only give accurate predictions when the content is written in English.

*Topic Modeling.* Topic modeling algorithms are based on statistical models that discover topics from a text corpus. We selected Latent Dirichlet Allocation (LDA) given its extensive use in the literature [6]. LDA takes a group of documents and treats each document as a combination of topics. Each topic is then defined from a collection of words. After creating a list of topics, the trained model assigns the probability of belonging to each topic to a document. Within each document, the probabilities sum up to one. The probabilities and topic distributions are helpful in comparing different large text corpora. Hence, we used this method to contrast the AfD's online content against the proposals included in their party manifesto.

For LDA to work, the number of topics (K) must be predefined. We decided to train the model on twenty topics. The algorithms for calculating the optimal number of topics did not converge, so we chose the most appropriate K after experimenting with different parameters. Moreover, the model requires two hyperparameters: $\alpha$, the prior of the topic distribution; and $\beta$, the prior of the word distribution. We set $\alpha = K/20$ and $\beta = 0.01$, as suggested by Griffiths and Steyvers [21]. For the implementation, we relied on *nltk* and *tmtoolkit*, two Python toolkits used for natural language processing.

In the next section, we discuss the data analysis and the results of these methods. First, we consider each social media platform separately and then we analyze them together with the proposed multi-platform approach.

# 4 FINDINGS

## 4.1 Facebook

For years, German political parties have used Facebook as their main form of online communication. They interact with users by

[1] https://botometer.iuni.iu.edu/#!/

creating posts that state their views and ideology on group pages. Of the seven parties, the AfD has the page with the most fans, with two times as many fans as the pages of CDU and SPD, which are the current ruling parties. After initially losing support in early 2015, the AfD's number of fans increased sharply when the *refugee crisis* arose [11]. The AfD's fan count almost doubled in a single year, from around 140,000 to 260,000 followers. This parallels the increase in support of the AfD shown in opinion polls. Both online and offline, the AfD grew to be an important political force in Germany.

Table 1 presents the results of our analysis of Facebook data. The most active party was the AfD, with 2,363 posts. At the same time, its post received the most comments. Each post had an average of 420 comments. In comparison, the CDU page had an average of 160 comments per post. Comments can have a positive or negative connotation. Hence, a high number of comments does not directly translate into party support. Negative comments were both in favor and against the posts' messages. Sentiment analysis of the comment corpus would not suffice to determine party support since the methods can only classify text into positive and negative categories. The context is necessary for understanding the nature of the comments.

**Table 1: Facebook statistics for the German political parties in the period from January 2015 to May 2018.**

| | posts | comments | likes | shares |
|---|---|---|---|---|
| *AfD* | 2,363 | 994,191 | 4,168,022 | 2,891,377 |
| *CDU* | 1,690 | 272,155 | 483,924 | 153,131 |
| *CSU* | 2,162 | 406,804 | 1,897,622 | 634,153 |
| *Die Grünen* | 1,127 | 142,473 | 625,689 | 411,073 |
| *Die Linke* | 1,367 | 140,489 | 903,629 | 437,920 |
| *FDP* | 2,211 | 118,277 | 755,000 | 192,974 |
| *SPD* | 1,992 | 247,095 | 892,198 | 421,025 |

The number of shares is more representative of the party's reach. When a user shares a post, it appears on the timelines of the user's Facebook friends. Posts with more shares have reached more users on the platform. The number of shares of the AfD's posts is larger than the sum of the shares of all posts from the rest of the parties. This is a clear signal of the wide reach of the AfD on Facebook and of its online popularity.

The CSU comes in second in terms of the numbers of comments and shares. The CSU is a conservative party that operates only in the state of Bavaria, while its close counterpart the CDU operates in the rest of Germany. The CSU is more conservative than the CDU on social issues and is closer to the political spectrum of the AfD [18]. These results suggest that users with right-wing ideologies are more politically active on Facebook. Although the CSU performed well in terms of Facebook activity, it has lost voter support over time. Since 2015, the CSU's approval rating has gone down ten points in the opinion polls.

Figure 1 shows the number of posts per month made by the pages of the German parties. The pattern is similar for all the parties. The large peak in the plot corresponds to the month of the 2017 parliamentary elections. In the months following the election, the AfD continued to post content on Facebook, whereas activity from the rest of the parties declined.

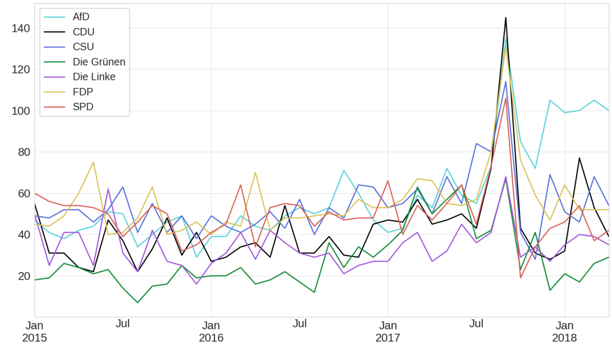


**Figure 1: German political parties' Facebook activity: Number of posts per month between January 2015 to May 2018.**

We further analyzed the AfD's posts. The format of these posts consists of a message and an image that combines a short text with a picture. The tone of these messages tends to be provocative and sometimes is even sensationalist. The topics discussed are controversial, which encourages users to engage with the posts and express personal opinions.

To perform a quantitative analysis of the posts, we preprocessed the text by removing stop words and punctuation marks. The most frequent nouns are AfD, Germany, politics, EU, Merkel, Euro, German people, SPD, and citizen. Indeed, the general message is that the AfD is on the side of Germany and its citizens, and it is against the Euro and the establishment parties, which are represented by *Merkel* and *SPD*. Another word appearing on several posts is the verb *teilen* (to share). This suggestion was the fifth-most used verb in the posts, indicating that part of the AfD's strategy is to viralize its content by explicitly asking fans to share it.

## 4.2 Twitter

All of the German political parties have a Twitter account that interacts with politicians, journalists, and other users. In contrast to Facebook, the AfD has the fewest followers on Twitter [55]. Nevertheless, this lack of followers does not imply that they are less successful on this platform. For example, more than 50% of the political conversation on Twitter on the day of the 2017 federal election was related to the AfD [23].

As in our analysis of Facebook data, we selected four corresponding measures on Twitter: number of tweets, likes, mentions and retweets (Table 2). We divided the number of tweets into two categories: all tweets and original tweets, the latter of which are those with content from only the party account, not including retweets from other users. The format of the AfD's original tweets resembles that of the Facebook posts, as a brief message together with a picture. Most of the tweets include a link to the corresponding Facebook post. However, given the character limit, the message is shorter on Twitter. The tweets also include hashtags. The top-used hashtags by the AfD Twitter account are AfD, TrauDichDeutschland, BTW17, Bundestag, Merkel, SPD, Gauland, CDU, FDP, and GroKo. TrauDichDeutschland is AfD's campaign slogan, BTW17 refers to the 2017 election, Gauland is one of the leaders of the AfD, and GroKo is the grand coalition between CDU, CSU, and SPD.

**Table 2: Twitter statistics in the period from July 2017 to July 2018.**

| | tweets | original tweets | mentions | likes | retweets |
|---|---|---|---|---|---|
| *AfD* | 9,193 | 2,092 | 368,005 | 638,886 | 269,445 |
| *CDU* | 4,911 | 3,097 | 345,192 | 117,437 | 39,726 |
| *CSU* | 2,886 | 1,622 | 233,012 | 70,474 | 14,812 |
| *Die Grünen* | 2,492 | 1,295 | 157,213 | 124,371 | 42,183 |
| *Die Linke* | 6,809 | 1,776 | 208,047 | 136,440 | 45,856 |
| *FDP* | 3,149 | 1,730 | 189,687 | 121,220 | 31,487 |
| *SPD* | 7,480 | 1,782 | 260,056 | 119,507 | 41,803 |

On Twitter, the AfD was also the most active party with the largest number of tweets. 77 percent of these tweets were retweets and most of them were from other regional AfD Twitter accounts or AfD politicians. SPD and Die Linke followed a similar pattern. CDU published more original tweets than the other parties, with 66 percent of the tweets being original content. Figure 2 shows the tweet activity over the one-year period. The AfD was more active than the other parties most of the time. The activity of all the parties went up during the election month and went down afterward. In contrast to the post activity on Facebook, the AfD did not continue to tweet at the same pace as during the months before the election.

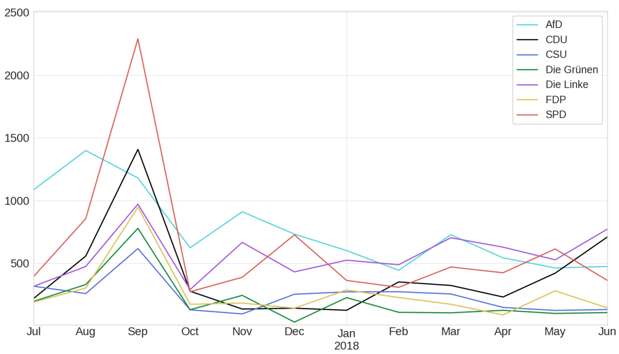


**Figure 2: German political parties' Twitter activity: Number of tweets per month during a one-year period.**

Like Facebook comments, mentions allow users to reply to a tweet or send a message directly to the party account by using the @ symbol and the screen name of the account. We excluded retweets that included a mention of a political party from this analysis. Additionally, if a tweet mentioned more than one party, each mention was counted separately. The results show that the AfD received the most mentions and CDU came in the close second place. Even though Die Grünen has the most followers, it had the fewest mentions. Note that the mentions in our data come from the sample that the Streaming API provides, which is only a selection of all mentions published to the platform over the period of interest.

While mentions can contain positive or negative messages about the party, likes and retweets serve as measures of support [24]. For both these measures, there is a large difference between the AfD values and those of the other parties. Like shares on Facebook, the AfD's tweets were retweeted more than the tweets of all other parties combined. Retweets and likes can only originate from the account's original tweets. On average, each of the AfD's original tweets was retweeted 129 times and had 305 likes. For comparison, each CDU tweet was, on average, retweeted 13 times and had 38 likes. This corresponds to a difference of one whole order of magnitude.

Not only the total number of retweets is of relevance, but also the information on which users retweeted the party accounts. With this data, we could investigate how many of the party retweets were published by social bot accounts. We obtained the Botometer score from all the users who had retweeted a political party during the month of September 2017. This subset of data included 111,919 retweets from 22,396 unique users. We assigned accounts that were closed by Twitter directly to the bot category. After classifying the users, we calculated the percentage of retweets from the bot accounts. Table 3 shows the results for each party. With almost 33 percent, the AfD is the party with the maximum number of bots retweeting each party's content. The CDU, CSU, and FDP follow with bot retweets between 20 and 25 percent. The parties with the lowest percentage of bot retweets are SPD, Die Grünen and Die Linke.

**Table 3: The number of retweets in September 2017 and the percentage of those that belong to accounts classified as bots.**

| | retweets | bots(%) |
|---|---|---|
| *AfD* | 43,633 | 32.92% |
| *CDU* | 14,603 | 21.15% |
| *CSU* | 3,738 | 24.02% |
| *Die Grünen* | 15,440 | 15.68% |
| *Die Linke* | 12,888 | 12.97% |
| *FDP* | 7,905 | 23.45% |
| *SPD* | 13,712 | 12.99% |

Note that these results do not provide an exact quantity of bot activity since the bot detection methods are not always accurate. Even the term *bot* is defined loosely across the literature [20]. Nevertheless, the comparison between parties is of relevance since the percentages vary considerably. Interestingly, center, center-right, and right-wing political parties have a higher proportion of bot retweets. Hence, we conclude that automated accounts on Twitter were more likely to spread messages from the center and right side of the German political spectrum.

## 4.3 YouTube

YouTube is different from the other platforms in its focuses on videos instead of text and images. YouTube allows political parties to publish video content that can then be shared on other social networks. In the first years after its conception, the AfD did not use a YouTube channel to spread its message. The channel AfD Kompakt was first created in October 2016 and is closely related to the AfD's member magazine of the same name. It claims in its description to be the AfD's official YouTube channel. Three months after the 2017 elections, the AfD faction in parliament created a new channel called AfD-Fraktion Bundestag. We considered both channels in our analysis since both seek to spread official AfD content. The other parties have only one YouTube channel with official videos.

Table 4 shows the party activity and user response to the parties on YouTube. The party activity varies, with Die Grünen having published the most videos in the period of interest. The difference between parties is greater when taking into account the number of comments on the videos. The difference is not entirely due to user interaction, but is affected by the fact that the comment sections of some videos were disabled. For CSU, FDP and Die Grünen, several videos have comments deactivated, and the AfD Kompakt channel does not allow any comments on its videos. The 8,080 comments on the AfD's videos are therefore all from its second channel. Even though this channel has been active only since December 2017, the AfD-Fraktion Bundestag's videos have more comments and likes than any of the other parties' channels.

**Table 4: Youtube statistics for the German political parties in the period from October 30 2016 to May 2018.**

| | videos | comments | likes | dislikes | views |
|---|---|---|---|---|---|
| *AfD* | 454 | 8,080 | 60,375 | 2,458 | 2,049,008 |
| *CDU* | 264 | 4,988 | 5,574 | 9,926 | 385,262 |
| *CSU* | 166 | 44 | 2,410 | 1,728 | 482,586 |
| *Die Grünen* | 479 | 665 | 27,480 | 26,199 | 5,283,833 |
| *Die Linke* | 204 | 5,791 | 20,043 | 6,912 | 814,219 |
| *FDP* | 63 | 910 | 341 | 90 | 853,673 |
| *SPD* | 236 | 3,573 | 22,469 | 22,348 | 2,418,132 |

A unique YouTube feature is the option to dislike a video. A dislike depends on the context of the video and does not always translate into opposition to the party. The AfD channels have the lowest ratio of dislikes to likes. Both Die Grünen and SPD have nearly equal numbers of likes and dislikes, whereas the CDU has considerably more dislikes than likes.

Video popularity is measured by the number of views. As with the publishing activity, Die Grünen's videos have the most views. The SPD has the channel with most views after Die Grünen even though they posted less than half the number of the videos. The AfD comes in third place for this measure. From all the parties, the most-seen videos are the campaign commercials. Most of the videos are of press conferences, campaign talks, and appearances in the German parliament by politicians of the party. These videos receive less attention than the others.

## 4.4 Instagram

Little research has been published about Instagram's potential in political campaigns [51]. Thomson and Greenwood [61] found that users in the US are less likely to engage with political images on Instagram. In contrast, a survey in Germany by Eckerl and Hahn [13] showed that the platform holds great potential for political communication. In the collected data, we observe active political campaigns from the German political parties. Each has an active Instagram account.

Table 5 shows that the AfD takes second place in terms of activity after the CSU. Even so, user interaction with the AfD's posts is greater than interaction with the CSU's posts. Die Linke has a similar number of posts and likes as the AfD, but their posts attract significantly fewer comments. Indeed, the number of comments on the AfD's posts is larger than the sum of the comments on the posts of all the other parties.

The AfD's strategy on Instagram is like its strategy on Facebook. Most of the party's Instagram posts are a subset of the Facebook posts. They include an image that embeds a short text and have a longer text in the image description. Since Instagram is more image oriented, the highlight of the information is embedded in the image. The top nouns used in the descriptions are the same as those used in the Facebook posts but are ranked in a different order. The subset of messages in these Instagram posts is representative of AfD's messages on Facebook.

**Table 5: Instagram statistics for all the German political parties in the period from January 2015 to May 2018**

| | posts | comments | likes |
|---|---|---|---|
| *AfD* | 870 | 68,399 | 469,380 |
| *CDU* | 229 | 4,702 | 122,353 |
| *CSU* | 958 | 4,880 | 154,291 |
| *Die Grünen* | 514 | 16,689 | 317,048 |
| *Die Linke* | 825 | 13,313 | 418,208 |
| *FDP* | 584 | 8,397 | 325,941 |
| *SPD* | 175 | 4,098 | 105,067 |

## 4.5 Social Media Comparison

In order to compare and summarize the results of the previous subsections, we applied the aforementioned multi-platform schema. Table 6 lists which features of each of the social media platforms

**Table 6: Features on each social media channel divided into the four categories from the multi-platform analysis. The collected data from the features in bold letters are more likely to follow a log-normal distribution and give better fits than the exponential, Poisson and power-law distribution according to the KS and Vuong tests. This is also the case for the features in italics, with the only difference that the log-normal distribution did not give a better fit than the exponential distribution.**

| | Facebook | Twitter | YouTube | Instagram |
|---|---|---|---|---|
| *party engagement* | posts | tweets | videos | posts |
| *user engagement* | **comments** | mentions | comments | **comments** |
| *user acceptance* | **likes** | *likes* | *likes* | *likes* |
| *message dissemination* | **shares** | **retweets** | *views* | – |

falls into each one of the four categories. The only missing entry on the table is the message dissemination measure for Instagram. This would correspond to the number of times the pictures from a party account have been observed. This information is private and is only accessible to the owner of a business account, and is even then restricted to the account's own posts.

With the aid of the four measures introduced in this table, we proceed to compare the level of effectiveness of the AfD's activity on social media. Unfortunately, a quantitative cross-platform comparison is not possible, since the period of data collection differs between the social media platforms. Across the social media channels, the AfD has a high party engagement, taking either the first or second place in this category. This is part of their aforementioned social media strategy. In terms of user engagement, the AfD surpasses the rest of the parties by a considerable margin. As mentioned before, the tone of the party's message, together with the controversial topics they tend to discuss, such as immigration and the economic crisis, encourage users to engage more with the posts. The same superiority is shown in user support. In this category, the difference between the AfD and the rest of the parties is even larger than the difference in user engagement. This suggests that the user base that supports the party is also regularly active on social media. Regarding the last category, the AfD spread its message on Facebook and Twitter more effectively than the other German political parties by an order of magnitude. The same dominance is not observed on YouTube. Nevertheless, the fact that the AfD decided to open a channel after entering the parliament suggests that they now consider YouTube as part of the party's social media strategy. In the next few years, the AfD may prove to be as successful in spreading their message on YouTube as it has been on the other two platforms.

For the measures of user engagement, user acceptance and message dissemination categories, we provided aggregated results that were independent of the party activity. We deliberately did not focus on efficiency measures like the average number of Facebook likes per post or the average number of views per YouTube video in this analysis since the social media features are far from following a normal distribution. The great majority of social media posts attract little user engagement, in contrast to the very few posts that draw large amounts of attention. Indeed, research has shown that most complex social media interactions follow a log-normal distribution [2, 32, 69]. We assessed this statement by testing the distributions of the features per unit of social media activity. This analysis considered the four social media channels and all seven German political parties[2]. We used bootstrapped Kolmogorov-Smirnov tests with 50 samples and a p-value of 0.05. In all cases, the tests failed to reject the null hypothesis that the data is generated from a log-normal distribution. We further compared the log-normal distribution with the exponential, power-law and Poisson distribution by implementing Vuong tests with a p-value of 0.05. In all cases, the log-normal distribution was a better fit than the Poisson or the power-law distributions. However, in some cases, the Vuong test failed to show that the log-normal distribution gave a better fit than the exponential distribution. The features with at least one case of a failing Vuong test are shown in Table 6 in italic letters. On the other hand, the features in bold letters fit better to the log-normal distribution than to the other three distributions for the data from all of the political parties.

### 4.6 Discourse Comparisons

The last step in our analysis consisted of exploring the difference between how the AfD presents its goals to followers online and the content of party's explicitly stated political intentions and motives. For this analysis, we only considered the Facebook posts, YouTube videos and Instagram posts created during the same one-year period over which we collected the Twitter data, so that the period of interest for all four channels was the same. With the help of topic modeling, we compared the topics between the different communication channels with the topics included in the party's manifesto.

For this analysis, we treated each post, tweet and video as a document. We divided the text of the manifesto into paragraphs and defined each of these paragraphs as a document. The resulting corpora include 1,113 Facebook posts, 9,213 tweets, 436 YouTube video captions, 866 Instagram posts, and 395 paragraphs from the manifesto. Preprocessing was applied to eliminate stop words, punctuation marks, and applying the Snowball stemming algorithm to the remaining words. We additionally removed the string "RT" from the retweets.

Since the topics are created algorithmically, the interpretation of each topic relies on human curation. To compare the five corpora, we selected the topics that were directly connected to the economy

[2]We did not evaluate the YouTube comments since many videos have them blocked, and the Twitter mentions since they do not always represent a response to a tweet.

or immigration. We then calculated the percentage of documents that included these topics. For this calculation, we summed up the probabilities of each document from the selected topics and then divided by the number of topics. Table 7 shows that economy and immigration are treated equally in the manifesto, whereas in the AfD's Facebook, Twitter and Instagram content, immigration topics are discussed significantly more than economic topics.

In the case of YouTube, only 18% of the total discourse is related to immigration or economic topics, and these are discussed in a similar proportion. This exception can be attributed to the fact that in contrast to the other social media platforms, there is a lack of platform-specific generated content. Most of the videos on the AfD's channels are political speeches, which represent the oral discourse of the right-wing party which is not necessarily related to its social media message. Further research into discourse comparisons could focus on the differences between the speeches given at campaign events and those given during parliament sessions.

By comparison, immigration topics are discussed more frequently on Twitter and Instagram than on Facebook. Retweets were included in the Twitter data, perhaps explaining the 6% difference between the Twitter and Facebook corpora. For Instagram, the AfD's content is mostly a subsample of the Facebook posts, which indicates that the AfD deliberately favored immigration-related content on this platform. With this discourse analysis, we prove quantitatively that the AfD-generated content on social media downplayed the party's economic proposals and focused on immigration topics, which validates the analysis published by Kim [30].

**Table 7: Percentage of documents related to topics discussing economy or immigration on different platforms.**

| | economy | immigration |
|---|---|---|
| Manifesto | 21% | 19.2% |
| Facebook | 4.5% | 16.1% |
| Twitter | 4.7% | 21.8% |
| Youtube | 9.9% | 8.2% |
| Instagram | 7.6% | 28% |

## 5 CONCLUSION

Since its foundation, the AfD has used social media as its primary communication tool. In this paper, we proved that the AfD's online activity prompted the most user interactions of any German political party. Our analysis covers four social platforms over a longer period of time than has been considered in previous research. We mapped the AfD's social media strategy and illustrated some of the differences between the party's online discourse and its manifesto. We also confirmed that that the spread of the AfD's message was boosted by automated accounts to some extent.

Although we cannot prove a direct connection between poll gains and social media dominance, we conclude that the AfD succeeded in spreading its message on social media. This message has also entered the limelight in traditional media and has permeated Germany's public agenda. The success of the AfD's social media campaign together with traditional media coverage was essential in spreading and stimulating anti-establishment sentiment, which partially explains the rise of a far-right party in Germany.

The AfD's effective social media campaign raises the question of what strategies the other political parties will take to improve their online activity. There is no simple solution that can attend to this question. Digital campaigns have to evolve and take into consideration the new trends. In particular, they have to be creative to attract the younger citizens and at the same time reach those citizens that are disappointed with current politics.

Overall, the results suggest that there exists a shift in Germany's online political communication induced by AfD's social media dominance. Future research should explore the causes and effects of this shift. For example, it is plausible that the shift has led to an increase in online polarization or to user discussions becoming more aggressive. These phenomena have to be analyzed taking into consideration the potential biased reality caused by automated accounts.

The main contribution in this paper is the unified multi-platform analysis that classifies social media features into four categories. We hope that this cross-platform analysis can help scholars in the study of political parties and their campaigns around the globe. We emphasize the necessity for continuous and rigorous research in this field for two reasons: to cope with recent changes in digital media that directly affect online political interaction, and to understand the emergence and rise to dominance of right-wing populist parties around the world.

## REFERENCES


[1] Kai Arzheimer. 2015. The AfD: Finally a Successful Right-Wing Populist Eurosceptic Party for Germany? *West European Politics* 38, 3 (2015), 535–556.

[2] Sitaram Asur, Bernardo A. Huberman, Gábor Szabó, and Chunyan Wang. 2011. Trends in Social Media : Persistence and Decay. *5th International AAAI Conference on Weblogs and Social Media* (2011).

[3] Justus Bender and Marvin Oppong. 2017. Frauke Petry und die Bots. *FAZ.* Retrieved March 01, 2019 from https://www.faz.net/aktuell/politik/digitaler-wahlkampf-frauke-petry-und-die-bots-14863763.html.

[4] Nicole Berbuir, Marcel Lewandowsky, and Jasmin Siri. 2016. The AfD and its Sympathisers: Finally a Right-Wing Populist Movement in Germany? *German Politics* 24 (Jan. 2016), 154–178.

[5] Knut Bergmann, Matthias Diermeier, and Judith Niehues. 2016. Die AfD - eine Partei der Besserverdiener? (2016).

[6] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

[7] Leticia Bode and Emily K. Vraga. 2018. Studying Politics Across Media. *Political Communication* 35, 1 (2018), 1–7.

[8] Michael Bossetta. 2018. The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly* 95, 2 (2018), 471–496.

[9] Oliver Decker, Alexander Yendell, Johannes Kiess, and Elmar Brähler. 2017. Polarisiert und radikalisiert? Medienmisstrauen und die Folgen für die Demokratie.

[10] Infratest dimap. 2017. WahlREPORT Bundestagswahl. Retrieved March 01, 2019 from https://www.infratest-dimap.de/umfragen-analysen/bundesweit/wahlreport-deutschland/2017/.

[11] Paul-Jasper Dittrich. 2017. Social networks and populism in the EU: Four things you should know.

[12] Julia Ebner. 2017. How Germany's far right took over Twitter - and tilted the election. *The Guardian.* Retrieved March 01, 2019 from https://www.theguardian.com/commentisfree/2017/sep/26/germany-far-right-election-afd-trolls.

[13] Thomas Eckerl and Oliver Hahn. 2018. Die Selfie-Seite der Macht: Instagram in der politischen Kommunikation in Deutschland. In *Strategische Politische Kommunikation im digitalen Wandel: Interdisziplinäre Perspektiven auf ein dynamisches Forschungsfeld*, Michael Oswald and Michael Johann (Eds.). Springer Fachmedien Wiesbaden, Wiesbaden, 237–257.

[14] Sven Engesser, Nicole Ernst, Frank Esser, and Florin Büchel. 2017. Populism and social media: How politicians spread a fragmented ideology. *Information,*

*Communication & Society* 20, 8 (2017), 1109–1126.
[15] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. 2013. Anger is More Influential Than Joy: Sentiment Correlation in Weibo. *arXiv preprint* arXiv:1309.2402 (2013).
[16] Emilio Ferrara, Onur Varol, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
[17] Kirill Filimonov, Uta Russmann, and Jakob Svensson. 2016. Picturing the Party: Instagram and Party Campaigning in the 2014 Swedish Elections. *Social Media + Society* 2, 3 (2016).
[18] Simon Franzmann. 2014. Die Wahlprogrammatik der AfD in vergleichender Perspektive. *Mitteilungen des Instituts für Deutsches und Internationales Parteienrecht und Parteienforschung* 20 (Jan. 2014), 115–124.
[19] Paolo Gerbaudo. 2018. Social media and populism: an elective affinity? *Media, Culture & Society* 40, 5 (2018), 745–753.
[20] Robert Gorwa and Douglas Guilbeault. 2018. Unpacking the Social Media Bot: A Typology to Guide Research and Policy. *Policy & Internet* (2018).
[21] T. L. Griffiths and M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences* 101, Suppl. 1 (April 2004), 5228–5235.
[22] Robert Grimm. 2015. The rise of the German Eurosceptic party Alternative für Deutschland, between ordoliberal critique and popular anxiety. *International Political Science Review* 36, 3 (2015), 264–278.
[23] Simon Hegelich. 2018. *Die digitale Revolution der Meinungsbildung Social Media im Wahlkampf.* Hans Seidel Stiftung, Munich, Germany.
[24] Simon Hegelich and Morteza Shahrezaye. 2015. The Communication Behavior of German MPs on Twitter: Preaching to the Converted and Attacking Opponents. *European Policy Analysis* 1 (2015), 155–174.
[25] Marc Helbling and Anke Tresch. 2011. Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies* 30, 1 (2011), 174–183.
[26] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13, 1 (2016), 72–91.
[27] Bente Kalsnes. 2016. The social media paradox explained: comparing political parties' Facebook strategy versus practice. *Social Media+ Society* 2, 2 (2016).
[28] David Karpf. 2017. *Analytic Activism: Digital Listening and the New Political Strategy.* Oxford University Press, Oxford, UK.
[29] Majid Khosravinik. 2017. Right Wing Populism in the West: Social Media Discourse and Echo Chambers. *Insight Turkey* 19 (Jan. 2017).
[30] Juho Kim. 2018. The radical market-oriented policies of the Alternative for Germany (AfD) and support from non-beneficiary groups – discrepancies between the party's policies and its supporters. *Asian Journal of German and European Studies* 3, 1 (21 May 2018), 6.
[31] Anders O Larsson and Bente Kalsnes. 2014. 'Of course we are on Facebook': Use and non-use of social media among Swedish and Norwegian politicians. *European Journal of Communication* 29, 6 (2014), 653–667.
[32] Kristina Lerman and Rumi Ghosh. 2010. Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. *arXiv preprint* arXiv:1003.2664 (2010).
[33] Azi Lev-On. 2012. YouTube Usage in Low-Visibility Political Campaigns. *Journal of Information Technology & Politics* 9, 2 (2012), 205–216.
[34] Viktor Mayer-Schnberger and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think.* John Murray Publishers, UK.
[35] Gianpietro Mazzoleni and Roberta Bracciale. 2015. Socially mediated populism: the communicative strategies of political leaders on Facebook. *Palgrave Communications* 4, 50 (2015).
[36] Juan Carlos Medina Serrano, Simon Hegelich, Morteza Shahrezaye, and Orestis Papakyriakopoulos. 2018. *Social Media Report: The 2017 German Federal Elections.* TUM University Press, Munich, Germany.
[37] Yves Meny and Yves Surel. 2002. *Democracies and the Populist Challenge.* Palgrave Macmillan UK.
[38] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M Carley. 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *ICWSM.* Proceedings of the seventh International Conference on Weblogs and Social Media, Cambridge, Massachusetts.
[39] Caroline Lego Munoz and Terri L. Towner. 2017. The Image is the Message: Instagram Marketing and the 2016 Presidential Primary Season. *Journal of Political Marketing* 16, 3-4 (2017), 290–318.
[40] Christian Nestler and Jan Rohgal. 2014. Eine deutsche Angst - Erfolgreiche Parteien rechts von der Union: Zur AfD und den gegenwärtigen Gelegenheitsstrukturen des Parteienwettbewerbs. *Zeitschrift für Politik* 61, 4 (Dec. 2014), 389–413.
[41] Lisa-Maria Neudert, Bence Kollanyi, and Philip N. Howard. 2017. Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter?
[42] Oskar Niedermayer and Jürgen Hofrichter. 2016. Die Wählerschaft der AfD: Wer ist sie, woher kommt sie und wie weit rechts steht sie? *Zeitschrift für Parlamentsfragen* 47 (Jan. 2016), 267–285.
[43] Pippa Norris. 2003. Preaching to the Converted?: Pluralism, Participation and Party Websites. *Party Politics* 9, 1 (2003), 21–45.
[44] Karoline Oelsner and Linette Heimrich. 2015. Social media use of German politicians: towards dialogic voter relations? *German Politics* 24, 4 (2015), 451–468.
[45] Orestis Papakyriakopoulos, Simon Hegelich, Morteza Shahrezaye, and Juan Carlos Medina Serrano. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society* 5, 2 (2018).
[46] Pew Research Center: Global Attitudes & Trends. 2018. News Media and Political Attitudes in Germany. Retrieved March 01, 2019 from https://www.pewglobal.org/fact-sheet/news-media-and-political-attitudes-in-germany/.
[47] Pew Research Center: Internet & Technology. 2018. Teens, Social Media & Technology 2018. Retrieved March 01, 2019 from https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/.
[48] Neil Postman. 1970. The reformed English curriculum. In *High school 1980: The shape of the future in American secondary education*, Alvin C. Eurich (Ed.). Pitman Publishing Corporation, New York, N.Y., 160–2168.
[49] Stephen A Rains and Steven R Brunner. 2015. What can we learn about social network sites by studying Facebook? A call and recommendations for research on social network sites. *New Media & Society* 17, 1 (2015), 114–131.
[50] Patricia Rossini, Jeff Hemsley, Sikana Tanupabrungsun, Feifei Zhang, and Jennifer Stromer-Galley. 2018. Social Media, Opinion Polls, and the Use of Persuasive Messages During the 2016 US Election Primaries. *Social Media + Society* 4, 3 (2018).
[51] Uta Russmann and Jakob Svensson. 2017. Interaction on Instagram?: Glimpses from the 2014 Swedish Elections. *IJEP* 8 (2017), 50–66.
[52] Sebastian Schelter, Felix Biessmann, Malisa Zobel, and Nedelina Teneva. 2016. Structural Patterns in the Rise of Germany's New Right on Facebook. In *ICDM Workshops.* IEEE Computer Society, Barcelona, Spain, 440–445.
[53] Franzisca Schmidt. 2018. Drivers of Populism: A Four-country Comparison of Party Communication in the Run-up to the 2014 European Parliament Elections. *Political Studies* 66, 2 (2018), 459–479.
[54] Rüdiger Schmitt-Beck. 2017. The 'Alternative für Deutschland in the Electorate': Between Single-Issue and Right-Wing Populist Party. *German Politics* 26, 1 (2017), 124–148.
[55] Statista. 2017. Ranking der beliebtesten Parteien bei Twitter nach der Anzahl der Follower in Deutschland im August 2017. Retrieved March 01, 2019 from https://de.statista.com/statistik/daten/studie/659207/umfrage/anzahl-der-twitter-follower-von-parteien-in-deutschland/.
[56] Stefan Stieglitz and Linh Dang-Xuan. 2013. Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining* 3, 4 (2013), 1277–1291.
[57] Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018. Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter. *Political Communication* 35, 1 (2018), 50–74.
[58] Jesper Strömbäck and Peter Van Aelst. 2013. Why political parties adapt to the media: Exploring the fourth dimension of mediatization. *International Communication Gazette* 75, 4 (2013), 341–358.
[59] Jennifer Stromer-Galley. 2014. *Presidential campaigning in the Internet age.* Oxford University Press, Oxford, U.K.
[60] Andree Thieltges, Orestis Papakyriakopoulos, Juan Carlos Medina Serrano, and Simon Hegelich. 2018. Effects of Social Bots in the Iran-Debate on Twitter. *arXiv preprint* arXiv:1805.10105 (2018).
[61] T.J. Thomson and Keith Greenwood. 2017. I "Like" That: Exploring the Characteristics That Promote Social Media Engagement With News Photographs. *Visual Communication Quarterly* 24, 4 (2017), 203–218.
[62] Anke Tresch, Jonas Lefevere, and Stefaan Walgrave. 2018. How parties' issue emphasis strategies vary across communication channels: The 2009 regional election campaign in Belgium. *Acta Politica* 53, 1 (2018), 25–47.
[63] Sonja Utz. 2009. The (potential) benefits of campaigning via social network sites. *Journal of computer-mediated communication* 14, 2 (2009), 221–243.
[64] José van Dijck and Thomas Poell. 2013. Understanding Social Media Logic. *Media and Communication* 1, 1 (2013), 2–14.
[65] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online Human-Bot Interactions: Detection, Estimation, and Characterization. *arXiv preprint* arXiv:1703.03107 (2017).
[66] Lucia Vesnic-Alujevic and Sofie Van Bauwel. 2014. YouTube: A Political Advertising Tool? A Case Study of the Use of YouTube in the Campaign for the European Parliament Elections. *Journal of Political Marketing* 13, 3 (2014), 195–212.
[67] Quang H. Vuong. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* 57, 2 (1989), 307–333.
[68] Markus Wagner and Thomas M Meyer. 2014. Which issues do parties emphasise? Salience strategies and party organisation in multiparty systems. *West European Politics* 37, 5 (2014), 1019–1045.
[69] Fang Wu and Bernardo A Huberman. 2007. Novelty and collective attention. *Proceedings of the National Academy of Sciences of the United States of America* 104, 45 (2007).

## 4.2 Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections

### Authors

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich

### In

International Conference on Social Media and Society. Association for Computing Machinery, New York, NY, USA, pages 111–121 (2020).
DOI: https://doi.org/10.1145/3400806.3400820

### Abstract

This study investigates the possibilities and limits presented by the newly created ad libraries from Facebook and Google to analyze online political campaigns. We selected Germany as a case study and focused on the months leading up to the 2019 elections to the European Parliament. We identified the political actors that were active advertisers, compared their spending, and contrasted the number of ad impressions with user engagement on their organic online content. From the political ads, we extracted the unique ads and manually analyzed a subsample of them. Furthermore, we explored regional and demographic distributions of users reached by the advertisements and used them as a proxy for the advertisers' targeting strategies. We also compared the success of the ad campaigns on boosted Facebook posts. We found that even though all the major German political parties engaged in online ad campaigns, they kept their attempts at microtargeting to a minimum. Although their Facebook-sponsored posts were more successful than normal posts, we did not find statistical significance for all the political parties. Interestingly, we noticed that the distribution of users reached by the right-wing party Alternative für Deutschland (AfD) diverges from that of the other parties. Finally, we discuss further challenges for enhancing transparency in online advertising.

### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections

Juan Carlos Medina Serrano
juan.medina@tum.de
Technical University of Munich
Munich

Orestis Papakyriakopoulos
Technical University of Munich
Munich, Bavaria
orestis.papakyriakopoulos@tum.de

Simon Hegelich
Technical University of Munich
Munich, Bavaria
simon.hegelich@hfp.tum.de

## ABSTRACT

This study investigates the possibilities and limits presented by the newly created ad libraries from Facebook and Google to analyze online political campaigns. We selected Germany as a case study and focused on the months leading up to the 2019 elections to the European Parliament. We identified the political actors that were active advertisers, compared their spending, and contrasted the number of ad impressions with user engagement on their organic online content. From the political ads, we extracted the unique ads and manually analyzed a subsample of them. Furthermore, we explored regional and demographic distributions of users reached by the advertisements and used them as a proxy for the advertisers' targeting strategies. We also compared the success of the ad campaigns on boosted Facebook posts. We found that even though all the major German political parties engaged in online ad campaigns, they kept their attempts at microtargeting to a minimum. Although their Facebook-sponsored posts were more successful than normal posts, we did not find statistical significance for all the political parties. Interestingly, we noticed that the distribution of users reached by the right-wing party Alternative für Deutschland (AfD) diverges from that of the other parties. Finally, we discuss further challenges for enhancing transparency in online advertising.

## CCS CONCEPTS

• **Information systems → Display advertising**; **Social networks**; • **Human-centered computing → Social network analysis**.

## KEYWORDS

political campaigns, ad libraries, social media, European elections, Facebook, Google, transparency

**ACM Reference Format:**
Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections. In *International Conference on Social Media and Society (SMSociety '20), July 22–24, 2020, Toronto, ON, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3400806.3400820


*SMSociety '20, July 22–24, 2020, Toronto, ON, Canada*

ACM ISBN 978-1-4503-7688-4/20/07…$15.00
https://doi.org/10.1145/3400806.3400820

## 1 INTRODUCTION

The creation of ad libraries represents the largest effort to date to introduce transparency in online advertising. Ad libraries are collections of online ads with corresponding information on who funded them, how much was spent, and general information on the users that saw them. Ad libraries can be supported with application programming interfaces (APIs), which provide a direct channel to extract information. Before ad libraries existed, online political ads were classified as dark ads as they were publicly inaccessible [44]. This enabled astute malicious actors to target susceptible groups. Online political manipulation was not thoroughly considered in the early years of the online platforms. However, there was a turning point after the 2016 US presidential election with the Cambridge Analytica scandal. The company was charged with sending customized ad messages based on psychographic profiles of US citizens using private data from Facebook [72]. Further reports on data misuse and misinformation campaigns on Facebook and other online platforms followed. This included divisive and polarizing Facebook ads from the Russian Internet Research Agency, which were intended to incite social conflict and polarize society in the US [62]. In response to these scandals and escalating international pressure, Facebook decided to increase transparency in online advertisement. The social media company created an ad archive report in the weeks before the 2018 US midterm elections and announced the first ad library API. Facebook made the ad library public in March 2019. Google and Twitter followed and launched their ad transparency libraries a few months before the 2019 European elections.

The elections for the European Parliament were held between May 23 and 26, 2019. They constituted the ninth elections since the first parliamentary elections in 1979. Twenty-eight countries participated in the process of selecting 751 Members of the European Parliament. For this study, we focused on Germany, the member state with the largest population. The case of Germany is of interest since 1) the country has strong data privacy laws [23]; 2) Facebook, Google, Instagram, and YouTube are still the primary channels for online political ads [21], in contrast to other countries where political actors also use messaging apps [61]; and 3) Germany has a far-right populist party that behaves differently than the other German parties [63].

The purpose of this paper is to show the extent to which Facebook's and Google's ad libraries can help in understanding online campaigns. For this purpose, we quantify the online advertising campaigns of the major German political parties in the months leading up to the 2019 European elections. We first compare political party spending and the differences between ad impressions and user interactions on organic content. Organic content refers to the content that was shared on Facebook pages or YouTube channels but was not promoted by ad campaigns. Second, we explore the number of unique ads per political party and give a general overview of the content of the ad. We then try to understand political parties' targeting strategies using the demographic and regional distributions of users reached by the advertisements. Finally, we quantify the ad campaigns' success by comparing user engagement between sponsored posts and organic posts on Facebook.

Studies by Edelson et al. [24] and Ghosh et al. [33] presented a first analysis of the ad libraries. They analyzed the different types of political advertisers in the US in 2018 and aimed to understand the targeting features that the advertisers used. However, neither study focused on the political parties, the content of the ads, or user interactions. Moreover, the studies explored the libraries when they were still in beta testing. This paper takes a further step toward the establishment of methodologies to analyze online political advertising with the help of the ad libraries in combination with additional data sources. This makes it possible to obtain a detailed overview of online political campaigns during election periods.

# 2 BACKGROUND

## 2.1 Online Advertising

With the widespread use of digital services, online advertisement is becoming increasingly ubiquitous. Worldwide spending on digital marketing is predicted to reach $333 billion in 2019, which accounts for roughly half of the global ad market [25]. There are numerous types of online advertising such as search advertising, social media advertising, and email advertising [35]. For all types of advertising, the business model is based on continuous data collection of personal information, monitoring of individuals' online behavioral patterns, and pursuit of customer manipulation. This new form of business model is at the core of what has been referred to as the "digital surveillance economy" [19]. Its main innovation relies on targeting users at an unprecedented granular level. Microtargeting is defined as the creation of customized messages delivered to groups or individuals that are predicted to be impacted by those messages [1]. It relies on the full gamut of big data algorithms to define differentiable and profitable customer groups [58]. Apart from targeting, online advertisement allows for increased measurability, since every user's response to ads can be easily tracked [35].

Social media platforms and search engines create user profiles by collecting as much information as possible on their users' online activity [9] from their interactions, such as likes, dislikes, comments, and connections, [6, 17] to their browsing behaviors via cookies [46]. With the collected data, advertisers are provided with fine-grained targeting features, including personally-identifiable information (PII), such as phone numbers and email addresses [71]. Facebook allows targeting of users' behavior, location, interests, demographics, and connections [26], and Google allows targeting of their website visits, interests, locations, and demographics [37]. After advertisers select a target audience, a matching mechanism decides which ads to deliver to a given user. There is a real-time auction to display an ad to a user, where the ad of the higher bidder is chosen [32].

A controversial category of online advertisement is political advertisement, given the possible influence of political ads on public opinion. In the last few years, political campaigns have increasingly embraced data-driven targeting strategies [39, 48, 57]. Online political advertisement as the main vehicle of these targeting strategies has reshaped modern-day politics [18]. Political actors are now able to identify and reach individual users who are more likely to be persuaded by them and to match their messages to the users' specific interests and vulnerabilities [74]. Narrower targeting allows politicians to tailor advertising on wedge issues, such as immigration and abortion, to sway pivotal voters [40]. A data-driven online campaign makes microtargeting a powerful technique and is currently the state of the art in political campaigning. In the US, political parties have made this possible by amassing a huge amount of personal data on voters' political affiliations and behavior [10]. Despite having the strictest data privacy regulations, Germany is also a fertile ground for political microtargeting [59]. Apart from political parties, other advertisers also activate ads with political intent. It is, however, not a straightforward task to define the borders of what constitutes a political ad. Facebook and Google have implemented different policies to decide which ads are added to the ad libraries. Facebook's archive includes all advertisements that have political content, are regulated as political advertising or are related to a national legislative issue. On the other hand, Google only includes advertisements either created by political actors or related to specific elections [24].

Few studies have examined the breadth of political parties' online advertisement, largely because of the unavailability of relevant datasets. However, there is extensive social science literature on political advertising—from studies that focus on analyzing the content of political ad campaigns [5, 66, 70] to research experiments that attempt to determine if political advertising is successful in persuading voters [31, 34, 51]. These studies agree that political advertising has the potential to persuade, but its real impact depends on the user characteristics. In this way, political advertising can both influence users' vote choice and prompt mobilization on election day [54]. Facebook conducted a randomized controlled trial of political mobilization messages delivered to 61 million users, which resulted in an increased turnout [14]. It is possible that the political advertising on social media platforms can motivate similar effects.

## 2.2 Data Privacy and Transparency

With the rise of online advertisement and microtargeting, serious concerns have been raised about data privacy and transparency [30, 46, 69, 74]. Prior work has uncovered the use of sensitive data for Facebook advertisement targeting purposes [16]. For example, Venkatadri et al. [71] noted the possibility of targeting individuals with sensitive PII, such as phone numbers provided for security purposes and phone numbers derived from friends' contact lists. Furthermore, attackers can exploit the vulnerabilities of the advertising interfaces to breach user privacy [47] or to employ discriminatory targeting [22, 65]. On the subject of transparency, Kreiss and Mcgregor [49] consider that both Facebook and Google have been opaque in their decision-making, following policies that are not transparent and that were applied without explicit justification.

Facebook and Google have taken steps to address the aforementioned concerns. Both platforms have introduced an ads preference page where users can see and correct the information that was inferred about them [29, 36]. They have also introduced information explaining why each ad was displayed. However, Andreou et al. [3] found the explanations on Facebook to be incomplete and sometimes misleading. Additionally, Datta et al. [22] reported incomplete representations of user profiles on Google's ad settings. Apart from these design flaws, it is essential to consider the user perspective. One concern is that users may not be aware of the advertising transparency mechanisms. A 2019 Pew Report [41] found out that 74% of Facebook users in the US did not know that the platform maintains a list of their interests and traits to target them. A second concern is that users may not care about transparency in the usage of their private data. The well-studied privacy paradox suggests that even though users are concerned about privacy online, their behaviors do not reflect these concerns [7]. For instance, Baum et al. [8] found that the presence of political ads on mobile apps does not consequentially deter users from choosing such an app. In the case of Germany, the findings of Taddicken et al. [68] indicate that privacy concerns hardly have an effect on users' self-disclosure.

### 2.3 Ad Effectiveness

The field of ad effectiveness is explored extensively in the literature as it constitutes the main goal of a successful marketing campaign. Kingsnorth [45] presents three types of measurement metrics in online advertisement: traffic metrics, such as impressions and click-through rate (number of clicks/impressions); conversion metrics, in which a conversion refers to the number of users who clicked on the ad and then successfully triggered a specific action such as buying a product or subscribing to a newsletter; and efficiency metrics, with the most common one being the return on investment (ROI), which is the revenue or profitability divided by the ad cost. These measures are connected to business value when the advertisement is directly linked to a physical product or a service. However, this is not the case when the ad's purpose is to raise customer awareness, such as promoting a brand or a political candidate [42]. Calculating a "political" ROI is challenging given that it is hard to find effective and standardized metrics for measuring the impact of online political advertising. Previous political campaigns in the US have used as metrics the number of dollars fundraised after a user clicked on a search ad, and the number of ad volunteer sign-ups gathered as a result of ad click-throughs [20]. However, this information is only available to the campaigns, which does not allow a comparison between different political parties. A useful set of metrics relies on user engagement. User engagement is defined as the quality of the user experience that emphasizes the positive aspects of the interaction and in particular what motivates a user to interact with a web application [52]. User engagement metrics differ from each online platform depending on their design. They range from the number of unique users, click-through rates, page views, and time spent on a website. A subclass of user engagement refers to online behavior metrics [52]. For social media, these include direct user interactions with the advertising posts, such as likes, comments, and shares. These metrics have already been used as a measure of success in literature. Poecze et al. [60] analyzed social media metrics to evaluate the effectiveness of YouTube videos. Lee et al. [50] and Jaakonmäki et al. [43] studied the relationship between ad content and user engagement to measure the effectiveness of social media content marketing campaigns on Facebook and Instagram, respectively. Additionally, previous research has found a positive relationship between user interactions and election results [11, 38]. This suggests that user interactions could also be a proxy for ad effectiveness. Unfortunately, there is no clear methodology to measure how online behaviors translate into offline actions. Research has found a strong connection between online user engagement and behavioral intentions, and simultaneously between the expression of intention and message-induced offline behaviors [2, 73]. Unsurprisingly, the online media platform's algorithms are programmed to optimize user engagement [15]. User engagement allows the collection of vast amounts of data. This contrast to the relatively scarce data on voting behavior as elections are held once or twice every two years. Nevertheless, it remains a challenge to correctly quantify ad effectiveness. With the extensive datafication of politics, new metrics could appear in the near future.

## 3 METHODOLOGY

### 3.1 Data Collection

We collected data from political ads in Germany using the newly created Facebook Ad Library API and the Google Cloud BigQuery API. The former includes ads shown both on Facebook and Instagram, whereas the latter has ads that appeared on Google Search and YouTube. We employed a daily procedure to collect data from both APIs to avoid missing ads that could be later removed. Moreover, this allowed us to circumvent the technical difficulties and bugs that appeared after a few days during the collection dates [55] and collect all the ads that targeted Germany. The collection period we selected began with the creation day for each API (beginning and middle of March 2019, for Facebook and Google respectively), and ended one week after the European elections. In total, the dataset comprises 50,794 Facebook ads and 34,197 Google ads. The content of the ads is often not unique since ad campaigns constantly deploy similar ads, which either contain a small modification, have a different target strategy, or are distributed on different platforms. However, each ad has a unique ID, which allows us to avoid duplicates in the daily collection process.

We also gathered the content and interactions from the Facebook posts and YouTube videos published by the main political parties in Germany. These are the Christian Democratic Union (CDU), the main center-right party; the Christian Social Union (CSU), the sister party of the CDU operating only in the state of Bavaria; the AfD, the far-right party; Bündnis/Die Grünen, the green party; the Free Democratic Party (FDP), a neoliberal centrist party; the Social Democratic Party of Germany (SPD); and Die Linke, the left party. For each party, we further obtained the data for the regional pages and channels from the 16 German states.

To collect data from Facebook and YouTube, we used CrowdTangle and the YouTube Data API, respectively. The collection period corresponds to the same dates we selected for the ads. The dataset consists of 11,496 Facebook posts and 1,059 YouTube videos. Unfortunately, we were not able to collect data from Instagram, given that most of the capabilities from its public API are deprecated. In addition, we exclude Twitter ads as we did not find political ads from national or regional German party accounts on the Twitter Transparency Center.

## 3.2 Content Analysis

The first step in our analysis was to manually match the advertisers on Facebook and Google to political parties. Each political party has a main advertiser and other smaller advertisers connected to accounts of German states, cities, and local politicians. We placed the advertisers that did not belong to a political party into the following categories: government; companies and organizations, which include NGOs and labor unions; and others.

Our analysis mainly focuses on the advertisement created by the main and regional advertisers for the aforementioned German political parties. From the ads created by these advertisers, we filtered out duplicate ads. This was not a trivial task since the APIs do not provide a direct way to compare images or videos. Identical images that belong to different ads are stored as copies in the ad archives and can only be compared visually. For this reason, we crawled the ad libraries with the help of Selenium and downloaded the images. It was not necessary to download the videos as the unique URL link (for Google, a link to a YouTube video) was contained in the HTML of the website.

To find unique advertisements, we used string matching between texts and the *Perceptual Hashing* algorithm [53] to compare images. This algorithm creates a unique hash for an image. Hashes can be then compared using the Hamming distance. Since advertisers often use the same image in different sizes or pixel qualities, the hashes are not completely identical. After several tries, we considered images similar if the Hamming distance between the respective hashes was smaller than 20. After identifying the list of unique ads, we analyzed a random sample of 30 ads per party and per platform to give us a general overview of the advertising content and party intentions. Our focus was to evaluate whether the ad content was personalized or not. Defining personalization is not straightforward as it has a complete taxonomy [12]. For us, personalization is correlated to the content breadth of messages spread by a political party. An ad campaign with similar messages and scarce content would be classified as not personalized.

## 3.3 Demographic and Regional Strategies

The Facebook advertisement data contains the demographic and regional distribution of the users reached by each advertisement. These distributions depend on the targeting distribution that the advertisers selected but are not the same. They also depend on Facebook's algorithm and the users who are active at a given time. However, since we do not have access to the specific targeting strategies, we take the distributions of reached users as proxy variables for the original targeting strategies. We consider this limitation in the discussion section.

The demographic distribution is divided according to gender (including unknown) and six age ranges, whereas the regional distribution corresponds to the 16 German states. We first averaged the regional and demographic distributions for all ads belonging to a given party. We then computed the correlation between each pair of political parties. We subsequently averaged the correlations for each party separately. This procedure resulted in a number between -1 and 1 that conveys the similarity between a party's strategy and the remaining parties. The higher the score, the more similar are the targeting distributions. A low score denotes dissimilarities to the rest of the parties or in the case of a negative value, opposite strategies. We excluded the CSU for the regional distribution as it is only represented in one of the German states.

Additionally, we compared the regional distribution with the electoral support in each state. We obtained the percentage of voter intention by opinion polls conducted between February and May 2019[1]. For each party, we divided the percentage of approval for the German states by the sum of the percentages to normalize the data and obtain a variable from 0 to 1 that represents state popularity. We then calculated the correlation between state popularity and the mean regional distribution. A positive correlation would indicate that a party targeted states with high voter support. On the other hand, a negative correlation would relate to a strategy targeting states with lower voter support. We also excluded the CSU from this analysis.

## 3.4 Campaign Performance

We seek to quantify the success of the German parties' ad campaigns. The success of an ad campaign can be measured through the performance of the ads, either by click-rate or by the extent of user engagement with the ad. We use the number of total interactions to measure success as the click-rate information is only available to the advertisers. It is impossible to quantify real success, in other words, if the users were politically persuaded, but analyzing user engagement is a common practice in marketing to estimate success.

The core idea of the method is to investigate whether increased sponsorships have a significant effect on the number of user interactions. The level of sponsorship can be quantified with the number of impressions as these two quantities correlate. However, they are not completely equivalent given that advertisers can decide to pay for either clicks or impressions. According to Facebook documentation [27], impressions correspond to the number of times the ads were on-screen for the target audience. This means that if an ad is shown twice to the same user, it counts as two impressions. Similarly, Facebook interactions are not unique since one user can comment, share, and choose one of six reactions (like, wow, angry, haha, love, and sad) for a given post.

Facebook divides advertisements into boosted posts and regular ads [28]. Boosted posts are already existing posts on a public Facebook page, whereas regular ads are created separately from scratch. To test performance, we focus only on boosted posts since it is possible to compare user engagement between sponsored and non-sponsored posts. The Facebook Ad Library API neither differentiates between the ad type nor provides the number of interactions per ad. To obtain this information, we matched the ad data with the collected posts from the Facebook pages. We used the text description and the image or video of the ads and posts to automatically match between them. Only the ads belonging to boosted posts had

[1] https://www.wahlrecht.de/umfragen/landtage/index.htm Retrieved May 27, 2019.

a direct match. In total, the seven German parties boosted 522 posts from their national and regional Facebook pages.

One limitation to the ads data is that the exact number of impressions is unknown. For each ad, the API only reports a lower and an upper bound. These bounds belong to predefined ranges by Facebook. Only the advertiser who launched the ads can access the real values for impressions. We use the middle value between the lower and upper bound as an approximation to the real value. Additionally, each post can be sponsored by more than one ad. Therefore, we define *total mean interaction* as the sum of the middle values of the ads that sponsor a given post.

To compare between sponsored and non-sponsored posts while considering the level of sponsorship, we divide the posts into four groups; not sponsored, low-level sponsor, middle-level sponsor, and high-level sponsor. We define the level boundaries by grouping the 522 sponsored posts into three even groups ordered according to the ads' total mean interaction.

The distributions of social media interactions are generally non-normal, and most of them follow a log-normal distribution [4]. By plotting the distribution of the total mean interactions, we confirm that the non-normality holds. Consequently, we perform Kruskal-Wallis tests, which are the non-parametric version of the one-way ANOVA. We then apply post-hoc pairwise Wilcox tests with Bonferroni corrections to compare between the non-sponsored group and each of the sponsored groups. Given that the dependent variable has intrinsic errors that we are not able to control, we set a stricter significance level of p <0.01.

One limitation of this approach is that it neither controls for the time of posting nor for post content. A complete controlled experiment would compare posts with similar content that were launched at the same point in time. However, the purpose of our analysis is only to test the general performance of the parties' ad campaigns and obtain an indicator of overall success.

Although a similar analysis on Google ads can be performed on sponsored YouTube videos, the dataset contains only 23 videos. The rest of the ads correspond exclusively to ad videos, or to another ad category on Google, such as text or image. We present the Kruskal-Wallis test results of the 23 videos combined, without distinguishing between parties. For YouTube, the number of total interactions consists of views, comments, likes, and dislikes from videos posted by the political parties.

# 4 RESULTS

## Ad Campaigns

The German political parties implemented ad campaigns on the platforms belonging to Facebook and Google. Table 1 shows the number of ads and the spending in Euros for the parties per platform. For Facebook, we differentiate between the federal and state advertisers, which we refer to as main advertisers, and the local advertisers, which correspond to cities and individual politicians. From the main advertisers, the CDU spent more money and activated more ads than the other parties. The SPD comes in a close second place. However, when the local advertisers are included, the SPD comes in first place. On the other hand, the CSU, AfD and Die Linke were the parties that allocated less money on ads from both Facebook categories and Google advertisers. The CSU is only active in the state of Bavaria and had no Google ads during the observed period. Although the period for the Google ads is two weeks shorter than the Facebook ads period, it is evident that most of the parties spent more money on the latter.

**Table 1: Facebook ads and spending in the period from March 1 to June 2, 2019 and Google ads from March 20 to June 2, 2019. The Facebook ads are divided into two categories: by country-wide and regional political pages, and city and individual political pages. Facebook political pages that spent less than 100 Euros are not included.**

| | Facebook | | Facebook (Local) | | Google | |
|---|---|---|---|---|---|---|
| | Ads | Spending | Ads | Spending | Ads | Spending |
| *AfD* | 48 | 22,278 | 255 | 9,029 | 17 | 23,400 |
| *CDU* | 17,449 | 296,801 | 1,755 | 53,141 | 33,120 | 261,200 |
| *CSU* | 27 | 60,816 | 66 | 6,611 | 0 | 0 |
| *FDP* | 5,456 | 138,762 | 970 | 60,691 | 259 | 32,600 |
| *Die Grünen* | 7,804 | 229,451 | 1,888 | 60,283 | 769 | 140,750 |
| *Die Linke* | 958 | 41,526 | 359 | 7,259 | 7 | 3,200 |
| *SPD* | 15,234 | 283,664 | 3,174 | 130,692 | 90 | 133,900 |

The Facebook local advertisers spent less money than the main advertisers for the seven parties. However, the amount spent by the local advertisers is still significant. We do not show the local advertisers on Google as we only found three city advertisers and each spent less than 2,000 Euros in the investigated period. The results that follow concentrate only on the main advertisers per party.

## User Interactions

The party spending translates into ad impressions, and then the impressions into clicks. Table 2 presents the upper and lower bound of total impressions from all the ads belonging to a given party. We opted to show the one percent of the bounds since click-through rates on Facebook are in this order of magnitude [56]. This is a rough approximation of users interacting with the ads, and the values' order of magnitude is more significant than the exact number. The impressions from Facebook include Instagram, and those from Google include YouTube. For comparison, we include the number of total interactions on Facebook posts and YouTube videos uploaded in the same period. These correspond to organic interactions on the parties' social media channels. For the AfD, Die Grünen and Die Linke, organic content exceeds the one percent upper bound of the number of impressions. This is not the case for the FDP and the SPD. Nevertheless, these interactions do not take into account the missing Instagram and Google search data. A comparison between the political parties reveals that the AfD has more interactions than the other German parties.

## Unique Ads

The perception of the number of ads created by the political parties changes when we take into consideration only the unique ads. The German political parties created little advertising content as

**Table 2: Lower and upper bound of impressions generated by ads. The 1% represents the order of magnitude of click-through rates. For Facebook, the interactions correspond to likes, shares, comments, and reactions. For YouTube, the interactions consist of views, comments, likes, and dislikes.**

| | Facebook/Instagram Impressions | | Facebook | Google/YouTube Impressions | | YouTube |
|---|---|---|---|---|---|---|
| | Lower (1%) | Upper (1%) | Interactions | Lower (1%) | Upper (1%) | Interactions |
| *AfD* | 17,600 | 39,019 | 2,632,650 | 27,600 | 276,200 | 5,674,531 |
| *CDU* | 143,780 | 639,673 | 256,493 | 148,600 | 4,659,300 | 3,412,524 |
| *CSU* | 23,750 | 55,109 | 197,422 | 0 | 0 | 42,041 |
| *FDP* | 169,850 | 576,462 | 268,756 | 74,200 | 751,200 | 260,839 |
| *Grünen* | 141,900 | 474,408 | 262,415 | 49,900 | 559,200 | 2,387,547 |
| *Linke* | 59,050 | 158,404 | 253,981 | 11,100 | 111,400 | 348,821 |
| *SPD* | 177,540 | 638,163 | 245,865 | 32,900 | 332,100 | 288,682 |

illustrated in Table 3. The table is divided by type of advertising. For Facebook, there are images, videos and events, and for Google there are images, YouTube videos and ads with only text that appear on Google searches. Even though the CDU activated by far the largest number of ads, its unique number of ads is distinctively small, for example, only six Google images and three YouTube videos. The FDP, Die Grünen and the SPD created more content for their campaigns. These results show that the political parties were mostly activating the same ads repeatedly during the months leading up to the 2019 European elections.

**Table 3: The number of unique ads and the total number of ads per political party. The ads are divided by their type and platform. The number of unique ads is presented in bold. For Facebook, some ads were not included in the three categories, were not found in the Ad Archive anymore, or were duplicate ads from the advertisers. For this reason, the exact number may differ from Table 1.**

| | Facebook | | | Google | | |
|---|---|---|---|---|---|---|
| | Image | Video | Event | Image | Video | Text |
| *AfD* | **19**/22 | **17**/17 | **2**/2 | **1**/1 | **6**/7 | **3**/9 |
| *CDU* | **34**/311 | **24**/12,316 | **4**/7 | **6**/28,139 | **3**/4,981 | /0 |
| *CSU* | **12**/17 | **7**/10 | /0 | /0 | /0 | /0 |
| *FDP* | **394**/3,991 | **68**/722 | **33**/58 | **27**/178 | **12**/80 | /0 |
| *Die Grünen* | **134**/1,606 | **78**/2,448 | **35**/53 | **3**/33 | **9**/63 | **215**/609 |
| *Die Linke* | **67**/111 | **41**/234 | **54**/167 | **1**/1 | **2**/2 | **4**/4 |
| *SPD* | **223**/7,748 | **66**/3,826 | **34**/75 | **1**/1 | **11**/11 | **54**/78 |

After inspecting a random batch of 30 ads per party and per platform, we found that many unique ads are similar to each other. The parties did not diversify their content and adopted similar content and format for their advertising. As an example, Figure 1 shows one representative ad for three of the political parties. From the ads we observed, we noticed that the CDU created fewer personalized messages than the other parties. Even though the AfD activated few ads, their advertising was more personalized by using long texts with several URLs included. The FDP, Die Grünen and the SPD concentrated on their main campaign proposals, and they presented a few ideas outside of the topics that distinguish these political parties. Interestingly, the FDP created several ads in languages other than German, for example, Russian and Greek, to try to reach different parts of the electorate.

## User Distributions and Polls

Figure 2 compares the regional and demographic distributions of users reached by the targeting parties. We compare the parties by providing a mean correlation value. The SPD has the highest value for the regional distribution, which means it is the most similar to the other parties. By contrast, the AfD has a correlation value close to zero, which corresponds to the party with the most dissimilar regional pattern. The difference is even larger for the demographic distribution, where the AfD's correlation value is negative. This represents a completely opposite demographic pattern to the audience who perceived the AfD's ads. The SPD also has a lower mean correlation value than the other parties, whereas the FDP has the highest value.

The left image from Figure 2 corresponds to the correlation between opinion poll values and the mean regional distribution. We observe that on average the AfD, SPD and Linke targeted and reached users in states with lower voter support. On the other hand, the CDU, Die Grünen and FDP targeted and reached on average the states where voter support was higher. The average correlation value shows a general tendency for the political parties. This tendency is hard to observe by only analyzing the individual regional distributions per ad as all parties targeted repeatedly the 16 German states.

## Effectiveness of Sponsored Posts

For the Facebook posts, we compare the relationship between impressions and the number of total interactions as a proxy of an ad campaign's success. For each political party, Table 4 shows the ratio between the mean value of total interactions from sponsored posts and the mean value from non-sponsored posts. The FDP has


**Figure 1: Example of three ads during the campaign. Right: FDP image on Google. Middle: CDU image on Facebook. Left: AfD video on Facebook.**

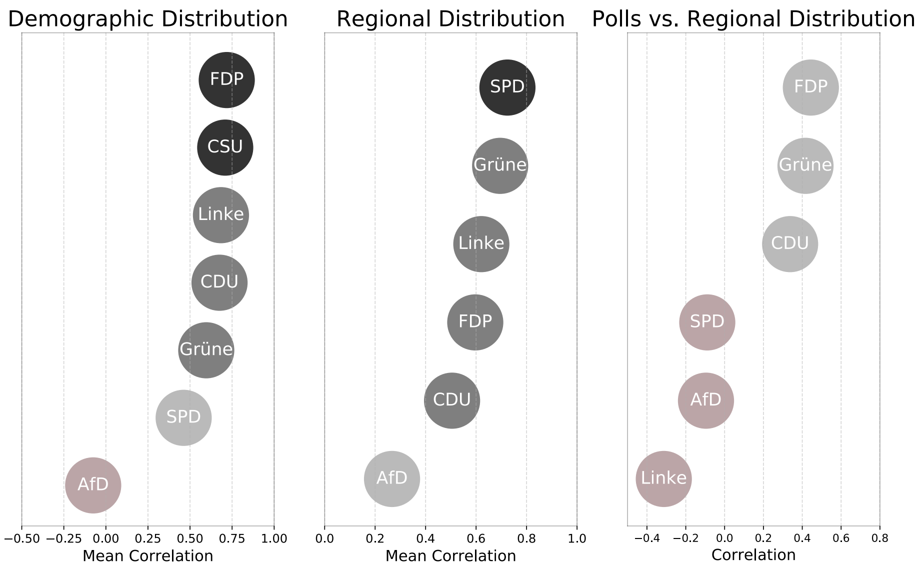


**Figure 2: Left and Center: Demographic and regional distributions of users reached by Facebook ads. The mean correlation value corresponds to overall distribution similarity. It is an average value of correlations between the ad distributions for one party and each of the rest. Right: Correlation between voter support from opinion polls and mean regional distribution.**

the biggest ratio value, whereas the CDU and CSU have the lowest values. On average, the sponsored posts from all the political parties had more interactions than the non-sponsored posts. However, this metric is not enough to show that there was a significant improvement given that it does not take into account the distributions of the interactions on the posts. Table 4 further presents the Kruskal-Wallis tests between four groups of posts: non-sponsored, low-sponsor, middle-sponsor and high-sponsor. The table also reports the results from the post-hoc pairwise Wilcoxon tests between the non-sponsored posts and each of the sponsored groups. For the CDU and CSU, there are no significant differences between groups, which is further confirmed by the three pairwise Wilcoxon tests. Even though there is significance in the group differences for Die Grünen and Die Linke, the pairwise tests reveal that the difference is only significant for highly sponsored ads. Finally, for the AfD, FDP, and SPD, the Kruskal-Wallis tests are significant as are the post-hoc tests between the non-sponsored posts and the other three groups. In the case of YouTube, we perform the statistical tests on the videos of the parties together because of the limited dataset of sponsored videos. We find a significant difference between the groups of videos ($p<.001$).

## Ad Categories

As a final analysis on the Facebook ad archive, we present the comparison between advertisers belonging to different categories. Given that Facebook considers advertisements related to both political issues and topics, there are other advertisers apart from the political parties in the archive. Table 5 shows the percentage of ads and total spending from the advertisers of the four categories. We observe that although 86% of the ads were created by the political parties, only 50% of the total spending can be credited to them. Private companies, NGOs, and labor unions represent 25% of the total ad spending on the social media platform. Further analysis of

**Table 4: The ratio between the means of total interactions from sponsored posts and non-sponsored posts. Additionally, statistical tests between the interactions of non-sponsored posts and sponsored posts divided in low, middle and high level of sponsorship. The p-value corresponds to the Kruskal-Wallis tests for the difference between the groups. In parenthesis, the post-hoc pairwise Wilcoxon tests with Bonferroni correction between non-sponsored posts and each of the other three groups (low, middle, high). Significance codes: *** p<0.001, ** p<.005, * p<0.01, • p>0.01**

| | Ratio of means | K-W tests | Pairwise W tests (low, middle, high) |
|---|---|---|---|
| *AfD* | 2.45 | p<.001 | (****,*,**) |
| *CDU* | 1.52 | p=.17 | (•,•,•) |
| *CSU* | 1.88 | p=.039 | (•,•,•) |
| *FDP* | 4.23 | p<.001 | (***,***,***) |
| *Die Grünen* | 2.19 | p<.001 | (•,•,***) |
| *Die Linke* | 2.84 | p<.001 | (•,•,***) |
| *SPD* | 3.49 | p<.001 | (***,**,***) |

**Table 5: Percentage of number of ads and spending from the advertising on Facebook. Advertisers are divided into four categories.**

| Category | Ads (%) | Spending (%) |
|---|---|---|
| *political parties* | 86.17 | 48.23 |
| *government* | 0.41 | 20.81 |
| *companies/organizations* | 10.68 | 25.72 |
| *other* | 2.73 | 5.23 |

political ads from other advertisers than political parties is outside the scope of this paper.

## 5 DISCUSSION

For the first time, it is possible to scrutinize the digital marketing campaigns of political parties. This opens up the possibility of holding political actors accountable for their online marketing spending and the messages they promote. The provided APIs helped us in obtaining the data through a direct and automatic process that allowed us to frequently update our database. However, APIs could present technical challenges for some users. During the studied period, we observed the technical difficulties and bugs that have been previously discussed [55]. Nevertheless, we collected data several times a day to have the most complete dataset possible. From the information perspective, we find two main limitations in the ad libraries. First, the spending and impressions per ad are only available in broad ranges with different sizes. This reduces the statistical models that can be implemented using the data. A better technique would be for the companies to provide the original quantities with some random noise added. In this way, well-known models, such as linear regression would present more accurate results. A second limitation is that there is no information on the original user-targeting strategies from the advertisers. We understand that such information would reveal campaign strategies to competitors, but this shortcoming limits the transparency potential of online political advertising.

Our findings in the case of Germany show that the major political parties were actively using the online advertisement platforms. In the three months leading up to the 2019 European elections, the CDU had the biggest ad campaign, with the most money spent and the largest number of ads. The SPD and Die Grünen also had very active campaigns on both platforms. Between the 2017 German federal election and the 2019 European elections, these two parties saw their electoral results change considerably. The SPD's acceptance slipped from 20.5% to 15.6%, whereas Die Grünen's approval increased from 8% to 20% [67]. We avoid comparing election results with ad campaigns as it is impossible to quantify a political conversion rate.

The possible impact of the online political campaigns for the 2019 European elections reduces when we consider the low number of unique ads created by the German political parties. We conclude that few attempts were made at implementing microtargeting strategies as this technique requires personalized messages that target different parts of the electorate. However, it remains possible that the correct users were targeted and have to some extent been mobilized to vote. However, it is unfeasible to calculate the success rate of online advertising and user voting. From the content of the ads, we observe that the political parties in Germany mostly advertised general messages to the targeted users. It is possible that since the European elections are not designed to tackle domestic issues, the political parties did not focus on diversifying their online advertising. Manually inspecting a random sample of advertising gave us a general impression of each party's strategy. A thorough qualitative analysis of the advertising would require the examination of theoretical frameworks designed for understanding political advertising.

For all the parties, user engagement is higher on average on the promoted Facebook posts. However, for the CDU and CSU, the statistical tests do not allow us to conclude that a significant difference exists. This could indicate a general lack of user interest in interacting with their content or that the boosted posts were less attractive than the other posts. If we take the ad content into consideration, the explanation could be related to the lack of personalized ads created by the two parties. By contrast, even low-sponsored posts showed a significant increase in interactions for the AfD, SPD, and FDP. These parties' advertising content was also more diverse and personalized. It is important to emphasize that this analysis only focuses on boosted posts and excludes the remaining Facebook ads. Moreover, the real success can only be measured by comparing click-through rates between the political parties' ad campaigns, which are not publicly available data. We want to emphasize that we are not linking ad effectiveness to approval ratings, as political advertising may have a weak indirect effect on political campaigns. We only focus on the success of a political marketing campaign without implying anything in the effects on real user persuasion. Another limitation in our analysis is the missing original targeting strategies. The comparison between distributions of users reached by Facebook ads can be interpreted as a proxy of the real targeting by the parties. However, this could be inaccurate given that the bias of Facebook's algorithm and other external factors cannot be quantified.

Throughout this study, the AfD's results deviate from the rest. The AfD advertisers did not spend a substantial amount of money on ads and launched few ads. They mostly relied on organic channels to spread their message. The interactions on the AfD's Facebook pages and YouTube channels are accordingly the highest among the political parties. This is consistent with research on the AfD's success on social media [63]. Additionally, the AfD's demographic and regional distributions on Facebook differ from those of the other parties. This could suggest that the AfD is targeting a separate group of the population. This disjunction could be considered as a proxy for the polarization of German society. Indeed, Shahrezaye et al. [64] reported that on Facebook the polarization between the AfD and the other political parties has increased in the last few years.

Overall, this study has presented the new opportunities and limitations of the ad libraries in order to explore the online advertising. They represent the first iteration of valuable tools necessary to achieve complete transparency. Under the current scenario, we present four challenges that remain to be addressed.

The first challenge relates to the definition of a political ad. Until now, no EU regulations state what constitutes a political ad on digital channels. Without a clear formalization, the responsibility of flagging political ads resides solely with private companies. As mentioned previously, Google only includes political actors in its ad archive, in contrast to Facebook's political content policy for selecting political ads. Hypothetically, if the non-political advertisers on Facebook would have activated political ads through Google's platforms in the same proportion as they did on Facebook, over 10% of ads and 45% of spending would be unaccounted for in Google's ad library.

Given the large number of ads on both platforms, the companies most probably employ automatic classification mechanisms in the selection of political ads. The accuracy of these algorithms is not presented in their transparency reports. A second challenge is that the classification system's vulnerabilities can be exploited by malicious actors. They would try to stay undetected by sponsoring ads that may appear to have non-political content but that are deliberately intended to influence and even polarize public opinion.

The third challenge is that the increase in transparency can make political actors feel subject to scrutiny. This can, in turn, lead them to change their digital political strategies. For example, they could focus on private communication channels like WhatsApp. The end-to-end encryption allows the spread of any form of political content to stay undetected. Although WhatsApp is still an unexplored channel for German political actors, this is not the same in other parts of the world, as in the case of the misinformation scandals in the 2018 Brazilian election [13].

The final challenge is that whereas online platform companies have responded to the backlash of election manipulation, government legislation has not kept up with the pace. As a consequence, the companies can make the first proposals and thus set the rules that define the problems. In such a scenario, the government will have to adapt to the proposed rules instead of leading the efforts that foster more transparency and privacy. Further steps need to be taken by both private and government institutions to deal with these challenges. Private institutions should provide services that enhance transparency and respect user privacy, whereas governmental institutions should focus on creating new regulations and raising public awareness.

## 6 CONCLUSION

In this paper, we explored methods to quantify the data from Facebook's and Google's newly created ad libraries and help further understand online political campaigns. We chose to examine the advertising from political parties in Germany in the months leading up to the 2019 European elections. We focused on spending, impressions, and the possible targeting strategies of the major German parties. We concluded that during the months leading up to the election, the parties engaged in few microtargeting attempts as they created few unique ads, with many of them lacking personalized messages. In this study, we also compared the ads' data with the political parties' organic content on Facebook and YouTube. For a selected group of ads, we introduced a way to measure success even without the ads' click-through rates. The boosted posts on Facebook attracted more user engagement than organic posts, although we did not find significant differences for all German political parties. Our findings also show that the distribution of users reached by the AfD's advertising deviated from that of the other parties, which could imply a targeting strategy that was contrary to other political parties' strategies. Furthermore, the far-right party relied mostly on promoting their organic content and not so much on sponsored ads. Finally, we introduced the challenges that arise with the transparency of ad libraries. We hope that this work helps future scholars to explore the ad libraries in different countries to understand the global political online advertising environment. A lack of continuous analysis and auditing undermines the full potential of transparency tools. It is essential to undertake further research on how to profit from these tools to enhance transparency and prevent malicious activity on online advertising platforms.

## REFERENCES

[1] Tom Agan. 2007. Silent marketing: Micro-targeting. *Penn, Schoen and Berland Associates* (2007).
[2] Saleem Alhabash, Juan Mundel, and Syed Ali Hussain. 2017. Social Media Advertising: Unraveling the Mystery Box. In *Digital Advertising*. Routledge, 285–299.
[3] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebooks Explanations. In *NDSS*. The Internet Society.
[4] Sitaram Asur, Bernardo A. Huberman, Gábor Szabó, and Chunyan Wang. 2011. Trends in Social Media: Persistence and Decay. In *ICWSM*. The AAAI Press.
[5] Andrew O Ballard, D Sunshine Hillygus, and Tobias Konitzer. 2016. Campaigning online: Web display ads in the 2012 presidential campaign. *PS: Political Science & Politics* 49, 3 (2016), 414–419.
[6] Oana Barbu. 2014. Advertising, microtargeting and social media. *Procedia-Social and Behavioral Sciences* 163 (2014), 44–49.
[7] Susan B. Barnes. 2006. A privacy paradox: Social networking in the United States. *First Monday* 11, 9 (2006).
[8] Katharina Baum, Stefan Meißner, Olga Abramova, and Hanna Krasnova. 2019. Do they Really Care about Targeted Political Ads? Investigation of User Privacy Concerns and Preferences. In *ECIS*. AIS.
[9] Anja Bechmann. 2013. Internet profiling: The economy of data intraoperability on Facebook and Google. *MedieKultur: Journal of media and communication research* 29, 55 (2013), 72–91.

[10] Colin J Bennett. 2016. Voter databases, micro-targeting, and data protection law: Can political parties campaign in Europe as they do in North America? *International Data Privacy Law* 6, 4 (2016), 261–275.

[11] Adam Bermingham and Alan Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*. Asian Federation of Natural Language Processing.

[12] Jan Blom. 2000. Personalization: A Taxonomy. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. ACM, 313–314.

[13] Anthony Boadle. 2018. Facebook's WhatsApp flooded with fake news in Brazil election. *Reuters*. Retrieved September 06, 2019 from https://www.reuters.com/article/us-brazil-election-whatsapp-explainer/facebooks-whatsapp-flooded-with-fake-news-in-brazil-election-idUSKCN1MU0UP.

[14] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (2012), 295–298.

[15] C-SPAN. 2019. Technology Companies and Algorithms. Video. Retrieved September 06, 2019 from https://www.c-span.org/video/?462071-1/technology-companies-algorithms.

[16] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. 2018. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. In *USENIX Security Symposium*. USENIX Association, 479–495.

[17] John Cheney-Lippold. 2018. *We are data: Algorithms and the making of our digital selves*. NYU Press.

[18] Jeff Chester and Kathryn C Montgomery. 2017. The role of digital marketing in political campaigns. *Internet Policy Review* 6, 4 (2017).

[19] Roger Clarke. 2019. Risks inherent in the digital surveillance economy: a research agenda. *Journal of Information Technology* 34, 1 (2019), 59–80.

[20] Michael Cornfield and Kate Kaye. 2009. Online Political Advertising. In *Politicking online: The transformation of election campaign communications*, Costas Panagopoulos (Ed.). Rutgers University Press, 163–175.

[21] Jeannette Cwienk. 2017. Wahlkampf auf Facebook, Twitter und Co. *DW* (2017). Retrieved January 20, 2020 from https://www.dw.com/de/wahlkampf-auf-facebook-twitter-und-co/a-37535854.

[22] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings. *Proceedings on privacy enhancing technologies* 2015, 1 (2015), 92–112.

[23] Eco. 2017. Germany: Land of Data Protection and Security - but why? *dotmagazine* (2017). Retrieved January 20, 2020 from https://www.dotmagazine.online/issues/security/germany-land-of-data-protection-and-security-but-why.

[24] Laura Edelson, Shikhar Sakhuja, Ratan Dey, and Damon McCoy. 2019. An Analysis of United States Online Political Advertising Transparency. *arXiv preprint* arXiv:1902.04385 (2019).

[25] Jasmine Enberg. 2019. Digital Ad Spending 2019 Global. *eMarketer* (2019). https://www.emarketer.com/content/global-digital-ad-spending-2019.

[26] Facebook. 2019. Find your audience. Retrieved September 06, 2019 from https://www.facebook.com/business/ads/ad-targeting.

[27] Facebook. 2019. Impressions. Retrieved September 06, 2019 from https://www.facebook.com/business/ help/675615482516035.

[28] Facebook. 2019. The Difference Between Boosted Posts and Facebook Ads. Retrieved September 06, 2019 from https://www.facebook.com/business/help/317083072148603.

[29] Facebook. 2019. Your ad preferences. Retrieved September 06, 2019 from https://www.facebook.com/ads/preferences/.

[30] Casey Fiesler and Blake Hallinan. 2018. "We Are the Product": Public Reactions to Online Data Sharing and Privacy Controversies in the Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, Article 53, 13 pages.

[31] Michael M Franz and Travis N Ridout. 2007. Does political advertising persuade? *Political Behavior* 29, 4 (2007), 465–491.

[32] Aritz Arrate Galán, José González Cabañas, Ángel Cuevas, María Calderón, and Rubén Cuevas Rumin. 2019. Large-Scale Analysis of User Exposure to Online Advertising on Facebook. *IEEE Access* 7 (2019), 11959–11971.

[33] Avijit Ghosh, Giridhari Venkatadri, and Alan Mislove. 2019. Analyzing Political Advertisers' Use of Facebook's Targeting Features. *IEEE Symposium on Security and Privacy. Workshop on Technology and Consumer Protection* (2019).

[34] Guy J Golan, Stephen A Banning, and Lisa Lundy. 2008. Likelihood to vote, candidate choice, and the third-person effect: Behavioral implications of political advertising in the 2004 presidential election. *American Behavioral Scientist* 52, 2 (2008), 278–290.

[35] Avi Goldfarb and Catherine E Tucker. 2011. Privacy regulation and online advertising. *Management science* 57, 1 (2011), 57–71.

[36] Google. 2019. Ad Settings. Retrieved September 06, 2019 from https://adssettings.google.com/.

[37] Google. 2019. Audience targeting. Retrieved September 06, 2019 from https://support.google.com/adwords/editor/answer/1052569?hl=en&ref_topic=2996408.

[38] Brian Heredia, Joseph Prusa, and Taghi Khoshgoftaar. 2017. Exploring the effectiveness of Twitter at polling the United States 2016 presidential election. In *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 283–290.

[39] Eitan D Hersh. 2015. *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press.

[40] D Sunshine Hillygus and Todd G Shields. 2008. *The persuadable voter: Wedge issues in presidential campaigns*. Princeton University Press.

[41] Paul Hitlin and Lee Rainie. 2019. Facebook Algorithms and Personal Data. *Pew Research Center: Internet & Technology* (2019). https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/.

[42] Donna L Hoffman and Marek Fodor. 2010. Can you measure the ROI of your social media marketing? *MIT Sloan Management Review* 52, 1 (2010), 41.

[43] Roope Jaakonmäki, Oliver Müller, and Jan Vom Brocke. 2017. The impact of content, context, and creator on user engagement in social media marketing. In *Proceedings of the 50th Hawaii international conference on system sciences*. AIS.

[44] Young Mie Kim, Jordan Hsu, David Neiman, Colin Kou, Levi Bankston, Soo Yun Kim, Richard Heinrich, Robyn Baragwanath, and Garvesh Raskutti. 2018. The stealth media? Groups and targets behind divisive issue campaigns on Facebook. *Political Communication* 35, 4 (2018), 515–541.

[45] Simon Kingsnorth. 2019. *Digital marketing strategy: An integrated approach to online marketing*. Kogan Page Publishers.

[46] Ralf Kneuper and Mark Perlitz. 2019. Datenschutz bei nutzungsbasierter Online-Werbung. *Datenschutz und Datensicherheit* 43, 8 (2019), 502–506.

[47] Aleksandra Korolova. 2010. Privacy Violations Using Microtargeted Ads: A Case Study. In *ICDM Workshops*. IEEE Computer Society, 474–482.

[48] Daniel Kreiss. 2016. *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.

[49] Daniel Kreiss and Shannon C. Mcgregor. 2019. The "Arbiters of What Our Voters See": Facebook and Google's Struggle with Policy, Process, and Enforcement around Political Advertising. *Political Communication* (2019), 1–24.

[50] Dokyun Lee, Kartik Hosanagar, and Harikesh S Nair. 2018. Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science* 64, 11 (2018), 5105–5131.

[51] Lynda Lee Kaid, Juliana Fernandes, and David Painter. 2011. Effects of political advertising in the 2008 presidential campaign. *American Behavioral Scientist* 55, 4 (2011), 437–456.

[52] Janette Lehmann, Mounia Lalmas, Elad Yom-Tov, and Georges Dupret. 2012. Models of user engagement. In *International conference on user modeling, adaptation, and personalization*. Springer, 164–175.

[53] Vishal Monga and Brian L Evans. 2006. Perceptual image hashing via feature points: performance evaluation and tradeoffs. *IEEE transactions on Image Processing* 15, 11 (2006), 3452–3465.

[54] Matthew P Motta and Erika Franklin Fowler. 2016. The content and effect of political advertising in US campaigns. In *Oxford Research Encyclopedia of Politics*.

[55] Mozilla. 2019. Facebook's Ad Archive API is Inadequate. *The Mozilla Blog* (2019). https://blog.mozilla.org/blog/2019/ 04/29/facebooks-ad-archive-api-is-inadequate/.

[56] Nanigans. 2019. Global Facebook Advertising Benchmark Report Q1 2019. (2019).

[57] David W Nickerson and Todd Rogers. 2014. Political campaigns and big data. *Journal of Economic Perspectives* 28, 2 (2014), 51–74.

[58] Robert W Palmatier and Kelly D Martin. 2019. Big Data's Marketing Applications and Customer Privacy. In *The Intelligent Marketer's Guide to Data Privacy*. Springer, 73–92.

[59] Orestis Papakyriakopoulos, Simon Hegelich, Morteza Shahrezaye, and Juan Carlos Medina Serrano. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society* 5, 2 (2018).

[60] Flora Poecze, Claus Ebster, and Christine Strauss. 2018. Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. *Procedia computer science* 130 (2018), 660–666.

[61] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis) Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference 2019*. ACM, 818–828.

[62] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19)*. ACM, 140–149.

[63] Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos, and Simon Hegelich. 2019. The Rise of Germany's AfD: A Social Media Analysis. In *Proceedings of the 10th International Conference on Social Media and Society (SMSociety '19)*. ACM, 214–223.

[64] Morteza Shahrezaye, Orestis Papakyriakopoulos, Juan Carlos Medina Serrano, and Simon Hegelich. 2019. Measuring the Ease of Communication in Bipartite Social Endorsement Networks: A Proxy to Study the Dynamics of Political Polarization. In *Proceedings of the 10th International Conference on Social Media and Society (SMSociety '19)*. ACM, 158–165.

[65] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. In *FAT (Proceedings of Machine Learning Research)*, Vol. 81. PMLR, 5–19.

[66] Emil Stjernholm. 2014. Political ads in the Swedish 2006 and 2010 parliamentary elections: Focus on the Sweden Democrats and the role of audiovisual media. *Online Journal of Communication and Media Technologies* 4, 2 (2014), 32–50.

[67] SZ. 2019. So hat Deutschland gewählt. *Süddeutsche Zeitung* (2019). Retrieved September 06, 2019 from https://www.sueddeutsche.de/politik/europawahlen-deutschland-ergebnisse-1.4456025.

[68] Monika Taddicken. 2014. The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication* 19, 2 (2014), 248–273.

[69] Zeynep Tufekci. 2015. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colo. Tech. LJ* 13 (2015), 203.

[70] Michail Vafeiadis, Ruobing Li, and Fuyuan Shen. 2018. Narratives in political advertising: An analysis of the political advertisements in the 2014 midterm elections. *Journal of Broadcasting & Electronic Media* 62, 2 (2018), 354–370.

[71] Giridhari Venkatadri, Elena Lucherini, Piotr Sapiezynski, and Alan Mislove. 2019. Investigating sources of PII used in Facebook's targeted advertising. *PoPETs* 2019, 1 (2019), 227–244.

[72] Julia Carrie Wong. 2019. Document reveals how Facebook downplayed early Cambridge Analytica concerns. *The Guardian*. Retrieved September 06, 2019 from https://www.theguardian.com/technology/2019/aug/23/cambridge-analytica-facebook-response-internal-document.

[73] Jing Zhang and En Mao. 2016. From online motivations to ad clicks and to behavioral intentions: An empirical study of consumer response to social media advertising. *Psychology & Marketing* 33, 3 (2016), 155–164.

[74] Frederik Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes H de Vreese. 2018. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review* 14, 1 (2018), 82–96.

# 5 Partisan Users

## 5.1 Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok

### Authors

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich

### In

12th ACM Conference on Web Science. Association for Computing Machinery, New York, NY, USA, pages 257–266 (2020).
DOI: https://doi.org/10.1145/3394231.3397916

### Abstract

TikTok is a video-sharing social networking service, whose popularity is increasing rapidly. It was the world's second-most downloaded app in 2019. Although the platform is known for having users posting videos of themselves dancing, lip-syncing, or showcasing other talents, user-videos expressing political views have seen a recent spurt. This study aims to perform a primary evaluation of political communication on TikTok. We collect a set of US partisan Republican and Democratic videos to investigate how users communicated with each other about political issues. With the help of computer vision, natural language processing, and statistical tools, we illustrate that political communication on TikTok is much more interactive in comparison to other social media platforms, with users combining multiple information channels to spread their messages. We show that political communication takes place in the form of communication trees since users generate branches of responses to existing content. In terms of user demographics, we find that users belonging to both the US parties are young and behave similarly on the platform. However, Republican users generated more political content and their videos received more responses; on the other hand, Democratic users engaged significantly more in cross-partisan discussions.

### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok

Juan Carlos Medina Serrano
Technical University of Munich
juan.medina@tum.de

Orestis Papakyriakopoulos
Technical University of Munich
orestis.p@tum.de

Simon Hegelich
Technical University of Munich
simon.hegelich@hfp.tum.de

## ABSTRACT

TikTok is a video-sharing social networking service, whose popularity is increasing rapidly. It was the world's second-most downloaded app in 2019. Although the platform is known for having users posting videos of themselves dancing, lip-syncing, or showcasing other talents, user-videos expressing political views have seen a recent spurt. This study aims to perform a primary evaluation of political communication on TikTok. We collect a set of US partisan Republican and Democratic videos to investigate how users communicated with each other about political issues. With the help of computer vision, natural language processing, and statistical tools, we illustrate that political communication on TikTok is much more interactive in comparison to other social media platforms, with users combining multiple information channels to spread their messages. We show that political communication takes place in the form of communication trees since users generate branches of responses to existing content. In terms of user demographics, we find that users belonging to both the US parties are young and behave similarly on the platform. However, Republican users generated more political content and their videos received more responses; on the other hand, Democratic users engaged significantly more in cross-partisan discussions.

## CCS CONCEPTS

• **Networks** → **Social media networks**; • **Human-centered computing** → **Social network analysis**.

## KEYWORDS

TikTok, political communication, social media, US politics

**ACM Reference Format:**
Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. In *12th ACM Conference on Web Science (WebSci '20), July 6–10, 2020, Southampton, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3394231.3397916

## 1 INTRODUCTION

Political communication is contingent on the available information channels. Because social media platforms are a fruitful space for socialization, politics largely takes place on them. Political candidates often exploit such media to interact with the electorate and to place targeted and personalized advertising. At the same time, partisan users utilize their social media accounts to engage in political discourse. Similarly, a large portion of society obtains its political news updates from social media sources.

Two factors shape the final forms taken by political communication on social media: *how* a platform is designed, and *who* uses this space for political purposes. The design of a social media service configures the available information channels for political discourse; its user-base shapes both the generated political content and dictates whether or not a platform can prevail as a significant political space.

Until now, researchers have considered Facebook, Twitter, and YouTube as the most politically relevant social media [48], given the vast number of users who engage daily with these platforms. Nevertheless, social media usage is dynamic. Users change or migrate from one service to another, and some platforms are abandoned as others become popular [31]. TikTok, a video-sharing social networking service, has recently witnessed a surge in its popularity. It became the world's second most downloaded app in 2019 [53].

### Motivation

Although researchers have extensively analyzed other popular social media platforms, both by explaining user political behavior and uncovering how platform design influences political communication, no study has focused on TikTok. The present study, therefore, aims to bridge this research gap, intending specifically to understand *who* uses TikTok for political purposes and *how* the design of this platform shapes the flow of political information. For this, we focus on US politics to answer the following research question:

> **RQ: What are the features of political communication on TikTok in terms of (a) partisan users, (b) interaction structure, and (c) diffused content?**

### Original Contributions

- We provide a first overview of political communication on TikTok by investigating videos of US Republican and Democratic partisans.
- We employ computer vision, natural language processing, and statistical tools to evaluate the ways in which partisans combine sound, video, and text to spread their messages.
- We study TikTok's interaction design and illustrate how partisans engage in political discussions. We show that TikTok fosters a novel type of political interactivity that is not available on other online social networks.
- We investigate user demographics and show that the partisan users are young and behave similarly regardless of their


*WebSci '20, July 6–10, 2020, Southampton, United Kingdom*

ACM ISBN 978-1-4503-7989-2/20/07…$15.00
https://doi.org/10.1145/3394231.3397916

political preferences. We find that Republican users are more active in creating political content and that they receive more responses. We further find that Republicans prefer to engage in video discussions with other Republicans, while Democrats are more open to reach out to users with opposing views.
- Finally, we discuss other issues related to political communication, privacy, and security on TikTok.

# 2 BACKGROUND & RELATED WORK

## 2.1 Political Communication on Social Media

The behavior of multiple actors determines the political communication that occurs on social media platforms. Politicians, partisans, and the general public interact constantly, generating complex communication patterns. The design and algorithms of the platforms influence these interactions along with malicious actors to generate a political landscape that can be difficult to understand. Researchers have therefore extensively analyzed various properties of Facebook, Twitter, YouTube, and other similar social media spaces to grasp the nature of politics on social media.

Several studies have investigated how politicians use social media for political purposes. They have shown that politicians use platforms in different ways, depending on the audiences and sociotechnical environments [21, 43, 46, 47]. They have also explained how services are used for personalized advertising campaigns [39] and whether the presence of politicians on these platforms affects their electoral popularity [19].

Other studies have explicitly focused on user political behavior. They have investigated how partisans of different political orientations use social media platforms [10, 44], analyzed how user activities are generally distributed [40] and evaluated the spread messages' content and polarity [15, 36, 38, 45]. Studies have also focused on the usage of social media in periods of social unrest [50], analyzed online social platforms as spaces for the coordination of social movements [49], investigated how different social groups behave, and scrutinized the conditions in which they help to polarize and segregate the citizenry [4, 6, 16].

Inseparable to political communication on social media is news consumption, with a large proportion of the public using the social media services as their primary information source pertaining to world events. Many studies have investigated agenda settings effects, as well as how different types of news coverage affect user behavior and contribute to attitude formation [9, 23, 29]. Moreover, researchers have also extensively studied how low credibility news and misinformation are diffused on social media platforms by real, fake, or automated accounts [3, 17, 22, 27, 28, 51].

An equally important role in understanding user behavior is the analysis of the content filtering algorithms employed by social media services. Thus, researchers have investigated social platforms as algorithmic ecosystems and have studied whether and how recommendation algorithms influence the public's political behavior and opinions [5, 32, 37, 40, 41]. In the same vein, scholars have evaluated platform design features and how they shape information diffusion [21, 33, 47]. These dimensions of political communication have been extensively analyzed for the US on different social media platforms [2, 20, 25, 26, 30, 52], but not yet on TikTok.

## 2.2 TikTok

TikTok was created by ByteDance, a Beijing-based tech company. The company had previously launched Douyin for the Chinese market in September 2016. Subsequently, the company launched TikTok in 2017 for markets outside China. Both services are similar, but they run on separate servers to comply with China's regulations. In 2018, TikTok merged with the social media app Musical.ly to create a larger video community. In October 2019, TikTok and Douyin jointly achieved 800 million monthly active users [34]. In the United States, 60% of TikTok's 26.5 million monthly active users are aged between 16 and 24 [24]. The platform is mainly accessible through a mobile app. Although it is possible to access posted videos from non-mobile devices, the features are limited as it is not possible to create content or read user comments.

TikTok offers users a unique method of sharing creative videos of themselves, their surroundings, or a compilation of external audiovisual content. The simplest videos consist only of text superimposed onto a colored background. Videos can then be more complex by including images, video clips, and sounds. The images and video footage can be altered using the app's voice effects, image filters, and video speed controllers. The maximum length of a video post is 60 seconds, and they can consist of a collection of shorter video clips that tell a story when they are combined. When the users post videos, they can add a caption with hashtags to describe their clips. Like Twitter, the most used hashtags represent topics that are trending on the platform, and like Instagram, the video clips are classified according to their hashtags.

TikTok is considered a social media platform because like Twitter and Instagram, its users have a social group of followers and other users they follow. However, the main feature that differentiates TikTok from other social media services, is the videos' background music, which represent the core message that the users want to convey. Users can select background music for their videos from a wide variety of music genres and can even create original sound clips. Any sound clip, including user voice messages, can be selected by other users to use in their videos. For many videos, the music serves as part of a dance routine, a lip-synching battle, or as the backdrop for a comedy skit. However, sound can also function as a story builder and can be used to deliver a specific message. For example, a famous original sound clip begins with La Roux's song, Bulletproof, remixed by Gamper & Dadoni. Then the music stops, four gunshots are heard, and a man's voice says "there's not any". Text messages appear in the part of the clip that plays the song, citing reasons why a particular cause should be supported, and the user points at the messages. When the gunshots begin, however, the user makes a gun sign with a hand and shoots at the text snippets containing the reasons for support. The user is implying that the citing reasons are invalid and that there is no real reason for people to support that cause.

Users consume content by viewing an algorithmically generated feed of videos on the so-called "For you" page, which is the landing place when users open the app. Although there is no explanation of how the algorithm work, the videos that appear to the user largely rely on a central recommendation algorithm instead of on the activities of the user's social network [12]. According to TikTok, the "For you" page is "a personalized video feed specifically for you based on what you watch, like, and share."[1] This contrasts with Facebook's and Twitter's feed, which rely mostly on the user's social graph and resemble more YouTube's recommendation system. Users can also search for hashtags, sounds and find the trending videos on the "Discover" page.

A unique feature on TikTok is the duet. A duet is a response video that allows users to reply directly to a video post with a video of their own. The original and duet videos play side by side and the music clip from the original video's audio is preserved. Since the audio does not change, the duet exhibits users responding through text snippets, images, or facial expressions. Users can also create a reply duet from an existing duet. In such instances, three videos appear together. Figure 1 shows a screenshot of a duet on TikTok. The original video is placed on the right and the duet video on the left. The screenshot displays the number of likes, comments, and shares the duet video attracts. The music that appears on the bottom is a music clip from a remixed song. In the image, both users point to a text snippet and communicate their perspectives as they dance to the same music. In this way, TikTok can be used to share opinions on controversial topics. In this study, we focus on political content to determine how TikTok users interact and show their support for a political party with the help of music, video, and image.

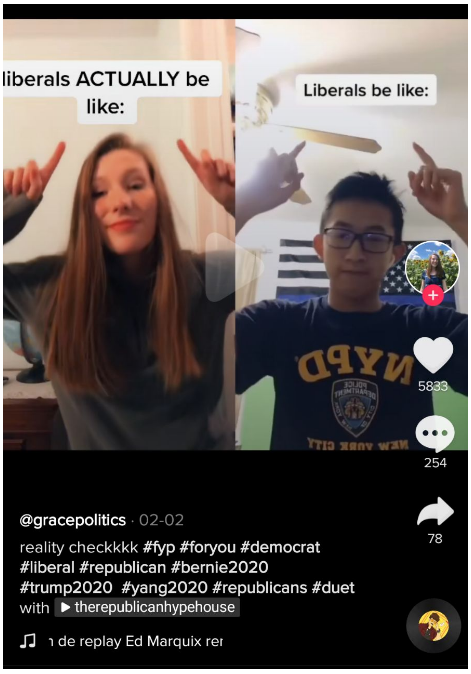


**Figure 1: Screenshot of a duet on TikTok. On the right is the original video and on the left, the video posted in response.**

[1]https://apps.apple.com/us/app/id835599320

# 3 DATA & METHODS

## 3.1 Data Collection

TikTok allows users to search for videos with a specific hashtag and view the most popular results. We decided to crawl the videos containing the hashtags *#republican* and *#democrat* on February 1, 2020. The hashtag search yields a limited number of videos and it is not clear how this limit is defined. Popularity may play a role as a hashtag search showcases the most popular videos. For the two hashtag queries, we obtained a different number of videos. The collection process resulted in 3,310 videos: 2,362 with the hashtag *republican*, and 1,831 with the hashtag *democrat*. Of the total, 350 were duets; thus, we also collected the original videos from them if they were not yet in the data. To expand the dataset, we searched for duets to the videos we had collected. Unfortunately, TikTok does not offer a search by video feature that directly links a video to its duets. However, it is possible to search by sound clip. This search shows videos that have employed the same sound. Our approach was to search for the sound of each video and to add the videos that were dueting to the videos in our dataset. As with search by hashtag, only a limited number of results can be obtained. This presents a limitation to collecting all the duets to a video, especially for videos that use extremely popular sounds. Nonetheless, searching for original sounds often yielded only the duets of the given video. After this procedure was complete, our dataset consisted of 7,825 TikTok videos. Most of the videos were created between October 2019 and February 2020. The oldest included video was posted in March 2019.

Before beginning the analysis, we manually labeled every original video and duet as pro-Republican, pro-Democrat, or nonpartisan. This coding was conducted by two of the main authors. Both authors labeled each video individually. For the cases where authors disagreed, the third author was consulted to resolve the coding conflict. Videos that directly supported or opposed a political party or a member of a party were classified accordingly. We labeled videos opposing one Democratic candidate but expressing support for another candidate of the same party as pro-Democrat. Videos in which users articulated their standpoints on issues such as abortion, guns, and LGBT rights, were not directly attributed to any political party. We only classified those videos that indicated a clear political affinity toward the Republican or Democratic parties as partisan. For example, social issue videos with both the *republican* and *democrat* hashtags were coded as nonpartisan. In total, we classified 5,946 videos as partisan posts. Apart from assigning partisanship, we grouped the partisan videos into four categories: 1) videos that included the user's face; 2) videos where the user appears but does not show its face or videos culled from other sources such as news clips; 3) videos solely comprised of images; and 4) videos that only showcased textual content. We used the same coding procedure as the classification of partisan videos. It was important for us to view all the videos to be able to understand TikTok's communication structures and the user behavior displayed on the platform.

## 3.2 Methods

TikTok videos are rich in features and extra pre-processing steps were required to extract the information for analysis. From the original videos, we started by extracting the text snippets from the videos. For this, we divided the videos into images every two seconds and then employed Tesseract, an optical character recognition engine. From the original videos that included the user's face, we used the images and processed them via Microsoft's Azure Face API[2], which allows emotions, gender, and age to be extracted. Additionally, we employed IBM's text to speech API[3] to extract the audio from the videos that contained original sounds created by the users.

To explore the difference in the usage of hashtags, we employed a measure of political valence introduced by Conover et al. [16]. They defined political valence $V$ of a hashtag $h$ as

$$V(h) = 2 * \frac{\frac{N(h,R)}{N(R)}}{\frac{N(h,D)}{N(D)} + \frac{N(h,R)}{N(R)}} - 1 \qquad (1)$$

where $N(h, D)$ and $N(h, R)$ indicate the number of appearances of a hashtag, and N(D) and N(R) represent the total number of hashtags in the Democratic and Republican videos, respectively. This equation bounds the valence between -1 for hashtags only used in Democratic videos and +1 for hashtags appearing only in Republican videos.

We created a graph of interactions where the users are the nodes and the directed edges represent duet interactions. This graph allows to measure the extent of cross-ideological (R⟶D, D⟶R) and intra-party interactions (R⟶R, D⟶D). We used another measure from Conover et al., which divides the observed number of interactions with the expected number of interactions. The expected value assumes that the source of the edge is preserved and the target of the edge is randomly assigned to one user, irrespective of the individual's political orientation. For example, the expected interactions between Republicans and Democrats is defined as:

$$E[R \longrightarrow D] = k_R * \frac{Users_D}{Users_D + Users_R} \qquad (2)$$

where $k_R$ refers to the number of edges originating from pro-Republican videos and $Users_D$ denotes the number of Democratic users.

Finally, we applied topic modeling to evaluate the content of video captions. We used a Latent Dirichlet Allocation algorithm [8] to extract the latent thematic topics and to calculate the empirical distribution of videos belonging to the identified topics $f(video|topic)$. Using this distribution, we computed and compared the amount of Democratic and Republican videos associated with each topic. In order to calibrate the number of topics and the latter model hyperparameters, we perform sensitivity analysis for various values and select the model with the highest topic coherence score, as suggested by Röder et al. [42].

## 4 RESULTS

We illustrate different features of political communication on TikTok (**RQ**) in the following three subsections. First, we describe the general statistics of the collected dataset, including user demographics and activities. Second, we describe and evaluate political interactions between partisan users. Finally, we analyze the content of different information channels used in TikTok videos to detail the nature of the political discourse.

[2]https://azure.microsoft.com/en-us/services/cognitive-services/face/

[3]https://www.ibm.com/cloud/watson-text-to-speech

### 4.1 Descriptive Statistics

Of the 5,946 partisan videos, 2,802 are original videos and 3,144 are duets. Table 1 shows the total number of videos classified as pro-Democrat or pro-Republican and the extent of the interactions generated by these videos. In our dataset, there are two times more Republican videos than Democratic videos. Overall, the Republican videos accumulate more likes, shares, and comments. We applied one-sided Mann-Whitney tests for each reaction to compare if there is a significant difference between the partisan videos. For likes, the Republican median is 497 and the Democratic median is 232 ($p < 0.00$). In terms of shares, the Republican median is 6 and the Democratic median is 3 ($p < 0.00$). With regard to the comments, the Republican median is 19 and the Democratic median is 13 ($p < 0.01$). All the tests are significant and show that Republican videos attracted more interactions in general.

**Table 1: Number of videos created by pro-Republican and pro-Democratic users and user engagement (likes, shares, and comments) with them.**

| | Videos | Users | Likes | Shares | Comm. |
|---|---|---|---|---|---|
| *Republican* | 3,987 | 1,957 | 15,533,963 | 817,728 | 500,514 |
| *Democrat* | 1,959 | 1,249 | 10,663,139 | 392,468 | 257,199 |

Of the original videos, 70% included the user's face, 22% featured other video content, 7% comprised only image content, and 1% only exhibited text. We performed feature extraction for the original videos that included the user's face. To this end, we divided the videos into pictures and obtained the features for each picture. With Microsoft's API, we were able to extract gender, age and emotions, which include anger, contempt, fear, happiness, sadness, and surprise. Afterward, we averaged the emotions and ages of all the pictures to obtain the mean values for each video. We used the mode for gender given that it is a categorical feature. For age and gender, we aggregated the videos per user to obtain unique values. We manually categorized the users, for which the mode of the gender was inconclusive. From the original Republican videos, 219 users are male and 187 are female, whereas for the Democratic videos, 84 users are male and 118 are female. We perform goodness of fit chi-squared tests to evaluate the male-female user balance. For the overall population, gender is balanced ($\chi^2$=0.01, p=0.935). The same applies for the Republican partisans ($\chi^2$=2.52, p=0.11). However, a significantly larger number of Democratic partisans posting original videos are female ($\chi^2$=5.72, p=0.016). Figure 2 portrays the cumulative age distribution of Democratic and Republican users. We observe that in general, the Democratic users are younger than Republican supporters. The percentage of Republicans between 16 and 24 years old is close to 60%, which mirrors the percentage of US TikTok users in the same demographic group. For the Democratic users, however, this percentage is closer to 70%, younger than the mean user age. Nevertheless, the majority of users creating political content are below 40 for both parties in our data sample.

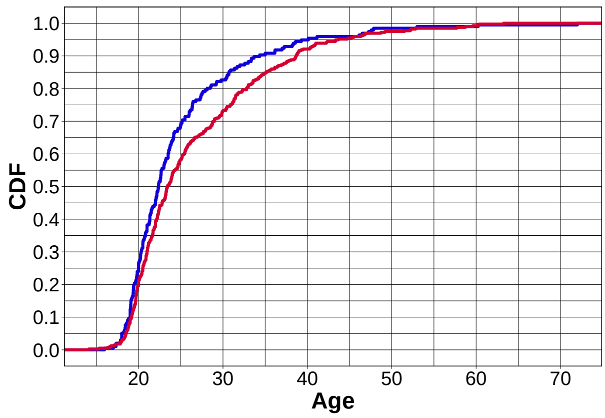


**Figure 2: Cumulative distribution of the users' age divided as Democratic (blue) and Republican (red) users.**

**Table 2: Average emotion expressed by TikTok users divided by partisanship.**

| Emotion | Democrat | Republican |
|---|---|---|
| anger | 0.021 | 0.019 |
| contempt | 0.013 | 0.014 |
| fear | 0.004 | 0.004 |
| happiness | 0.217 | 0.212 |
| neutral | 0.619 | 0.635 |
| sadness | 0.047 | 0.044 |
| surprise | 0.074 | 0.067 |

Table 2 illustrates the average emotion expressed for the posted videos per party. We do not observe significant differences between the two groups. We conclude that users on TikTok express themselves similarly irrespective of the party they support. Interestingly, happiness and surprise have higher averages than anger or sadness. We presume that this finding is related to the nature of TikTok comedy skits. Indeed, we observed many videos that relied heavily on sarcasm: smiling and dancing individuals confronting users supporting the views of the opposite political party and mocking or belittling them.

## 4.2 Interaction Structure

Political communication on social media is influenced directly by the design of the platform by determining the interaction structure between users. The interactions can be ordered in consecutive levels of communication, with each increasing level representing a more direct response. We identify four levels of communication on TikTok. The first level corresponds to an indirect response when a user views a video. Although there is no active reaction from the user, the message is received and processed. Additionally, from the data perspective, the view counter on the video increases, and this metric can influence TikTok's recommendation algorithm. The second level of communication consists of a basic response that involves liking the video or sharing it. The next level constitutes written responses to a video through user comments. On other social media platforms, this is the highest level of user response to a posted element. On TikTok, however, a fourth level of communication allows users to respond with a video, a feature referred to as duet. The duet shows more similarities to face-to-face communication than to a written response, which can make the interaction between users feel more personal.

The structure of duets directly affects how political communication takes place on TikTok. The duets follow a tree structure, where users create branches by responding to other videos. We depict this tree structure of communication in Figure 3. On top of the tree, there is a political issue, which partisan users use as their motive for the production of pro-Democrat or pro-Republican videos. Connected on the second depth level are the original content videos. The third depth level represents the duets to the original videos. The next level nodes on the tree denote duets posted in response to previous duets. It is possible to continue to react with duets further than the three duet depth levels displayed in Figure 3. Inner nodes on the duet nodes have been included in the illustration to represent that duets, previous duets, and original videos appear side by side on TikTok. A user interacting with a deep level duet video can directly observe the complete communication chain on-screen without needing to scroll down. This differentiates duets from any feature available on other social media networks.

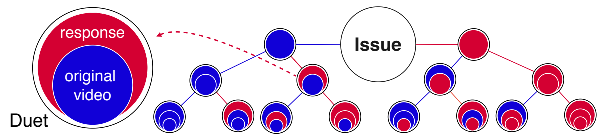


**Figure 3: Communication tree for TikTok duets.**

Using the manually labeled videos, we were able to quantify the duet interactions between partisan users. Table 3 shows the percentage of partisan and cross-ideological interactions. It also includes the observed interactions divided by the expected interactions as presented in Formula 2. We observe that 77% of the Republican duets represent responses to Republican users. In contrast, more than 80% of the Democratic duets were directed toward Republican supporters. This inverted dynamic also appears in the ratio between observed and expected interactions. Democrat-Republican and Republican-Republican interactions present a ratio larger than one (1.35 and 1.28 respectively). We include the ratio, as it is a standardized measure that can be used to compare to the outcomes of other studies. For example, Conover et al. [16] evaluated political communication in the US on Twitter and found that retweets had higher ratios than one for intra-partisan interactions (1.70, 2.32) and lower ratios than one for cross-partisan interactions (0.03, 0.05). Mentions displayed a similar but less pronounced effect for both parties (1.23, 1.31 for partisan exchanges, and 0.68, 0.77 for cross-ideological interactions). Therefore, the authors found that user behavior with regard to Twitter mentions and retweets was unrelated to the political party. In contrast, we find that TikTok duets represent a party-specific structure. This difference shows the importance of studying the effects of different platform designs on the political communication that occurs between users.

**Table 3: Interactions between partisan and cross-ideological users, including the percentage and ratio between observed and expected interactions.**

| | Percentage | | Ratio | |
|---|---|---|---|---|
| | →D | →R | →D | →R |
| Democrat | 19% | 81% | 0.48 | 1.35 |
| Republican | 22.6% | 77.4% | 0.57 | 1.28 |

We portray the duet interaction between users in Figure 4. Each node represents a user and the edges represent two users of a duet. Blue nodes indicate Democrats and red nodes designate Republicans. We only find one account that posted both pro-Democrat and pro-Republican videos, and this account was omitted from the analysis. The graph shows a tight Republican cluster in the middle with some Democratic users interacting with this community. The boundaries evince large clusters of Democratic users responding to Republican accounts. These users are separated from the main cluster because they did not interact with accounts other than a specific Republican user. The graph confirms the result of high partisan interactions among Republicans and the high cross-ideological interactions from Democrats to Republicans.

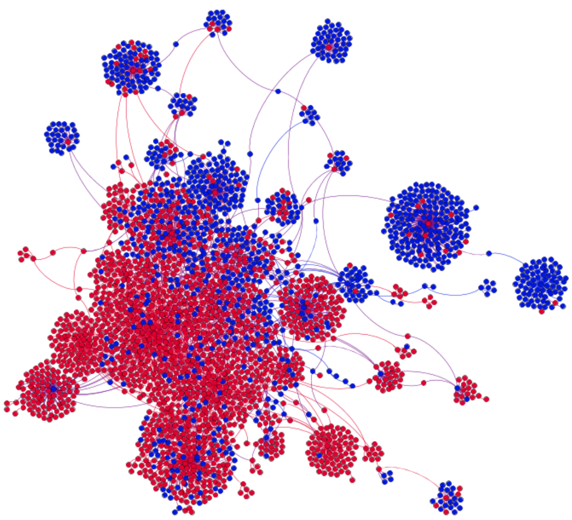

**Figure 4: Graph of duet interactions between partisan users. Red nodes correspond to Republicans and blue nodes to Democrats. Purple edges depict cross-ideological exchanges.**

## 4.3 Content Analysis

The content analysis of the various information channels on TikTok provided important insights on how political communication takes place on the platform. Each information channel was used differently by partisans. Partisan users generally avoided political statements in their profile descriptions, except for some users whose username explicitly stated their political affiliations. Most users added links to their social media handles on other platforms such as Twitter or Instagram, and some provided their Venmo accounts for fans to support them financially.

In contrast to the profile descriptions, video captions were extremely politicized. Users usually inserted as many political hashtags as possible from across the political spectrum, to ensure the visibility of their content. Their hashtag selections were also influenced by partisanship. Users inserted specific hashtags more often in accordance with their political orientation. To reveal this, we divided the valency spectrum in five equidistant groups and grouped the hashtags according to valences. Figure 6 presents the top ten words per group. Democratic partisans more frequently used hashtags related to the impeachment of Donald Trump, Bernie Sanders, Elizabeth Warren, and LGBT issues. Republican partisans more often used hashtags related to Trump's campaign slogans and phrases used by the alt-right to assign credibility to information such as #facts, #maketheswitch, and #openyoureyes. Regardless of partisanship, users added to their videos platform-specific hashtags such as #foryou, #foryourpage, and #xyzbca. These hashtags are irrelevant to the political discussions but constitute a cardinal aspect of TikTok interactions.

The topic modeling algorithm provided more detailed information about partisan interests for the specific period on TikTok. The model optimization process for the video captions yielded ten topics on which both Republican and Democrats generated content. Figure 5 shows that although some topics were more prevalent in posts by Democrats or by Republicans, both groups engaged in the same discussions. The topics that were discussed more or less equally by both sides related to social issues such as religion and abortion, guns and the second amendment, as well as discussions associated with daily political developments. Democrats generated content related to their party, about Trump's impeachment trial, as well as social diversity. In comparison, Republicans created more

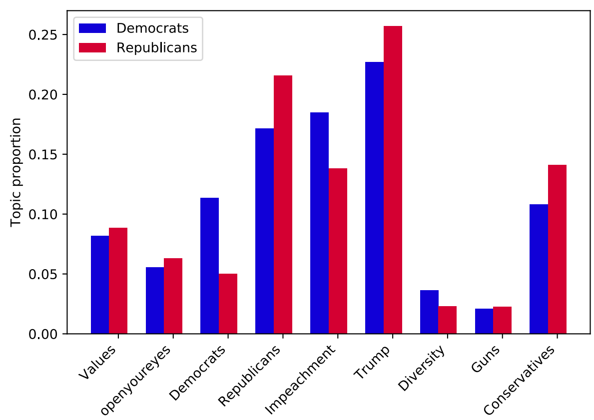


**Figure 5: Ten predominant topics appearing in videos captions on TikTok and their distribution between Democratic and Republican users.**

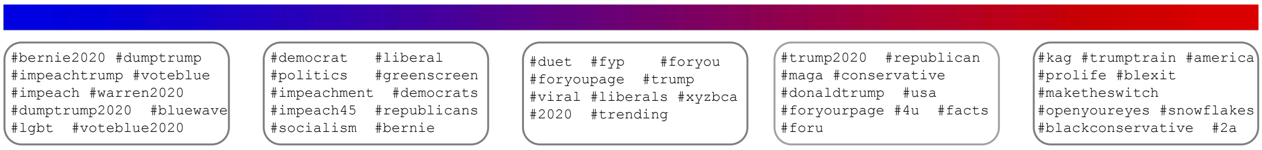


**Figure 6: Political valence of hashtags on the Democratic-Republican spectrum.**

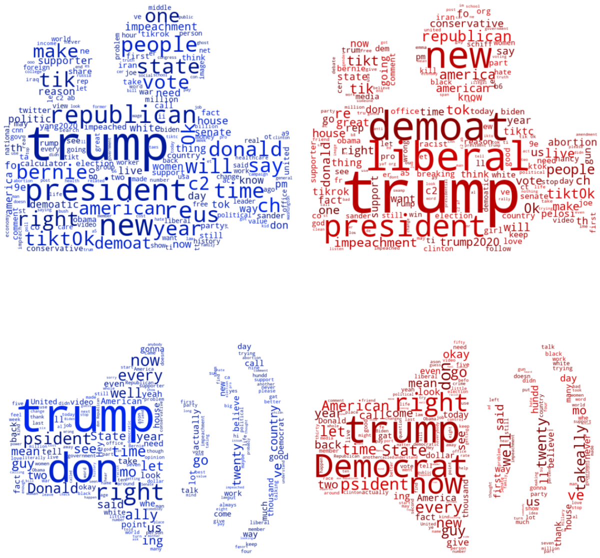

**Figure 7: Most frequent words appearing in embedded texts snippets (up) and user audio messages (down).**

videos about their party, conservative values in general, and about Donald Trump's activities.

The analysis of audio and the text embedded in the videos evidenced particular ways in which these communication channels were used (Figure 7). Both Republican and Democratic partisans used sound and embedded text to call out their opponents, setting the stage for the audiovisual discourses. Donald Trump was the focal character of these channels and was most frequently mentioned in the videos. Nevertheless, the embedded text contained additional information about the opinions of the partisans, and often mentioned other political candidates, expressing support for them or criticizing them. Furthermore, the text also included issues of interest to the partisan users, such as Donald Trump's impeachment, abortion rights, healthcare, and the second amendment. These results illustrate that partisans used audio, video, captions, and user descriptions in different ways, creating a complex multichannel information flow in their political interactions.

Our final analysis assessed the extent of political content on TikTok in comparison to non-political content. TikTok's search tool reports the number of total user views for videos that include a particular hashtag. This allowed us to compare between political and non-political hashtags. Thus, we searched for a selected number of hashtags on February 1, 2020. Table 4 shows hashtags of several political actors, including their names and their names plus *2020*. We also include three popular personalities and the most popular hashtags on TikTok, *#foryou* and *#foryoupage*, for comparison. Among the political hashtags, *#Trump2020* leads by a substantial margin, with a total of 1.1 billion views. The political hashtag that comes second with 166.3 million views is *#Bernie2020*. Interestingly, the hashtags with the names of the candidate and *2020* have more views than hashtags that included only the candidate's name. The large difference between *#Trump2020* and the rest of the Democratic candidates corroborates our finding pertaining to the greater number of pro-Republican videos seen in our data and evidences the absence of a possible sample bias effect in the collection procedure. Apart from politics, hashtags referring to singers Billie Eilish and Shawn Mendes have total views comparable to Trump-related videos, whereas Greta Thunberg videos have the number of views in the same order of magnitude as the videos related the Democratic candidates. The most famous hashtag on TikTok has 1,687 more views than *#Trump2020*. As a rough cross-platform comparison, we include the number of Instagram posts for the same hashtags in Table 4. Instagram's search tool displays the number of posts for a hashtag but not the number of views. The number of posts cannot be compared to the number of views as they represent two different quantities. The proportions on Instagram between politicians and popular personalities are similar to the proportion in views on

**Table 4: Number of total views for TikTok videos with a given hashtag. Number of Instagram posts that include the same hashtags.**

| Hashtag | TikTok Views | Instagram Posts |
|---|---:|---:|
| *#foryou* | 1,687B | - |
| *#foryoupage* | 968B | - |
| *#trump* | 730.3M | 13M |
| *#trump2020* | 1.1B | 1.2M |
| *#bernie* | 34.8M | 564K |
| *#bernie2020* | 166.3M | 216K |
| *#biden* | 4.8M | 102K |
| *#biden2020* | 1.9M | 26.9K |
| *#warren* | 3.5M | 253K |
| *#warren2020* | 11.6M | 38.9K |
| *#billieeilish* | 3.5B | 5.6M |
| *#shawnmendes* | 1.4B | 9.5M |
| *#gretathunberg* | 100.5M | 381K |

TikTok. However, there are more videos with only the politicians' names than hashtags including *2020*. This result could signify that there is more content focused on the 2020 US presidential campaign on TikTok than on Instagram. In sum, we conclude that not only does US political content takes place on TikTok, but it also accounts for a large ecosystem on the platform.

## 5 DISCUSSION

The results of this study demonstrate that a new form of political communication takes place on TikTok. Communication still preserves its decentralized character as on other social media platforms, with users generating, sharing, and diffusing information. However, TikTok users do not just merely circulate content and comment it; they *become* the content. In contrast to Facebook and Twitter, where users exchange news articles in the form of URLs and articulate their political opinions through comments or feedback posts, TikTok users become active presenters of political information. Every TikTok user is a performer who externalizes personal political opinion via an audiovisual act, with political communication becoming a far more interactive experience than on YouTube or Instagram. Since every user seeks increased popularity to disseminate their messages more widely, they create short political spectacles resulting in the realization of *politics as entertainment*. Unlike television media where news anchors and political pundits are the showmen and women, everybody on TikTok is one. It is not a coincidence, therefore, that this intensive audiovisual universe attracts young users who actually "play" their politics on the platform.

The duet function is one of the main reasons why political communication is so interactive on TikTok. Users can employ a variety of elements to respond to videos posted by other members. These features range from simple facial reactions to text snippets that serve as fact-checking points. Some users even overwrite the original video's text to "correct" the other user's stance on a topic and showcase opposing arguments. Moreover, the audience interacting with duet videos can directly compare the different points of view. This duet structure contrasts with other social media platforms where public exchanges take the form of written responses that appear as a list under the original post. Duets also allow users to exhibit their creativity, in showing support or making counterarguments. For these reasons, we argue that duets are the closest feature on social media to an actual online public debate.

Given that TikTok's design introduces a novel way of conducting politics, it is reasonable to ask how this framework can transform other aspects of political communication. In this study, we illustrated how political partisans generate content and interact on TikTok. However, multiple other dimensions of political communication should also be investigated. For example, although news media URLs are not diffused as in other social platforms, many news media agencies already have TikTok profiles to broadcast reports to the public. The same applies to a handful of political candidates who use TikTok as a new medium of reaching the electorate. Recently, TikTok followed Twitter's suit and banned the placement of political advertisements on the platform [14]. These phenomena and decisions interfere directly with political campaigning and opinion formation; therefore, researchers should investigate these actions more comprehensively in the future. This also applies to researchers who plan to study the general user behavior on the platform. Scholars should seek to uncover whether the platform design and the deployed recommendation algorithms result in the polarization or segregation of social groups, and whether hyperactive user behavior has an agenda-setting effect on the platform. Although TikTok primarily involves real users who reveal themselves in front of the camera, it is equally important to study whether and how any misinformation attempts take place, as well as how users present controversial issues on the platform. Researchers should undertake the task of determining whether offensive and discriminatory content prevails on the platform.

Besides the aspect of political communication, further political issues regarding user privacy and security should be addressed. Given TikTok's open nature, such concerns for the users have already been raised [7] but require more in-depth evaluation. Although users can create private videos visible only to their friends, the platform is mainly geared toward the production of viral videos. This means that data is easily reachable for data mining processes. Indeed, TikTok is a rich information source because its content reveals the personal features of users through the immediacy of audiovisual media to their appearances, personalities, traits, vocal attributes, and points of views. Moreover, users creating and interacting with political content can be classified by their partisanship as we did in this study. With the manually classified videos, a machine learning algorithm may be trained to identify political content and to automatically assign partisanship to a TikTok user. This information can then be employed for political or advertising purposes as it is already prevalent on other platforms like Facebook [35]. However, the potential risks are higher on TikTok because advances in facial recognition technologies make it possible to identify individual users and match them with citizenship records. Although this scenario is also possible on other platforms, active TikTok users are open to publicly share their biometric data. There is a greater danger, therefore, for TikTok users to become incorporated in electoral or other databases that can be exploited for varied purposes. Political campaigns and third parties may be eager to collect data on young people as many of them are first-time voters or are still not old enough to vote. As such, they are in the process of creating their political identity and the information they perceive on social media platforms can permeate their eventual ideology.

TikTok can potentially redefine political communication as a new public arena for civic discourse. While other social media platforms are highly dependent on the friend structure and can thus foster echo-chambers [11, 18], TikTok's open structure may allow a more cross-partisan dialogue. Whether this assumption holds can only be answered through future research. Even if the hypothesis is proven true, the caveat that political confrontation can be counterproductive exists, and maybe especially applicable to a platform that is highly focused on the virality and humor of its content. Videos with sarcastic and ridiculing content can exacerbate bullying and other harmful behaviors that particularly afflict teenagers [1]. The research community should conduct further analyses that include psychological examinations of the influence exercised by the platform on the youth.

### Ethical Concerns

While conducting this study, we encountered serious ethical questions that must be taken into consideration. First, we crawled TikTok and explicitly collected public data, a portion of which concerned the behavior of young users. Given their age, young users may not be yet be fully aware of the consequences of putting themselves in the public sphere. To maintain data protection, we deleted the collected materials after the analyses. However, we preserved the video ids to allow the replication of this study. These ids can be found in our GitHub repository[4].

We encountered further ethical issues during the use of the Microsoft Azure image recognition APIs for the detection of age and gender. First, researchers have shown that such algorithms can potentially misclassify minorities and social groups [13]. Second, the algorithms for gender classification only provide binary male/female inferences and automatically neglect the existence of other genders. These issues should be kept in mind when this study is read and addressed in the future to promote ethical and inclusive research.

## 6 CONCLUSION

In this paper, we studied political communication on TikTok for the first time. We focused on videos related to US politics and evaluated textual, aural and visual information extracted from them. We analyzed the different levels of communication made possible by the platform design and especially concentrated on TikTok's unique duet feature. We then investigated the duet interactions from pro-Republican and pro-Democrat users. In our sample, we find a larger collection of Republican videos, which, on average, attracted more interactions than Democratic videos. Although we find that Democratic users are younger than Republican users, the majority of the users in our data are below 40 years old. We observed that Republican users generated duet videos from users who professed the same ideology more often, whereas Democratic users interacted more with cross-ideological users. Irrespective of their political preferences, however, partisan users expressed themselves in similar ways. Finally, we identified that political content appears to be a relevant aspect of TikTok's ecosystem. Further research is needed to understand how political content is disseminated on this novel social media platform. It would be especially beneficial if prospective studies examined the platform's design and its recommendation system because they are pivotal to the creation of user communities and the shaping of political interactions. Only through rigorous auditing can it be ensured that TikTok represents an open and unbiased arena for political communication.

## REFERENCES


[1] Sophia Alim. 2016. Cyberbullying in the world of teenagers and social media: A literature review. *International Journal of Cyber Behavior, Psychology and Learning* 6, 2 (2016), 68–95.

[2] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[3] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics* 6, 2 (2019).

[4] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.

[5] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[6] Pablo Barberá. 2014. How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University* 46 (2014).

[7] Emily Birnbaum. 2020. Agencies play catch-up over security concerns with TikTok. *The Hill* (2020). Retrieved February 10, 2020 from https://thehill.com/policy/technology/484406-agencies-play-catchup-over-security-concerns-with-tiktok.

[8] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine learning research* 3, Jan (2003), 993–1022.

[9] Pablo J Boczkowski, Eugenia Mitchelstein, and Mora Matassi. 2018. "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New media & society* 20, 10 (2018), 3523–3539.

[10] Leticia Bode and Kajsa E. Dalrymple. 2016. Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter. *Journal of Political Marketing* 15, 4 (2016), 311–332.

[11] Andrei Boutyline and Robb Willer. 2017. The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology* 38, 3 (2017), 551–569.

[12] Ryan Broderick. 2019. Forget The Trade War. TikTok Is Chinas' Most Important Export Right Now. *BuzzFeed.News* (2019). Retrieved February 10, 2020 from https://www.reuters.com/article/us-tiktok-cfius-exclusive/exclusive-u-s-opens-national-security-investigation-into-tiktok-sources-idUSKBN1XB4IL.

[13] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. 77–91.

[14] Blake Chandlee. 2019. Understanding our policies around paid ads. *TikTok Newsroom* (2019). Retrieved February 10, 2020 from https://newsroom.tiktok.com/en-us/understanding-our-policies-around-paid-ads.

[15] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Mean birds: Detecting aggression and bullying on Twitter. In *Proceedings of the 2017 ACM on web science conference*. ACM, 13–22.

[16] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on Twitter. In *Fifth international AAAI conference on weblogs and social media*. The AAAI Press.

[17] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.

[18] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific reports* 6 (2016), 37825.

[19] Robin Effing, Jos Van Hillegersberg, and Theo Huibers. 2011. Social media and political participation: Are Facebook, Twitter and YouTube democratizing our political systems?. In *International conference on electronic participation*. Springer, 25–35.

[20] Gunn Enli. 2017. Twitter as arena for the authentic outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *European journal of communication* 32, 1 (2017), 50–61.

[21] Gunn Sara Enli and Eli Skogerbø. 2013. Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. *Information, communication & society* 16, 5 (2013), 757–774.

[22] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[23] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.

[24] Echo Wang Alexandra Alper Greg Roumeliotis, Yingzhi Yang. 2019. U.S. opens national security investigation into TikTok - sources. *Reuters* (2019). Retrieved February 10, 2020 from https://www.reuters.com/article/us-tiktok-cfius-exclusive/exclusive-u-s-opens-national-security-investigation-into-tiktok-sources-idUSKBN1XB4IL.

[25] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.

[26] Wendy Hall, Ramine Tinati, and Will Jennings. 2018. From Brexit to Trump: Social media's role in democracy. *Computer* 51, 1 (2018), 18–27.

[27] Simon Hegelich and Dietmar Janetzko. 2016. Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet. In *Tenth International AAAI Conference on Web and Social Media*. The AAAI Press.


[4] https://github.com/JuanCarlosCSE/TikTok/blob/master/tiktok_ids.txt

[28] Philip N Howard and SC Woolley. 2016. Political communication, computational propaganda, and autonomous agents-Introduction. *International Journal of Communication* 10, 2016 (2016).

[29] Gary King, Benjamin Schneer, and Ariel White. 2017. How the news media activate public expression and influence national agendas. *Science* 358, 6364 (2017), 776–780.

[30] Daniel Kreiss and Shannon C McGregor. 2018. Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 US presidential cycle. *Political Communication* 35, 2 (2018), 155–177.

[31] Shamanth Kumar, Reza Zafarani, and Huan Liu. 2011. Understanding user migration patterns in social media. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*. The AAAI Press.

[32] Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic authority: the ethics, politics, and economics of algorithms that interpret, decide, and manage. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1057–1062.

[33] Momin M Malik and Jürgen Pfeffer. 2016. Identifying platform effects in social media data. In *Tenth International AAAI Conference on Web and Social Media*. The AAAI Press.

[34] Andrew Meola. 2020. Analyzing Tik Tok user growth and usage patterns in 2020. *Business Insider* (2020). Retrieved February 10, 2020 from https://www.businessinsider.com/tiktok-marketing-trends-predictions-2020.

[35] Jeremy Merril. 2016. Liberal, Moderate or Conservative? See How Facebook Labels You. *The New York Times* (2016). Retrieved February 10, 2020 from https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html.

[36] Raphael Ottoni, Evandro Cunha, Gabriel Magno, Pedro Bernardina, Wagner Meira Jr, and Virgílio Almeida. 2018. Analyzing right-wing YouTube channels: Hate, violence and discrimination. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 323–332.

[37] Laura Hazard Owen. 2019. One year in, Facebook's big algorithm change has spurred an angry, Fox News-dominated - and very engaged! - News Feed. *NiemanLab* (2019). Retrieved February 10, 2020 from https://www.niemanlab.org/2019/03/one-year-in-facebooks-big-algorithm-change-has-spurred-an-angry-fox-news-dominated-and-very-engaged-news-feed/.

[38] Orestis Papakyriakopoulos, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. 2020. Bias in word embeddings. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, 446–457.

[39] Orestis Papakyriakopoulos, Simon Hegelich, Morteza Shahrezaye, and Juan Carlos Medina Serrano. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society* 5, 2 (2018).

[40] Orestis Papakyriakopoulos, Juan Carlos Medina Serrano, and Simon Hegelich. 2020. Political communication on social media: A tale of hyperactive users and bias in recommender systems. *Online Social Networks and Media* 15 (2020).

[41] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira Jr. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, 131–141.

[42] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. ACM, 399–408.

[43] Juan Carlos Medina Serrano, Simon Hegelich, Morteza Shahrezaye, and Orestis Papakyriakopoulos. 2018. *Social Media Report: The 2017 German Federal Elections*. TUM University Press.

[44] Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos, and Simon Hegelich. 2019. The Rise of Germany's AfD: A Social Media Analysis. In *Proceedings of the 10th International Conference on Social Media and Society*. ACM, 214–223.

[45] Stefan Stieglitz and Linh Dang-Xuan. 2012. Political communication and influence through microblogging–An empirical analysis of sentiment in Twitter messages and retweet behavior. In *2012 45th Hawaii International Conference on System Sciences*. IEEE, 3500–3509.

[46] Stefan Stieglitz and Linh Dang-Xuan. 2013. Social media and political communication: a social media analytics framework. *Social network analysis and mining* 3, 4 (2013), 1277–1291.

[47] Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political communication* 35, 1 (2018), 50–74.

[48] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: A review of the scientific literature (March 19, 2018)* (2018).

[49] Zeynep Tufekci and Christopher Wilson. 2012. Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of communication* 62, 2 (2012), 363–379.

[50] Onur Varol, Emilio Ferrara, Christine L Ogan, Filippo Menczer, and Alessandro Flammini. 2014. Evolution of online user behavior during a social upheaval. In *Proceedings of the 2014 ACM conference on Web science*. ACM, 81–90.

[51] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[52] Chris Wells, Dhavan V Shah, Jon C Pevehouse, JungHwan Yang, Ayellet Pelled, Frederick Boehm, Josephine Lukito, Shreenita Ghosh, and Jessica L Schmidt. 2016. How Trump drove coverage to the nomination: Hybrid media campaigning. *Political Communication* 33, 4 (2016), 669–676.

[53] Katie Williams. 2020. TikTok Was Installed More Than 738 Million Times in 2019, 44% of Its All-Time Downloads. *Sensor Tower* (2020). Retrieved February 10, 2020 from https://sensortower.com/blog/tiktok-revenue-downloads-2019.

# 6 Misinformation and Conspiracy Theories

## 6.1 NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube

### Authors

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich

### In

Proceedings of the 1st Workshop on NLP for COVID-19 at ACL. Association for Computational Linguistics. ACL Anthology (2020).
URL: https://www.aclweb.org/anthology/2020.nlpcovid19-acl.17

### Abstract

We present a simple NLP methodology for detecting COVID-19 misinformation videos on YouTube by leveraging user comments. We use transfer learning pre-trained models to generate a multi-label classifier that can categorize conspiratorial content. We use the percentage of misinformation comments on each video as a new feature for video classification. We show that the inclusion of this feature in simple models yields an accuracy of up to 82.2%. Furthermore, we verify the significance of the feature by performing a Bayesian analysis. Finally, we show that adding the first hundred comments as tf-idf features increases the video classifier accuracy by up to 89.4%.

### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube

**Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich**
Technical University of Munich, Germany
{juan.medina, orestis.p}@tum.de, simon.hegelich@hfp.tum.de

## Abstract

We present a simple NLP methodology for detecting COVID-19 misinformation videos on YouTube by leveraging user comments. We use transfer learning pre-trained models to generate a multi-label classifier that can categorize conspiratorial content. We use the percentage of misinformation comments on each video as a new feature for video classification. We show that the inclusion of this feature in simple models yields an accuracy of up to 82.2%. Furthermore, we verify the significance of the feature by performing a Bayesian analysis. Finally, we show that adding the first hundred comments as tf-idf features increases the video classifier accuracy by up to 89.4%.

## 1 Introduction

The COVID-19 health crisis was accompanied by a misinfodemic: The limited knowledge on the nature and origin of the virus gave ample space for the emergence of conspiracy theories, which were diffused on YouTube, and online social networks. Although YouTube accelerated attempts to detect and filter related misinformation, it yielded moderate results (Li et al., 2020; Frenkel et al., 2020).

In this study, we present a simple NLP-based methodology that can support fact checkers in detecting COVID-19 misinformation on YouTube. Instead of training models on the videos themselves and predicting their nature, we exploit the vast amount of available comments on each YouTube video and extract features that can be used in misinformation detection. Our methodology comes with the advantage that labeling comments is simpler and faster than video labeling. Additionally, no complex neural architecture is needed for the classification of videos.

Our study provides the following contributions:

- We create a multi-label classifier based on transfer learning that can detect conspiracy-laden comments. We find that misinformation videos contain a significantly higher proportion of conspiratorial comments.
- Based on this information, we use the percentage of conspiracy comments as feature for the detection of COVID-19 misinformation videos. We verify its efficiency by deploying simple machine learning models for misinformation detection. We employ the videos' title and the first 100 comments to validate feature significance.
- We show that including the first hundred comments as tf-idf features in the classifier increases accuracy from 82.2% to 89.4%.

## 2 Related Work

Previous research studies have extensively investigated the possibilities and limits of NLP for detecting misinformation. Researchers have provided theoretical frameworks for understanding the lingual and contextual properties of various types of misinformation, such as rumors, false news, and propaganda (Li et al., 2019; Thorne and Vlachos, 2018; Rubin et al.; Zhou and Zafarani, 2018). Given the general difficulty in detecting misinformation, scientists have also developed dedicated benchmark datasets to evaluate the effectiveness of NLP architectures in misinformation-related classification tasks (Pérez-Rosas et al., 2018; Hanselowski et al., 2018). Given the vast amount of misinformation appearing in online social networks, various research studies propose case-specific NLP methodologies for tracing misinformation. For example, Della Vedova et al. (2018) and Popat et al. (2018) combined lingual properties of articles and other meta-data for the detection of false news. Volkova et al. (2017), Qazvinian et al. (2011) and Kumar and Carley (2019) created special architectures that take into consideration the microblogging structure of online social networks, while De Sarkar et al. (2018) and Gupta et al. (2019) exploited sentence-level semantics for misinformation detection.

Despite the deployment of such architectures for fact checking, locating malicious content and promptly removing it remains an open challenge (Gillespie, 2018; Roberts, 2019). In the case of COVID-19 misinformation, a large share of conspiratorial contents remain online on YouTube and other platforms, influencing the public despite content moderation practices (Li et al., 2020; Frenkel et al., 2020; Ferrara, 2020). Given this, it is important to develop case-specific NLP tools that can assist policymakers and researchers in the process of detecting COVID-19 misinformation and managing it accordingly. Towards this end, we illustrate how NLP-based feature extraction (Shu et al., 2017; Jiang et al., 2020) based on user comments can be effectively used for this task. User comment data has been employed to annotate social media objects (Momeni et al., 2013), infer the political leaning of news articles (Park et al., 2011), and to predict popularity (Kim et al., 2016). Previous studies explicitly employed comments as proxies for video content classification (Huang et al., 2010; Filippova and Hall, 2011; Eickhoff et al., 2013; Doğruöz et al., 2017). However, only Jiang and Wilson (2018) have analyzed user content to identify misinformation. However, they focused on linguistic signals and concluded that users' comments were not strong signals for detecting misinformation.

## 3 Methodology and Experiments

### 3.1 Dataset

The first step of the study consisted of obtaining a set of YouTube videos that included either misinformation or debunking content. We decided to search for YouTube videos through user-generated content on social media platforms. For this, we queried the Pushshift Reddit API (Baumgartner et al., 2020), and Crowdtangle's historical data of public Facebook posts (Silverman, 2019) using the query "COVID-19 OR coronavirus". Additionally, we downloaded the COVID-19 Twitter dataset developed by Chen et al. (2020). The total dataset included over 85 million posts generated between January and April 2020. We significantly reduced this dataset by querying the posts with "biowarfare OR biological weapon OR bioweapon OR man-made OR human origin". From the remaining posts, we extracted and expanded the URLs. We identified 1,672 unique YouTube videos. 10% of these videos had been blocked by YouTube as of April 2020. For the rest of the videos, we watched them, excluded the non-English videos, and manually labeled them as either misinformation, factual, or neither. To label a video as misinformation, we validated that its message was conveying with certainty a conspiracy theory regarding the origin of the coronavirus, as a man-made bioweapon or being caused by 5G. We did not classify videos that questioned its origin but showed no certainty about a hoax (which included well-known and verified news media videos) as misinformation. We classified as factual those videos that included debunking of conspiracy theories or presented scientific results on the origins and causes of COVID-19. We labeled the rest of the videos as neither. Two of the authors (JCMS, OP) performed the labeling procedure independently. For the cases where the labels did not agree, the third author was consulted (SH).

Afterward, we collected the comments on both misinformation and factual videos using YouTube's Data API[1]. For this study, we only included videos with more than twenty comments. The final dataset consisted of 113 misinformation and 67 factual videos, with 32,273 and 119,294 total comments respectively. We selected a ten percent random sample of the comments from the misinformation videos and proceeded to label them. This labeling procedure was performed in the same manner as the video classification to ensure data quality. For each comment, we collected two labels. First, we gave a label if the comment expressed agreement (1) or not (0). Agreement comments included comments such as "this is the video I was looking for", or "save and share this video before YouTube puts it down". The second label considered if comments amplified misinformation with a conspiracy theory/misinformation comment (1) or without one (0). Comments that questioned the conspiracies (such as "could it be a bioweapon?") were not labeled as misinformation. 19.7% of the comments in the sample were labeled as conspiracy comment and 12.5% as agreement comment. Only 2.2% of the comments were classified as both agreement and conspiratorial. Although both agreement and conspiracy labeled comments express the same message of believing in the misinformation

[1]https://developers.google.com/youtube/v3

content from the videos, we decided to keep them apart due to their different linguistic properties. To compare the collection of agree-labeled comments and conspiracy-labeled comments, we tokenized and created a bag-of-words model. 19.4% of the processed tokens appear on both collections. However, only 1.95% of the tokens have more than four occurrences in the two collections. We applied $\chi^2$ tests for each of these remaining words and observed that 50% occur in significantly different proportions. In the end, only 0.96% of the vocabulary has a significant similar number of occurrences in the two datasets. The YouTube comments dataset without user data can be accessed in this GitHub repository[2], alongside a Google Colab notebook with the code.

### 3.2 Classification of Users' Comments

We first performed a multi-label classification on the 10% sample of the misinformation videos' comments. We split the annotated data into training (80%) and test (20%) datasets. We employed state-of-the-art neural transfer learning for the classification by fine-tuning three pre-trained models: XLNet base (Yang et al., 2019), BERT base (Devlin et al., 2018) and RoBERTa base (Liu et al., 2019). The fine-tuning consists of initializing the model's pre-trained weights and re-training on labeled data. We ran the models for four epochs using the same hyperparameters as the base models. For the experiments, we used 0.5 as a decision threshold. Additionally, we trained two simpler models as baselines: a logistic regression model using LIWC's lexicon-derived frequencies (Tausczik and Pennebaker, 2010) as features, and a multinomial Naive Bayes model using bag-of-words vectors as features. Table 1 shows the average micro-$F_1$ scores for the three transformer models after performing the fine-tuning five times. RoBERTa

[2]https://github.com/JuanCarlosCSE/YouTube_misinfo

| | Agree | | Conspiracy | |
|---|---|---|---|---|
| | Train | Test | Train | Test |
| **LIWC** | 88.7 | 88.6 | 81 | 78.2 |
| **NB** | 94.2 | 82.4 | 94.3 | 78.8 |
| **XLNet** | 97±0.1 | 93.1±0.3 | 93.9±0.5 | 84.8±0.6 |
| **BERT** | **98.5**±0.1 | 93.3±0.5 | 96.3±0.3 | 83.8±0.9 |
| **RoBERTa** | 98.1±0.2 | **93.9**±0.4 | **96.4**±0.3 | **86.7**±0.5 |

Table 1: Train and test micro $F_1$ scores (mean and standard deviation) from multi-label classification models: LIWC with logistic regression and Naive Bayes as baselines, and three transformer models with five runs.

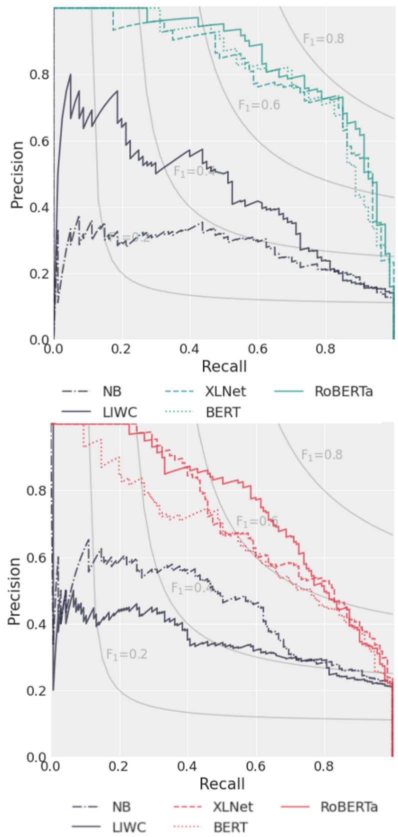


Figure 1: Precision and recall curves for binary $F_1$ scores for the conspiracy (upper figure) and agreement (lower figure) label. The plot shows the results for three neural-transfer classifiers.

is the best performing model for the training and test dataset on the conspiracy classification as for the test data on the agreement label. BERT is the best performing model only for the training data on the agree label. The three transformer models outperform the baseline models. This predictive superiority is more evident in the precision-recall curves (with corresponding binary-$F_1$ scores) of the five models on the test data (Figure 1).

We employed the fine-tuned RoBERTa model to predict the labels of the remaining comments from the misinformation and factual videos. We then calculated the *percentage of conspiracy comments* per video. We also obtained this percentage for the agreement label. Figure 2 shows the resulting density distributions from misinformation and factual videos. We observed a difference between the distributions from the two types of videos. We confirmed this by performing Welch's t-test for independent samples. For the conspiracy comments percentage, the t-test was significant (p<0.000), indicating that the samples came from different distributions. The t-test was not significant for the agreement percentage (p>0.1).

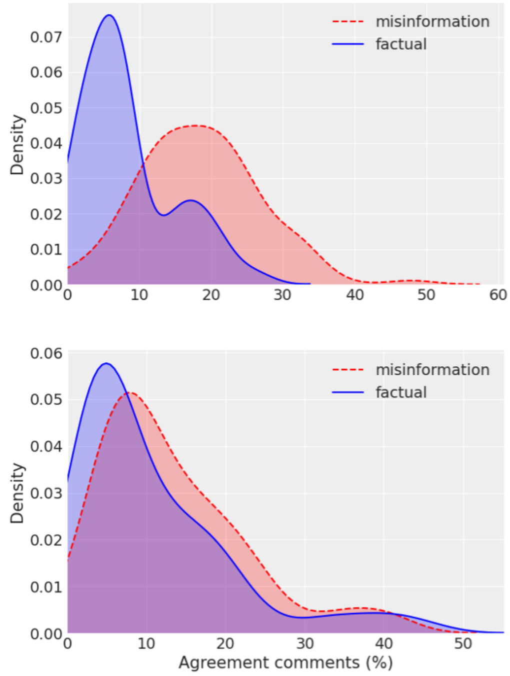


Figure 2: Probability densities of misinformation and factual videos regarding the percentage of conspiratorial comments (top) agreement comments (bottom).

### 3.3 Classification of YouTube Videos

The next step consisted of classifying the set of YouTube videos to detect misinformation. For this, we employed the percentage of conspiracy comments of each video as a feature. Additionally, we extracted content features from the videos' titles and from the raw first hundred comments per video (or all the comments for videos with fewer than 100 comments). For this, we preprocessed the titles and comments with tokenization, removal of stopwords, and usage of the standard term frequency-inverse document (tf-idf) weighting for word frequencies to create a document term matrix, whose columns serve as input features. We selected six feature settings for our experiments: each of the set of features alone and the three possible combination between them . For each setting, we employed three classification models: logistic regression, linear support vector machine (SVM), and random forest. We performed 10-fold cross-validation and report the mean accuracy in Table 2. We avoided grid search to find better hyperparameters as we did not have a test dataset. We observe that the SVM model has the highest accuracy for all the settings except for one. The conspiracy feature

| | LR | SVM | RF |
|---|---|---|---|
| **title** | 62.7 | 62.7 | 62.7 |
| **conspiracy %** | 62.7 | **81.1** | 72.2 |
| **comments** | 66.7 | **83.9** | 82.8 |
| **title + conspiracy %** | 64.4 | 77.7 | **82.2** |
| **comments + conspiracy %** | 73.3 | **89.4** | 84.44 |
| **all** | 73.3 | **84.4** | 82.7 |

Table 2: Classification accuracy for logistic regression, linear support vector machines, and random forest models for six feature settings. Results show the average of 10-k cross-validation.

alone achieves an accuracy of 81.1. Using the tf-idf comment features the accuracy is slightly better with 83.9. However, the conspiracy feature and comments combined achieve the highest accuracy of 89.4. We observe that the models with all the features combined have lower accuracy than the models omitting the title features. This may explain that the title is not a good feature. Using the title feature alone does not improve the baseline accuracy of 62.7. Interestingly, the accuracy for the best model is still high (85.5%) when taking into consideration only videos with less than 100 comments. This implies that our methodology is appropriate for the early detection of misinformation videos.

### 3.4 Bayesian Modeling

To find the statistical validity of the conspiracy percentage feature, we turned to Bayesian modeling as it allows us to obtain the full posterior distribution of feature coefficients. We performed inference on three Bayesian logistic regression models using a Hamiltonian Monte Carlo solver. A simple model considered only the conspiracy percentage feature. A second model included this feature and the ten most relevant word features from the random forest model trained only on the title and conspiracy percentage. A third model included the conspiracy feature, and the top ten most relevant words from the linear SVM trained on the conspiracy feature and the first 100 comments. The first column of Table 3 and 4 shows the importance of each of the features in the random forest and linear SVM model, respectively. The two tables also show the statistics of the posterior probability distributions of the model coefficients: the mean, standard deviation, and the 1% and 99% quantiles. For the three models, the coefficients distribution converged (the $\widehat{R}$ diagnostic (Vehtari et al., 2019) was equal to one). We specifically selected logistic regression models for their interpretability. We observe that for the model based on the title word features, the posterior distribution of the conspiracy percentage feature coefficient is the only one that does not include zero in its 98% highest posterior density interval (Table 3). Although this is not equivalent to traditional p-values, it conveys significance in a Bayesian setting. The model based on the 100 comments word features (Table 4) maintains the conspiracy feature as significant. However, three coefficients from the word features also avoid zero in their 98% interval. The model's coefficients are negative for *covid19* and *lab*, and positive for *god*.

Finally, we compare the three Bayesian models using the WAIC information criteria, which estimates out-of-sample expectation and corrects for the effective number of parameters to avoid overfitting (Watanabe and Opper, 2010). Figure 3 shows the resulting deviance of the three models. We observe that the second model is slightly better than the simple model. However, the differences are included in the standard error of the title words feature model. This is not true for the simple model and the model including the comments features. In this case, the full model outperforms the model based only on the conspiracy feature. This indicates that there is important information in the videos' first hundred comments that is not explained by the conspiracy percentage feature on its own.

## 4 Discussion

We have leveraged large quantities of user comments to extract a simple feature that is effective in predicting misinformation videos. Given that the classifier is also accurate for videos with few comments, it can be used for online learning. For example, the user comments of videos containing *coronavirus* can be tracked and classified as they are posted. High levels of conspiracy comments could then indicate that the video includes misinformation claims. For this to work, it is not necessary to have a conspiracy classifier with perfect accuracy given that the percentage of conspiracy comments feature is based on aggregating the classification results from all the comments. An improved classifier would be able to define a threshold that allows a balanced number of false positives and true negatives. The average percentage of conspiratorial comments would be maintained, irrespective of the wrong classifications. On the other hand, the accuracy of the video classifier is more critical. We found that using simple classifiers on the raw content of the videos' first 100 comments significantly improves the accuracy of misinformation video detection from 82.2 to 89.4. However, in large-scale settings, it may be prohibitive to store the raw comments and continuously perform batch classification. In contrast, the conspiracy percentage feature only requires storing one conspiracy comment counter per video. Future research could leverage the video content to increase the classifier accuracy. The detection of misinformation on social media remains an open challenge, and further research is needed to understand how the COVID-19 misinfodemic spread to prevent future ones.

| | RF | mean | SD | 1% | 99% |
|---|---|---|---|---|---|
| conspiracy % | 19.2 | **28.25** | 4.8 | 18.19 | 39.94 |
| *coronavirus* | 2.95 | -7.45 | 3.4 | -15.57 | 0.01 |
| *covid19* | 2.81 | -5.17 | 2.4 | -11.08 | 0.10 |
| *china* | 1.42 | -4.28 | 3 | -11.23 | 2.63 |
| *man* | 1.24 | -6.04 | 2.8 | -12.25 | 0.52 |
| *bioweapon* | 1.24 | 4.81 | 5.5 | -6.40 | 19.32 |
| *conspiracy* | 1.1 | -4.24 | 3.7 | -13.96 | 3.72 |
| *new* | 1.03 | -5.13 | 5.4 | -18.93 | 6.39 |
| *update* | 0.87 | -0.15 | 2.5 | -6.57 | 5.69 |
| *cases* | 0.83 | -12.37 | 6.3 | -26.75 | 2.10 |
| *outbreak* | 0.72 | -1.25 | 2.9 | -8.31 | 5.66 |

Table 3: Top eleven features from the random forest model with the conspiracy and title as feature with the statistics of the coefficients' posterior probability distributions. The first column shows the percentage of feature importance.

| | svm | mean | SD | 1% | 99% |
|---|---|---|---|---|---|
| conspiracy % | 2.82 | **34.96** | 6.2 | 20.56 | 50.09 |
| *virus* | 0.93 | -6.70 | 5.3 | -19.64 | 4.82 |
| *covid19* | 0.84 | **-28.8** | 10 | -54.33 | -6.20 |
| *god* | 0.75 | **19.29** | 7.6 | 3.39 | 37.54 |
| *allah* | 0.73 | -40.09 | 26 | -103.18 | 1.32 |
| *china* | 0.72 | -4.64 | 3.9 | -14.60 | 3.76 |
| *gates* | 0.69 | 3.39 | 16 | -32.39 | 42.94 |
| *amir* | 0.68 | -8.57 | 6.6 | -24.66 | 5.81 |
| *lab* | 0.68 | **-20.70** | 8.2 | -40.57 | -2.28 |
| *cases* | 0.66 | -22.41 | 14 | -57.26 | 8.48 |
| *trump* | 0.63 | 14.53 | 9.6 | -7.23 | 36.92 |

Table 4: Top eleven features from the SVM model with conspiracy and first 100 comments as features with the statistics of the coefficients' posterior probability distributions. The first column shows the SVM coefficients.

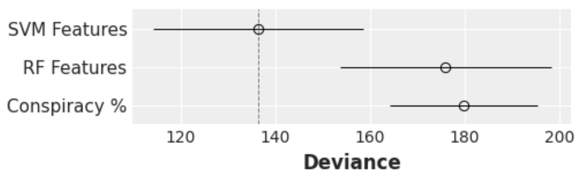


Figure 3: Deviance using WAIC as model selection metric. Black error bars represent the standard error.

## References


Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 830–839.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372*.

Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. Attending sentences to detect satirical fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3371–3380, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Marco L Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro. 2018. Automatic online fake news detection combining content and social signals. In *2018 22nd Conference of Open Innovations Association (FRUCT)*, pages 272–279. IEEE.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

A Seza Doğruöz, Natalia Ponomareva, Sertan Girgin, Reshu Jain, and Christoph Oehler. 2017. Text based user comments as a signal for automatic language identification of online videos. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 374–378.

Carsten Eickhoff, Wen Li, and Arjen P De Vries. 2013. Exploiting user comments for audio-visual content indexing and retrieval. In *European Conference on Information Retrieval*, pages 38–49. Springer.

Emilio Ferrara. 2020. What types of covid-19 conspiracies are populated by Twitter bots? *First Monday*.

Katja Filippova and Keith B Hall. 2011. Improved video categorization from text metadata and user comments. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 835–842.

Sheera Frenkel, Ben Decker, and Davey Alba. 2020. How the 'plandemic' movie and its falsehoods spread widely online.

Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Pankaj Gupta, Khushbu Saxena, Usama Yaseen, Thomas Runkler, and Hinrich Schütze. 2019. Neural architectures for fine-grained propaganda detection in news. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 92–97, Hong Kong, China. Association for Computational Linguistics.

Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance-detection task. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1859–1874, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Chunneng Huang, Tianjun Fu, and Hsinchun Chen. 2010. Text-based video content classification for online video-sharing sites. *Journal of the American Society for Information Science and Technology*, 61(5):891–906.

Shan Jiang, Miriam Metzger, Andrew Flanagin, and Christo Wilson. 2020. Modeling and measuring expressed (dis) belief in (mis) information. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 315–326.

Shan Jiang and Christo Wilson. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–23.

Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8):e0161197.

Sumeet Kumar and Kathleen Carley. 2019. Tree LSTMs with convolution units to predict stance and rumor veracity in social media conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5047–5058, Florence, Italy. Association for Computational Linguistics.

Heidi Oi-Yee Li, Adrian Bailey, David Huynh, and James Chan. 2020. YouTube as a source of information on covid-19: a pandemic of misinformation? *BMJ Global Health*, 5(5).

Quanzhi Li, Qiong Zhang, Luo Si, and Yingchi Liu. 2019. Rumor detection on social media: Datasets, methods and opportunities. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 66–75, Hong Kong, China. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Elaheh Momeni, Claire Cardie, and Myle Ott. 2013. Properties, prediction, and prevalence of useful user-generated comments for descriptive annotation of social media objects. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 113–122.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. DeClarE: Debunking fake news and false claims using evidence-aware deep learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 22–32, Brussels, Belgium. Association for Computational Linguistics.

Vahed Qazvinian, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599.

Sarah T Roberts. 2019. *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. Fake news or truth? using satirical cues to detect potentially misleading news.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

B Silverman. 2019. Crowdtangle for academics and researchers.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3346–3359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2019. Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc. *arXiv preprint arXiv:1903.08008*.

Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 647–653.

Sumio Watanabe and Manfred Opper. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12).

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5753–5763.

Xinyi Zhou and Reza Zafarani. 2018. Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*.

# 7 Discussion

In this last chapter, I provide a short discussion on the major topics I treated in this thesis. First, I provide a summary by revisiting the main framework from Chapter 1 and tie it to the research publications from the previous chapters. Afterward, I present the main implications and challenges from this thesis's contributions. Finally, I give an outline of the most important future work in the area of social media and political communication.

## 7.1 Main Framework (Revisited)

To provide a discussion of the previous papers, I present the connection between them and the framework I proposed in Section 1.2.2. Figure 7.1 shows the same illustration as in the introduction but with an additional labeling that corresponds to the presented publications. I will refer to them by either their short name in the following discussion or the label number from Figure 7.1.

*The Rise of the AfD* (1) and *Exploring Political Ad Libraries* (2) deal with similar parts of the framework. They focus on the political communication from political actors to users. However, each of them examines a different channel of communication on social media; (1) tackles the organic content and (2) the sponsored content (advertisement). A problem I discuss in (2) is that political advertising on Facebook is hard to distinguish from normal organic content. This presents the user with the challenge of differentiating the messages from political actors and, thus, complicating the transparency of the overall communication. Both papers focus on the main German political parties and their online messages on multiple social media platforms. They try to provide a cross-platform analysis to understand the complete political strategies, including the messages that political actors share and the users they want to reach through their market targeting.

In Figure 7.1, (1) also appears on the bad-natured agents part of the framework. This is due to my comparing the percentage of bots that retweeted the content from the German political parties on Twitter. The far-right party, the Alternative für Deutschland, accounted for the highest number of bots retweeting their content. Although the method the study used is far from perfect, the large difference between the AfD and the rest of the political parties is remarkable. Davis et al. [188] found similar results on fake Facebook accounts that share AfD content. This shows that part of the success of the AfD on social media is attributable to fake personas. However, this cannot completely explain their popularity. Other reasons attributable to real users explain AfD's success on social media, as mentioned in (1). This success refers only to organic channels and is not reflected in sponsored activity. In (2), I find that the AfD was the party that spent significantly less money on advertising. The populist party relied on its normal traffic to spread its messages. Additionally, the regional and demographic distribution of users that the party targeted differed significantly from the distributions of the other political parties. This shows AfD's strategy to reach a different part of the population that it perceives as potential voters.

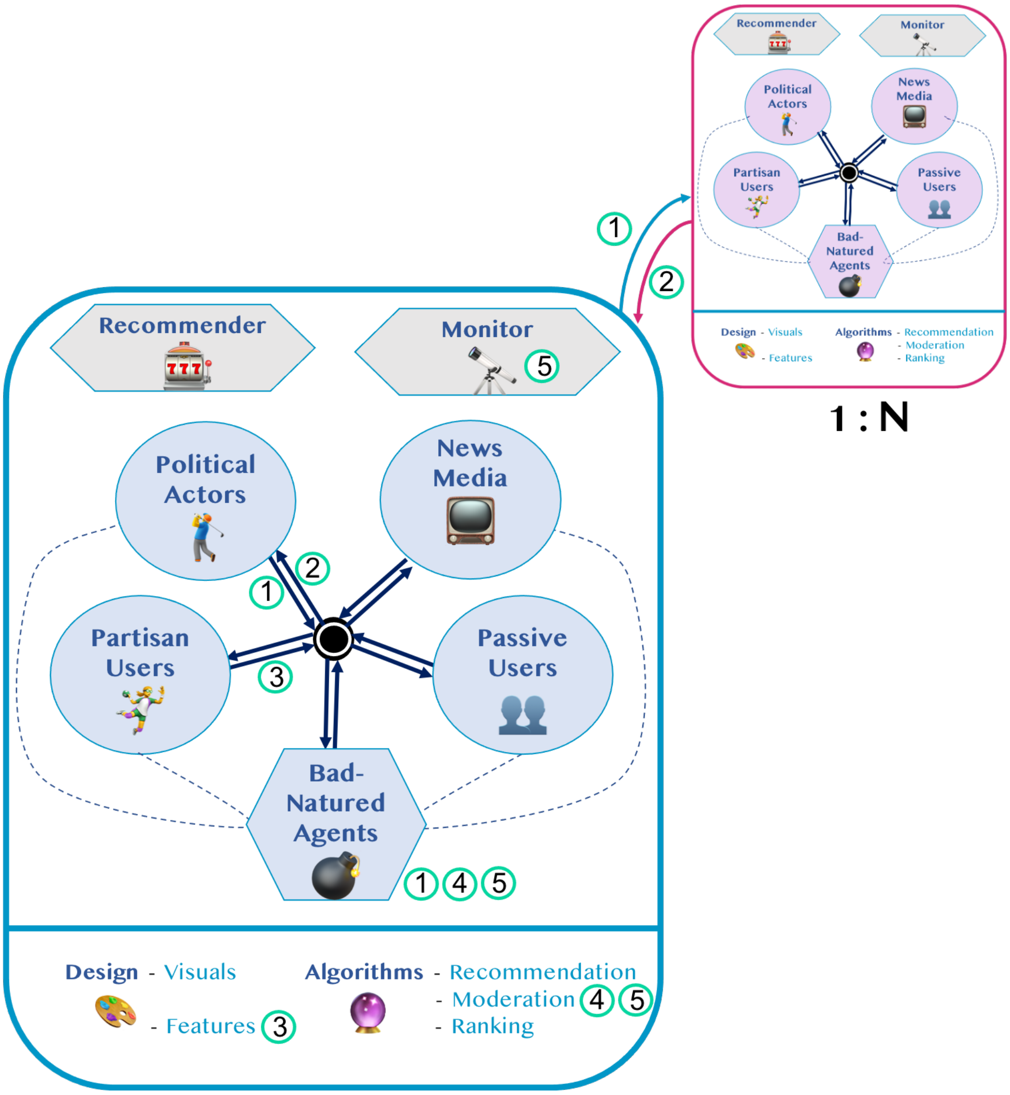


**Figure 7.1:** Main framework of this thesis with a labeling corresponding to the publications that comprise this work. (1) The Rise of the AfD, (2) Exploring Political Ad Libraries, (3) Dancing to the partisan beat, (4) NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube, (5) Coordinated and Suspended Accounts on Twitter (Appendix)

Apart from political actors, partisan users create political content to engage and mobilize other users. In *Dancing to the Partisan Beat* (3), I present the first analysis of political communication on TikTok. The focus centers on U.S. politics, the 2020 U.S. election, and the large number of politically centered videos. TikTok puts the users in the center of the political communication. They share the news in a similar manner with political pundits. This means that, unlike Twitter or Facebook, the users are not passively retweeting and sharing pieces of news; rather, they become the news. They present their political ideologies and interact with other users in a more direct manner that they can in the comment section of other social media platforms. A major focus of this study is on the design of the platform, the features it presents, and how these influence political communication. As Figure 7.1 shows, design plays an important role in the user communication. The study analyses a new TikTok feature, the duet, in which users respond to other videos with a video alongside the original. Thus, a complete communication channel appears on screen. This type of feature is analogous to online political debates. The creativity of users makes political communication entertainment, similar to that of late-night shows. Young users (under 34 years) are the most active on TikTok, and the impact of TikTok politics on the formation of their political identity may be significant. In this ways, TikTok has transformed political communication on social media. As new platforms emerge in which to foster political communication (e.g., Clubhouse, Twitch), the transformation will continue.

Although promising as digital public squares, social media platforms also foster bad-natured agents. These actors try to misinform, spread junk news, and pollute political communication. The urgency of identifying these accounts and containing them has increased, especially when online misinformation translates into offline events that attack democratic institutions. A clear example is the U.S. Capitol siege after months of misinformation on election fraud [189]. Similarly, the COVID-19 infodemic [190] created a difficult panorama for public policy and governmental actions that require citizen compliance. In *NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube* (4), I present a new methodology to identify conspiratorial videos on YouTube, based on analyzing user comments. This methodology works as an early-detection method for videos that are still not popular enough to be taken into consideration for moderation. This is indeed helpful in stopping the spread of videos that can become viral on other social media platforms. Content moderation is one of the three important algorithmic features of the proposed framework, defining what is allowed in political communication. The control of who decides what are the moderation rules is part of the debate on how much should be moderated. This debate has come to public attention especially after the banning of President Donald Trump from Twitter, a few days before he left office [191].

In the Appendix section, I attach the paper *Coordinated and Suspended Accounts on Twitter* (5), which focuses on the same part of the framework as (4). However, this study pays further attention to Twitter's content moderation, examining which accounts it suspended in the months leading up to four general elections (Germany 2017, Mexico 2018, Greece 2019, and the UK 2019). Some of these accounts had a large follower base, implying a possible influence on the political network. However, I did not investigate whether the followers were true accounts or part of a bot farm. Instead, I investigated whether the suspended accounts had engaged in inauthentic coordinated behavior. Those accounts either retweeted or tweeted the same tweets in recognizable time patterns. Although each country has a different political spectrum, I found similar patterns of coordinated partisan accounts. In accordance with my previous findings, most of the coordinated and suspended accounts in the German election supported the AfD.

The main framework of this thesis provides a road map for how to study the different elements of political communication on social media. The ecosystem is changing according to the software and design decisions of a handful of tech companies, defining political communication in the twenty-first century. Even though governments are reacting to this by implementing new legislation to regulate social media companies, doing so accurately faces many challenges. The algorithms and the scale of the data points translate into an environment that is hard to understand, much less regulate. The next sections deal with the challenges and future work that improving the regulation of political discourse requires.

## 7.2 Implications & Challenges

The present work provided insights on how different actors produce and consume political communication on social media. Through an array of case studies, I exemplified the different parts of the main framework that appears in Section 1.2.2. The focus of this study relied on the online political spectrum in Germany and the United States. At the end of Chapters 1 and Chapter 2, I presented the main contributions of this work. As a follow-up, I discuss here three major implications of these contributions and the challenges associated with them.

First, the social media dominance of the far-right party, the AfD, implies that populist rhetoric influences political discourse in the German online ecosystem. A large number of the party's messages are inflammatory and sensationalist, even containing hate speech. Social media companies rewarded them indirectly as they increase user engagement. This means that the one who shouts louder often define political discourse, and a more aggressive political style will win the algorithmic game. In this way, social media does not represent an equitable space for political discourse. This translates into a fractured ecosystem, where users unknowingly take part in political echo chambers that confront them with confirmatory bias [192]. It is an open challenge to regulate social media as spaces for political communication, a challenge that must understand how the recommendation algorithms influence agenda setting and content priming. Although the AfD has not employed advanced microtargeting strategies in social media advertising (nor have the rest of the German political parties), this may change soon in future elections. The current political ad libraries provide some transparency but do not completely provide the political parties' microtargeting techniques. This is especially dangerous if parties decide to use data-driven techniques to send opposite messages to different users.

Second, young citizens are redefining how political communication takes place on online channels. In a matter of months, TikTok went from a dancing and lip-syncing platform to a digital space filled with short political debates. Even though TikTok had no intention of getting involved in politics [193], the demand for political content increased as the platform became popular around the globe. On this platform, partisan users are at the center of the political discourse. They are the news, and they influence other users. Young audiences look up to these *influencers* to form an opinion and replicate what they see. Thus, *influencers* play the role of political gatekeepers, especially as young people are in the process of discovering their political orientation. The advantages of such a politically-laden platform is that political engagement increases, which is a sign of a healthy democracy. However, a large portion of political videos use sarcasm and belittling to respond to users with different views. Hate speech is a challenge on the platform, especially when it is subtle. It is hard for machine learning algorithms to detect this negative behavior. Users are young and sensitive to mocking, which can take a higher toll on their well-being than it does on older users.

Third, natural language processing algorithms can help to detect conspiratorial and misinformation videos based on the users' comments. This implies that identifying harmful video content is possible without processing the video itself. This can save time; video as input is harder to analyze, especially with an immense number of videos being uploaded to a platform. However, the main challenge is that performing content moderation solely based on an algorithm is not possible. The false-positives classifications would affect users sharing factual content. Even for large social media companies, the scale of moderation needed to stop misinformation outpaces the resources that these companies are willing to spend on this problem [194]. Additionally, classifying content as misinformation is not always straightforward, as its definition remains vague. It also depends on multiple factors, and the source of truth may change with time. No matter how good the social media companies become at protecting political communication from harmful and inaccurate content, the problem remains of having a few stakeholders with dominant power over online speech. In the final section, I present the future work necessary to understand and regulate this inequality.

## 7.3 Future work

*I'm less convinced that this [disinformation] is a problem of information systems and increasingly convinced that this is a problem of power and responsibility*

**- Ethan Zuckermann [195]**

Social media owners have become key stakeholders in the political process around the world. This increases the necessity to audit them and regulate them regarding political communication. Ethan Zuckermann is right in pointing out that the disinformation and related problems are not the sole responsibility of the technological system as there are inherent problems in society. Problems such as polarization, bias, and radicalization are mostly potentiated through social media and create a real threat to democratic systems. I am convinced that the solution not only depends on understanding the platforms' algorithms but also requires a larger understanding of the underlying problems.

Unfortunately, the data available is largely insufficient to tackle these problems. As researchers, we have little information on passive users (the only part of the main framework that I did not cover in the case studies presented). Their passive actions remain only at the hand of the technology companies. Questions about their change in beliefs, the pages they visit, or the political campaigns that reached them are unanswerable. Most of the research on social media is based on active users and their interactions. However, they constitute the minority of online users. Additionally, due to privacy concerns, monitoring the political communication of active users on private channels is not possible. Journalists and researchers only obtain access through infiltration. This all means that the percentage of information that can be audited from outside the social media companies is only a fraction of the complete communication, and this data is probably statistically biased.

As mentioned at the beginning of this thesis, the tech companies agreed to more transparency in 2019. One of the largest efforts to share data with researchers came with the announcement of the Social Science One initiative. However, Hegelich [196] argues that the data shared has little value for understanding the real effects of sharing behavior on elections. At the same time, the transparency tools for political advertising create enough data to monitor political advertisement, but not to understand the real targeting schemes. The NYU ad observatory tried to enhance the dataset by crawling personal data through a browser extension, but Facebook stopped their efforts [197]. At the beginning of 2021, Facebook decided to share the data on targeting with selected researchers [198], changing and giving for the first time access to this information. However, it may result in a path similar to the Social Science One cooperation. On the side of content moderation, the tech companies often publish their efforts to contain coordinated inauthentic behavior, fake accounts, or the public accounts they decided to ban from the platform. However, the underlying problem is that even the best content moderation will not stop the demand for misinformation or heated political discussion, and the users will migrate to other platforms.

Future research should focus on understanding the effects of social media on the passive users. This can only succeed with the help of large online field experiments to understand how mobilization, polarization, or political influence change over time. The experiments should also focus on the side-effects of exposure to microtargeted advertising. Combining results in this area with research on the effect of algorithms on users is the only path to understanding the complex ecosystem of social media and its implications for the offline world. This research can help in finding auditing measures that make tech companies responsible for the data algorithms. Creating guidelines for auditing without taking into consideration the data and the real sociological effects would probably prove inefficient. In the following, I mention some notables examples of the research that focus on these points. Luca et al. [199] investigated why readers choose to read clickbait stories over news from respected sources. Brady et al. [200] found that positive social feedback for outrage expressions on social media increased the likelihood of future outrage expressions. Levy [201] performed a field experiment and found that social media algorithms may limit exposure to counter-attitudinal news and, thus, increase polarization. Theocharis et al. [202] found that maintaining a Facebook account had negative consequences for reports of political and civic participation.

At the same time, future research is needed to swiftly study new social media platforms that come into popularity, especially platforms whose ecosystem naturally fosters political communication. Only agile research that adapts to new channels of communication can keep pace with the digital transformation. In my analysis of TikTok, I realized that this platform has fewer echo chambers than others, as the focus of political communication is based on cross-partisan interactions. The “For you” page’s algorithm seems to show videos from different political ideologies. However, this is a subjective perception and not based on experiments. Future research should also compare the political discourse between older and newer social media platforms, to understand the technological factors that determine toxicity, polarization, and bias. This is especially important for platforms mostly consisting of young audiences. My opinion is that young users interacting and expressing their political opinions have a positive effect on democracy. However, platforms show a biased look at how political communication works. Politics is displayed as entertainment, and sometimes it appears that the best way to win an argument is by belittling your opponent or finding alternative facts. A polarized political conversation has many side-effects for young users who deal with depression [203], cyberbullying [204], and harassment [205]. The study of political communication on new social media channels should take these aspects into consideration.

This work shed light on the study of political communication in a multiplatform social media environment. It presented a framework for understanding the different stakeholders and the challenges that face quantifying the interactions between them. It provided key insights into the political landscape in Germany and the United States. The study cases enrich the literature of political discourse on social media. The challenges presented in this last section allow the reader an overview of the challenges that lie ahead in this area of research.

# Bibliography


[1] J. C. M. Serrano, M. Shahrezaye, O. Papakyriakopoulos, and S. Hegelich. The Rise of Germany's AfD: A Social Media Analysis. In *Proceedings of the 10th International Conference on Social Media and Society*, pages 214–223, 2019.

[2] J. C. M. Serrano, O. Papakyriakopoulos, M. Shahrezaye, and S. Hegelich. The Political Dashboard: A Tool for Online Political Transparency. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 983–985, 2020.

[3] J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich. Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections. In *International Conference on Social Media and Society*, pages 111–121, 2020.

[4] J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich. Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. In *12th ACM Conference on Web Science*, pages 257–266, 2020.

[5] J. C. M. Serrano, O. Papakyriakopoulos, and S. Hegelich. NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, 2020.

[6] O. Papakyriakopoulos, M. Shahrezaye, A. Thieltges, J. C. M. Serrano, and S. Hegelich. Social media und microtargeting in Deutschland. *Informatik-Spektrum*, 40(4):327–335, 2017.

[7] O. Papakyriakopoulos, S. Hegelich, M. Shahrezaye, and J. C. M. Serrano. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, 5(2):2053951718811844, 2018.

[8] M. Shahrezaye, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. Estimating the Political Orientation of Twitter Users in Homophilic Networks. In *2019 AAAI Spring Symposium Series*, 2019.

[9] O. Papakyriakopoulos, M. Shahrezaye, J. C. M. Serrano, and S. Hegelich. Distorting Political Communication: The Effect Of Hyperactive Users In Online Social Networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 157–164. IEEE, 2019.

[10] M. Shahrezaye, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. Measuring the Ease of Communication in Bipartite Social Endorsement Networks: A Proxy to Study the Dynamics of Political Polarization. In *Proceedings of the 10th International Conference on Social Media and Society*, pages 158–165. ACM, 2019.

[11] O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. Political communication on social media: A tale of hyperactive users and bias in recommender systems. *Online Social Networks and Media*, 15:100058, 2020.

[12] O. Papakyriakopoulos, S. Hegelich, J. C. M. Serrano, and F. Marco. Bias in Word Embeddings. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 446–457, New York, NY, USA, 2020. Association for Computing Machinery.

[13] O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. The spread of COVID-19 conspiracy theories on social media and the effect of content moderation. *Harvard Kennedy School Misinformation Review*, 2020.

[14] E. O. Wilson. Oxford Essential Quotations, 2016. Retrieved December 20, 2020 from https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00016553.

[15] Wearesocial. Digital in 2020. *https://wearesocial.com/*, 2020. Retrieved January 6, 2021 from https://wearesocial.com/digital-2020.

[16] C. T. Carr and R. A. Hayes. Social media: Defining, developing, and divining. *Atlantic journal of communication*, 23(1):46–65, 2015.

[17] L. Rainie, A. Smith, K. L. Schlozman, H. Brady, and S. Verba. Social media and political engagement. *Pew Internet & American Life Project*, 19:2–13, 2012.

[18] N. Gustafsson. The subtle nature of Facebook politics: Swedish social network site users and political participation. *New Media & Society*, 14(7):1111–1127, 2012.

[19] R. Effing, J. Van Hillegersberg, and T. Huibers. Social media and political participation: are Facebook, Twitter and YouTube democratizing our political systems? In *International conference on electronic participation*, pages 25–35. Springer, 2011.

[20] B. D. Loader and D. Mercea. Networking democracy? Social media innovations in participatory politics: Brian D. Loader and Dan Mercea. In *Social Media and Democracy*, pages 12–21. Routledge, 2012.

[21] N. Fenton and V. Barassi. Alternative media and social networking sites: The politics of individuation and political participation. *The Communication Review*, 14(3):179–196, 2011.

[22] S. Joseph. Social media, political change, and human rights. *BC Int'l & Comp. L. Rev.*, 35:145, 2012.

[23] G. Wolfsfeld, E. Segev, and T. Sheafer. Social media and the Arab Spring: Politics comes first. *The International Journal of Press/Politics*, 18(2):115–137, 2013.

[24] R. Chen and Y. Sakamoto. Perspective matters: Sharing of crisis information in social media. In *2013 46th Hawaii International Conference on System Sciences*, pages 2033–2041. IEEE, 2013.

[25] J. Van Dijck. Facebook as a tool for producing sociality and connectivity. *Television & New Media*, 13(2):160–176, 2012.

[26] D. Linders. From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government information quarterly*, 29(4):446–454, 2012.

[27] M. Muller, K. Ehrlich, T. Matthews, A. Perer, I. Ronen, and I. Guy. Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2815–2824, 2012.

[28] F. Zuiderveen Borgesius, J. Möller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodo, and C. H. de Vreese. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1):82–96, 2018.

[29] E. D. Hersh. *Hacking the electorate: How campaigns perceive voters.* New York: Cambridge University Press, 2015.

[30] D. Kreiss. *Prototype Politics: Technology-Intensive Campaigning and the Data of Democracy.* New York: Oxford University Press, 2016.

[31] P. N. Howard and S. Woolley. Political communication, computational propaganda, and autonomous agents-introduction. *International Journal of Communication*, 10(2016), 2016.

[32] J. A. Tucker, A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*, 2018.

[33] N. Just and M. Latzer. Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2):238–258, 2017.

[34] E. Ruppert, E. Isin, and D. Bigo. Data politics. *Big Data & Society*, 4(2):2053951717717749, 2017.

[35] N. Seaver. Knowing Algorithms. *digitalSTS: A Field Guide for Science & Technology Studies*, page 412, 2019.

[36] E. Bozdag. Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3):209–227, 2013.

[37] D. Lazer. The rise of the social algorithm. *Science*, 348(6239):1090–1091, 2015.

[38] L. Taylor. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2):2053951717736335, 2017.

[39] D. Beer. The social power of algorithms, 2017.

[40] P. N. Howard and M. R. Parks. Social media and political change: Capacity, constraint, and consequence, 2012.

[41] W. L. Bennett. The personalization of politics: Political identity, social media, and changing patterns of participation. *The ANNALS of the American Academy of Political and Social Science*, 644(1):20–39, 2012.

[42] E. Zuckerman. New media, new civics? *Policy & Internet*, 6(2):151–168, 2014.

[43] C. Montag and S. Hegelich. Understanding Detrimental Aspects of Social Media Use: Will the Real Culprits Please Stand Up. *Front. Sociol. 5: 599270*, 2020.

[44] P. Lewis and P. Hilder. Leaked: Cambridge Analytica's blueprint for Trump victory. *The Guardian*, 2018. Retrieved January 10, 2021 from https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory.

[45] C. Gartenberg. What is Facebook? Just ask Mark Zuckerberg. *The Verge*, 2019. Retrieved January 10, 2021 from https://www.theverge.com/2019/3/8/18255269/facebook-mark-zuckerberg-definition-social-media-network-sharing-privacy.

[46] M. D. Kearney, S. C. Chiang, and P. M. Massey. The Twitter origins and evolution of the COVID-19 "plandemic" conspiracy theory. *Harvard Kennedy School Misinformation Review*, 1(3), 2020.

[47] D. G. Lilleker. *Key concepts in political communication*. Sage, 2006.

[48] H. P. Bang. *Governance as social and political communication*. Manchester University Press, 2003.

[49] F. Esser and B. Pfetsch. *Comparing political communication: Theories, cases, and challenges*. Cambridge University Press, 2004.

[50] R. E. Denton and G. C. Woodward. *Political communication in America*. Praeger Publishers, 1990.

[51] B. McNair. *An introduction to political communication*. Taylor & Francis, 2017.

[52] P. M. Coelho, P. P. Correia, and I. G. Medina. Social media: A new way of public and political communication in digital media. *International Journal of Interactive Mobile Technologies (iJIM)*, 11(6):150–157, 2017.

[53] M. Karlsson and J. Åström. Social media and political communication: Innovation and normalisation in parallel. *Journal of Language and Politics*, 17(2):305–323, 2018.

[54] W. L. Bennett and B. Pfetsch. Rethinking political communication in a time of disrupted public spheres. *Journal of communication*, 68(2):243–253, 2018.

[55] P. Iosifidis and M. Wheeler. Modern political communication and web 2.0 in representative democracies. *Javnost-The public*, 25(1-2):110–118, 2018.

[56] P. Van Aelst, J. Strömbäck, T. Aalberg, F. Esser, C. De Vreese, J. Matthes, D. Hopmann, S. Salgado, N. Hubé, et al. Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, 41(1):3–27, 2017.

[57] S. Stier, A. Bleier, H. Lietz, and M. Strohmaier. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political communication*, 35(1):50–74, 2018.

[58] S. Hegelich and M. Shahrezaye. The communication behavior of German MPs on Twitter: Preaching to the converted and attacking opponents. *European Policy Analysis*, 1(2):155–74, 2015.

[59] O. Varol, E. Ferrara, F. Menczer, and A. Flammini. Early detection of promoted campaigns on social media. *EPJ Data Science*, 6(1):13, 2017.

[60] N. Xi, D. Ma, M. Liou, Z. C. Steinert-Threlkeld, J. Anastasopoulos, and J. Joo. Understanding the Political Ideology of Legislators from Social Media Images. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 726–737, 2020.

[61] S. Lewandowsky, M. Jetter, and U. K. Ecker. Using the president's tweets to understand political diversion in the age of social media. *Nature communications*, 11(1):1–12, 2020.

[62] I. Himelboim, R. W. Lariscy, S. F. Tinkham, and K. D. Sweetser. Social media and online political communication: The role of interpersonal informational trust and openness. *Journal of Broadcasting & Electronic Media*, 56(1):92–115, 2012.

[63] S. Kruikemeier, G. Van Noort, R. Vliegenthart, and C. H. De Vreese. Getting closer: The effects of personalized and interactive online political communication. *European journal of communication*, 28(1):53–66, 2013.

[64] K. Holt, A. Shehata, J. Strömbäck, and E. Ljungberg. Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveller? *European Journal of Communication*, 28(1):19–34, 2013.

[65] S. Shugars and N. Beauchamp. Why keep arguing? Predicting engagement in political conversations online. *Sage Open*, 9(1):2158244019828850, 2019.

[66] M. Barnidge. Exposure to political disagreement in social media versus face-to-face and anonymous online settings. *Political Communication*, 34(2):302–321, 2017.

[67] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

[68] P. Barberá. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis*, 23(1):76–91, 2015.

[69] D. Ruths and J. Pfeffer. Social media for large studies of behavior. *Science*, 346(6213):1063–1064, 2014.

[70] P. Barberá and Z. C. Steinert-Threlkeld. How to Use Social Media Data for Political Science Research. *The SAGE Handbook of Research Methods in Political Science and International Relations*, page 404, 2020.

[71] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[72] P. Barberá and G. Rivero. Understanding the political representativeness of Twitter users. *Social Science Computer Review*, 33(6):712–729, 2015.

[73] M. Klašnja, P. Barberá, N. Beauchamp, J. Nagler, and J. Tucker. Measuring public opinion with social media data. In *The Oxford Handbook of Polling and Survey Methods*. 2017.

[74] N. Beauchamp. Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61(2):490–503, 2017.

[75] S. Stieglitz and L. Dang-Xuan. Social media and political communication: A social media analytics framework. *Social network analysis and mining*, 3(4):1277–1291, 2013.

[76] J. Pfeffer. Visualization of political networks. *The Oxford Handbook of Political Networks*, 1:277, 2017.

[77] B. Kalsnes. The power of likes: Social media logic and political communication. 2016.

[78] H. D. Lasswell. *Propaganda technique in world war I*, volume 170. MIT press, 1971.

[79] P. Dahlgren. The internet, public spheres, and political communication: Dispersion and deliberation. *Political communication*, 22(2):147–162, 2005.

[80] D. V. Shah, D. M. McLeod, H. Rojas, J. Cho, M. W. Wagner, and L. A. Friedland. Revising the communication mediation model for a new political communication ecology. *Human Communication Research*, 43(4):491–504, 2017.

[81] R. Karlsen. Followers are opinion leaders: The role of people in the flow of political communication on and beyond social networking sites. *European Journal of Communication*, 30(3):301–318, 2015.

[82] P. J. Shoemaker, T. P. Vos, and S. D. Reese. Journalists as gatekeepers. *The handbook of journalism studies*, 73, 2009.

[83] J. Guberman, C. Schmitz, and L. Hemphill. Quantifying toxicity and verbal violence on Twitter. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, pages 277–280, 2016.

[84] S. Stieglitz, F. Brachten, B. Ross, and A.-K. Jung. Do social bots dream of electric sheep? A categorisation of social media bot accounts. *arXiv preprint arXiv:1710.04044*, 2017.

[85] L. Reichart Smith, K. D. Smith, and M. Blazka. Follow Me, What's the Harm: Considerations of Catfishing and Utilizing Fake Online Personas on Social Media. *J. Legal Aspects Sport*, 27:32, 2017.

[86] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. M. Menczer. Detecting and tracking political abuse in social media. In *Fifth international AAAI conference on weblogs and social media*. Citeseer, 2011.

[87] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer. Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*, pages 273–274, 2016.

[88] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[89] A. Rauchfleisch and J. Kaiser. The false positive problem of automatic bot detection in social science research. *Berkman Klein Center Research Publication*, (2020-3), 2020.

[90] S. Thielman. Facebook news selection is in hands of editors not algorithms, documents show. *The Guardian*, 2016. Retrieved December 30, 2020 from https://www.theguardian.com/technology/2016/may/12/facebook-trending-news-leaked-documents-editor-guidelines.

[91] K. Roose. Facebook reverses postelection algorithm changes that boosted news from authoritative sources. *The New York Times*, 2020. Retrieved December 20, 2020 from https://www.nytimes.com/2020/12/16/technology/facebook-reverses-postelection-algorithm-changes-that-boosted-news-from-authoritative-sources.html.

[92] M. E. McCombs and D. L. Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.

[93] I. Rogstad. Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. *Journal of Information Technology & Politics*, 13(2):142–158, 2016.

[94] E. G. Nilsson. Design patterns for user interface for mobile applications. *Advances in engineering software*, 40(12):1318–1328, 2009.

[95] J. R. Hayes. User interface design for online social media. 2014.

[96] B. Shneiderman and C. Plaisant. *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education India, 2010.

[97] D. Benyon. *Designing interactive systems: A comprehensive guide to HCI, UX and interaction design*. Pearson Edinburgh, 2014.

[98] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296, 2010.

[99] K. Conger. Twitter Has Labeled 38% of Trump's Tweets Since Tuesday. *The New York Times*, 2020. Retrieved December 20, 2020 from https://www.nytimes.com/2020/11/05/technology/donald-trump-twitter.html.

[100] T. Gillespie. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press, 2018.

[101] L. Bode. Political news in the news feed: Learning politics from social media. *Mass communication and society*, 19(1):24–48, 2016.

[102] U. Klinger and J. Svensson. The emergence of network media logic in political communication: A theoretical approach. *New media & society*, 17(8):1241–1257, 2015.

[103] M. Shahrezaye. *Understanding big social networks: Applied methods for computational social science.* PhD thesis, Technische Universität München, 2019.

[104] O. Papakyriakopoulos. *Political machines: machine learning for understanding the politics of social machines.* PhD thesis, Technische Universität München, 2020.

[105] S. Issenberg. *The victory lab: The secret science of winning campaigns.* Crown, 2012.

[106] M. M. Franz and T. N. Ridout. Political advertising and persuasion in the 2004 and 2008 presidential elections. *American Politics Research*, 38(2):303–329, 2010.

[107] C. Cadwalladr. The great British Brexit robbery: how our democracy was hijacked. *The Guardian*, 2017. Retrieved December 15, 2020 from https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.

[108] K. Endres. The accuracy of microtargeted policy positions. *PS: Political Science & Politics*, 49(4):771–774, 2016.

[109] S. Kruikemeier, M. Sezgin, and S. C. Boerman. Political microtargeting: Relationship between personalized advertising on Facebook and voters' responses. *Cyberpsychology, Behavior, and Social Networking*, 19(6):367–372, 2016.

[110] B. Bodó, N. Helberger, and C. H. de Vreese. Political micro-targeting: A Manchurian candidate or just a dark horse? *Internet Policy Review*, 6(4), 2017.

[111] B. C. Schipper and H. Woo. Political awareness, microtargeting of voters, and negative electoral campaigning. *Microtargeting of Voters, and Negative Electoral Campaigning (May 2, 2017)*, 2017.

[112] B. Zarouali, T. Dobber, G. De Pauw, and C. de Vreese. Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media. *Communication Research*, 2020.

[113] E. Ferrara, O. Varol, C. A. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Commun. ACM*, 59(7):96–104, 2016.

[114] S. Hegelich and D. Janetzko. Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet. In *ICWSM*, pages 579–582. AAAI Press, 2016.

[115] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer. The spread of low-credibility content by social bots. *Nature communications*, 9(1):4787, 2018.

[116] P. N. Howard and B. Kollanyi. Bots, #Strongerin, and #Brexit: Computational Propaganda during the UK-EU Referendum. *arXiv preprint*, arXiv:1606.06356, 2016.

[117] A. Bessi and E. Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. 21, 11 2016.

[118] F. Brachten, S. Stieglitz, L. Hofeditz, K. Kloppenborg, and A. Reimann. Strategies and Influence of Social Bots in a 2017 German state election - A case study on Twitter. *arXiv preprint*, arXiv:1710.07562, 2017.

[119] Z. Gilani, E. Kochmar, and J. Crowcroft. Classification of Twitter Accounts into Automated Agents and Human Users. In J. Diesner, E. Ferrari, and G. Xu, editors, *ASONAM*, pages 489–496. ACM, 2017.

[120] A. Thieltges, O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich. Effects of Social Bots in the Iran-Debate on Twitter. *arXiv preprint*, arXiv:1805.10105, 2018.

[121] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *ICWSM*, pages 280–289. AAAI Press, 2017.

[122] S. Kudugunta and E. Ferrara. Deep neural networks for bot detection. *Inf. Sci.*, 467:312–322, 2018.

[123] Z. Gilani, R. Farahbakhsh, and J. Crowcroft. Do Bots impact Twitter activity? In R. Barrett, R. Cummings, E. Agichtein, and E. Gabrilovich, editors, *WWW (Companion Volume)*, pages 781–782. ACM, 2017.

[124] A. Badawy, E. Ferrara, and K. Lerman. Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. In U. Brandes, C. Reddy, and A. Tagarelli, editors, *ASONAM*, pages 258–265. IEEE Computer Society, 2018.

[125] A. Spangher, G. Ranade, B. Nushi, A. Fourney, and E. Horvitz. Analysis of Strategy and Spread of Russia-sponsored Content in the US in 2017. *arXiv preprint*, arXiv:1810.10033, 2018.

[126] S. Zannettou, T. Caulfield, E. D. Cristofaro, M. Sirivianos, G. Stringhini, and J. Blackburn. Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. *arXiv preprint*, arXiv:1801.09288, 2018.

[127] D. L. Linvill and P. L. Warren. Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building, 2018.

[128] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[129] V. L. Rubin, Y. Chen, and N. J. Conroy. Deception detection for news: three types of fakes. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, page 83. American Society for Information Science, 2015.

[130] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[131] R. Faris, H. Roberts, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. 2017.

[132] J. C. Medina Serrano, S. Hegelich, M. Shahrezaye, and O. Papakyriakopoulos. *Social Media Report: The 2017 German Federal Elections.* TUM University Press, Munich, Germany, 2018.

[133] A. Hanselowski, A. PVS, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych. A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*, 2018.

[134] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro. Some Like it Hoax: Automated Fake News Detection in Social Networks. *arXiv preprint*, arXiv:1704.07506, 2017.

[135] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explorations*, 19(1):22–36, 2017.

[136] K. Popat, S. Mukherjee, A. Yates, and G. Weikum. DeClarE: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint*, arXiv:1809.06416, 2018.

[137] N. Ruchansky, S. Seo, and Y. Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806. ACM, 2017.

[138] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv preprint*, arXiv:1806.00749, 2018.

[139] J. Zhang, L. Cui, Y. Fu, and F. B. Gouza. Fake news detection with deep diffusive network model. *arXiv preprint*, arXiv:1805.08751, 2018.

[140] K. Shu, H. R. Bernard, and H. Liu. Studying Fake News via Network Analysis: Detection and Mitigation. *arXiv preprint*, arXiv:1804.10233, 2018.

[141] X. Zhou and R. Zafarani. Fake News: A Survey of Research, Detection Methods, and Opportunities. *arXiv preprint*, arXiv:1812.00315, 2018.

[142] S. Zannettou, T. Caulfield, E. De Cristofaro, N. Kourtelris, I. Leontiadis, M. Sirivianos, G. Stringhini, and J. Blackburn. The web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources. In *Proceedings of the 2017 Internet Measurement Conference*, pages 405–417. ACM, 2017.

[143] M. H. Ribeiro, S. Jhaver, S. Zannettou, J. Blackburn, E. De Cristofaro, G. Stringhini, and R. West. Does Platform Migration Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. *arXiv preprint arXiv:2010.10397*, 2020.

[144] S. Zannettou, B. Bradlyn, E. De Cristofaro, H. Kwak, M. Sirivianos, G. Stringini, and J. Blackburn. What is gab: A bastion of free speech or an alt-right echo chamber. In *Companion Proceedings of the The Web Conference 2018*, pages 1007–1014, 2018.

[145] N. Wiener. *The human use of human beings: Cybernetics and society*. Number 320. Da Capo Press, 1988.

[146] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. *arXiv preprint arXiv:1306.5204*, 2013.

[147] A. Bruns. After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11):1544–1566, 2019.

[148] C. B. Williams and G. J. Gulati. Digital advertising expenditures in the 2016 presidential election. *Social Science Computer Review*, 36(4):406–421, 2018.

[149] Facebook. Spending Tracker: US 2020 Presidential Race, 2020. Retrieved December 17, 2020 from https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US.

[150] S. Perez. TikTok explains its ban on political advertising. *TechCrunch*, 2019. Retrieved December 18, 2020 from https://techcrunch.com/2019/10/03/tiktok-explains-its-ban-on-political-advertising/.

[151] J. Koetsier. Massive TikTok Growth: Up 75% This Year, Now 33X More Users Than Nearest Direct Competitor. *Forbes*, 2020. Retrieved December 18, 2020 from https://www.forbes.com/sites/johnkoetsier/2020/09/14/massive-tiktok-growth-up-75-this-year-now-33x-more-users-than-nearest-competitor/.

[152] L. Matsakis. TikTok Finally Explains How the For You Algorithm Works. *Wired*, 2020. Retrieved December 18, 2020 from https://www.wired.com/story/tiktok-finally-explains-for-you-algorithm-works/.

[153] T. Lorenz. The Political Pundits of TikTok. *The New York Times*, 2020. Retrieved December 18, 2020 from https://www.nytimes.com/2020/02/27/style/tiktok-politics-bernie-trump.html.

[154] M. Cohen. Meet The Politician Who Lives on TikTok. *Politico*, 2020. Retrieved December 17, 2020 from https://www.politico.com/news/magazine/2020/08/14/politicians-using-tiktok-matt-little-395620.

[155] N. M. Krause, I. Freiling, B. Beets, and D. Brossard. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research*, pages 1–8, 2020.

[156] A. LaFrance. The Prophecies of Q. *The Atlantic*, 2020. Retrieved December 18, 2020 from https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/.

[157] A. Romano. Twitter released 9 million tweets from one Russian troll farm. Here's what we learned. *Vox*, 2018. Retrieved December 18, 2020 from https://www.vox.com/2018/10/19/17990946/twitter-russian-trolls-bots-election-tampering.

[158] R. M. Alvarez. *Computational Social Science*. Cambridge University Press, 2016.

[159] L. Sloan and A. Quan-Haase. *The SAGE handbook of social media research methods*. Sage, 2017.

[160] K. Imai. *Quantitative social science: An introduction*. Princeton University Press, 2018.

[161] I. Foster, R. Ghani, R. S. Jarmin, F. Kreuter, and J. Lane. *Big Data and Social Science: Data Science Methods and Tools for Research and Practice*. CRC Press, 2020.

[162] F. Meyer. Everything AI?, 2020. Retrieved December 20, 2020 from https://techxplore.com/news/2020-09-ai.html.

[163] A. Gordon. Data science, classification and related methods, 1998.

[164] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Social science. Computational social science. *Science (New York, NY)*, 323(5915):721–723, 2009.

[165] F. Wu and B. A. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences of the United States of America*, 104(45), 2007.

[166] S. Asur, B. A. Huberman, G. Szabó, and C. Wang. Trends in Social Media : Persistence and Decay. *5th International AAAI Conference on Weblogs and Social Media*, 2011.

[167] Statext. Kolmogorov-Smirnov Test, 2013. Retrieved December 12, 2020 from https://www.statext.com/practice/KolmogorovSmirnovT04.php.

[168] OpenDataScience. Logistic Regression with Python, 2019. Retrieved December 12, 2020 from https://medium.com/@ODSC/logistic-regression-with-python-ede39f8573c7.

[169] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[170] W. S. Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[171] A. Burkov. *The hundred-page machine learning book*, volume 1. Andriy Burkov Quebec City, Can., 2019.

[172] J. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.

[173] A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[174] B. Lambert. *A student's guide to Bayesian statistics*. Sage, 2018.

[175] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC press, 2013.

[176] O. Martin. *Bayesian analysis with Python*. Packt Publishing Ltd, 2016.

[177] J. Pearl. *Causality*. Cambridge university press, 2009.

[178] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[179] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[180] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, and M. Welling. Fast collapsed Gibbs sampling for latent Dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 569–577, 2008.

[181] Mathworks. Introducing Deep Learning with MATLAB, 2018.

[182] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[183] Stanford. Feature extraction using convolution, 2013. Retrieved December 20, 2020 from http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution.

[184] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30:5998–6008, 2017.

[185] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805, 2018.

[186] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. Improving language understanding by generative pre-training, 2018.

[187] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*, 2020.

[188] T. Davis, S. Livingston, and M. Hindman. Suspicious Election Campaign Activity on Facebook. School of Media & Public Affairs: The George Washington University, 2019.

[189] L. Leatherby and A. Singhvi. Critical Moments in the Capitol Siege. *The New York Times*, 2021. Retrieved January 17, 2021 from https://www.nytimes.com/interactive/2021/01/15/us/trump-capitol-riot-timeline.html.

[190] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala. The COVID-19 social media infodemic. *Scientific Reports*, 10(1):1–10, 2020.

[191] P. Noor. Should we celebrate Trump's Twitter ban? Five free speech experts weigh in. *The Guardian*, 2021. Retrieved January 10, 2021 from https://www.theguardian.com/us-news/2021/jan/17/trump-twitter-ban-five-free-speech-experts-weigh-in.

[192] K. Weidner, F. Beuk, and A. Bal. Fake news and the willingness to share: A schemer schema and confirmatory bias perspective. *Journal of Product & Brand Management*, 2019.

[193] A. Bhat. TikTok battles to stay 'apolitical' ahead of US election. *Financial Times*, 2020. Retrieved January 24, 2021 from https://www.ft.com/content/ebffedde-8432-41a3-94ae-72289ef82ba5.

[194] S. T. Roberts. *Behind the screen: Content moderation in the shadows of social media*. Yale University Press, 2019.

[195] E. Zuckerman. Fixing disinformation won't save us, 2021. Retrieved January 24, 2021 from https://ethanzuckerman.com/2021/01/18/fixing-disinformation-wont-save-us/.

[196] S. Hegelich. Facebook needs to share more with researchers. *Nature*, 579(7800):473–473, 2020.

[197] J. Horwitz. Facebook Seeks Shutdown of NYU Research Project Into Political Ad Targeting. *The Wall Street Journal*, 2020. Retrieved January 31, 2021 from https://www.wsj.com/articles/facebook-seeks-shutdown-of-nyu-research-project-into-political-ad-targeting-11603488533.

[198] R. Klar. Facebook to grant access to targeting information about political ads. *The Hill*, 2021. Retrieved January 31, 2021 from https://thehill.com/policy/technology/535694-facebook-to-grant-access-on-targeting-information-about-political-ads.

[199] M. Luca, K. Munger, J. Nagler, and J. A. Tucker. You Won't Believe Our Results! But They Might: Heterogeneity in Beliefs About the Accuracy of Online Media. *Journal of Experimental Political Science*, pages 1–11, 2021.

[200] W. J. Brady, K. McLoughlin, T. N. Doan, and M. Crockett. How social learning amplifies moral outrage expression in online social networks. 2021.

[201] R. Levy. Social media, news consumption, and polarization: evidence from a field experiment. *News Consumption, and Polarization: Evidence from a Field Experiment (July 16, 2020)*, 2020.

[202] Y. Theocharis and W. Lowe. Does Facebook increase political participation? Evidence from a field experiment. *Information, Communication & Society*, 19(10):1465–1486, 2016.

[203] L. Y. Lin, J. E. Sidani, A. Shensa, A. Radovic, E. Miller, J. B. Colditz, B. L. Hoffman, L. M. Giles, and B. A. Primack. Association between social media use and depression among US young adults. *Depression and anxiety*, 33(4):323–331, 2016.

[204] E. Whittaker and R. M. Kowalski. Cyberbullying via social media. *Journal of school violence*, 14(1):11–29, 2015.

[205] M. Anderson, J. Jiang, et al. Teens, Social Media & Technology 2018. *Pew Research Center*, 31(2018):1673–1689, 2018.

# A Misinformation in the 2017 German Election

This Appendix section provides excerpts from the publication **Social Media Report: The 2017 German Federal Elections**. The focus is on the insights regarding online manipulation on Twitter in the months leading up to the 2017 German Election.

### In

TUM.University Press DOI:10.14459/2018md1452635 (**Not peer reviewed)**

### Authors

Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos, Simon Hegelich

### Abstract

This report presents thorough research on social media platforms during the months before the 2017 German federal election. The focus is on Facebook and Twitter given their increasing role in online political communication. Over 350 million tweets and 37 thousand Facebook posts related to German politics were collected and analyzed. This work takes an overlook at the online interaction between users and political parties. Moreover, it tries to identify disinformation and manipulation techniques.

### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

## ONLINE MANIPULATION

One of the most powerful advantages of monitoring social media is to be aware of possible online manipulation. Although it is difficult to control this from the beginning, the efforts against it should focus on avoiding its propagation. By monitoring different online channels, it is also possible to inform the general public of current means of manipulation.

In Germany, online manipulation has emerged as an important political issue, fueled by fears of election meddling. In June 2017, German lawmakers passed a law against hate speech and fake news, the Netzwerkdurchsetzungsgesetz. This law enforces social network platforms to take down defamatory and hate speech content in a 24-hour period to avoid a fine of up to 50m Euro. Given that German policy-makers are eager to find measures of controlling mass manipulation online, it is important to understand what happened during the election period. For this section, we focus on three main aspects of online manipulation: fake news, social bots and intervention from foreign actors.

### Fake News

One of the most controversial topics regarding online manipulation is the spread of fake news and its effect on political perception. Fake news is defined as "fabricated information that mimics news media content in form but not in organizational process or intent" [1]. Its goal can be to spread misleading information or simply false information to deceive people. There exists little research to date that sheds true light on the effect of fake news on political events. Additionally, there is no tool that can automatically classify between real and fake news. Human curation is needed to understand which news is false or true, and even then it can be difficult to make a correct decision without the help of professional fact-checkers. For this reason, a complete review of all news shared in the millions of collected tweets is a difficult task. We decided to adopt two approaches to overcome this difficulty and have a general overview of the news that was shared during the election period.

The first approach involved looking at a relatively large sample of our data to find whether the most commonly-shared news included fake news. For this purpose, we selected the tweets that had mentioned at least one German political party and extracted the URLs that were shared in the tweets (the URL extraction procedure and the analysis of online media are the main topics of Chapter 5). The sample comprises more than 5 million tweets . We manually looked at the top 100 news shared in these tweets. Additionally, we checked the number of shares on Facebook for every media link found in the sample of tweets. We also manually analyzed the 100 most shared news on Facebook.

From the sample of tweets, none of the 100 most commonly-shared news was fake, but interestingly 51 of the top 100 were AfD-related. Most of them are polarizing stories that are meant to trigger political discussions. This reinforces the theory that populist parties thrive to be in the media limelight to gain as much publicity as possible [2]. We used the same URLs and checked their "popularity" on Facebook. The top results were different from those on Twitter. From the 100 most commonly-shared news, nine of them appear not to be completely accurate. As seen in Table 1, they are all related to the topic of refugees in Germany. Part of the message that is conveyed can be considered as misleading. They can be categorized as sensationalist stories with the purpose of triggering negative emotions to refugees. Most of them use certain facts incorrectly and show only one side of the story. Nevertheless, this is insufficient to categorize them as fake news. The study of fake news is not a trivial topic and requires extensive field knowledge.

A similar study regarding the 2016 US election [3] concluded that "at least in the 2016 election [the fake news 'framing'] seems to have played a relatively small role in the overall scheme of things." According to the study, disinformation and propaganda from dedicated partisan sites played a stronger role in the election. By looking at the most-commonly shared news on both Twitter and Facebook, we argue that a similar conclusion applies to the German

| Headline | Media | Shares |
|---|---|---|
| BKA vertuscht Straftaten von 600.000 Flüchtlingen! | Journalistenwatch | 107,740 |
| Syrer mit vier Frauen und 23 Kindern erhält monatlich ca. 30.030 € | Denken macht frei | 97,228 |
| Claudia Roth fordert mehr Flüchtlinge für Europa | Berlin Journal | 92,021 |
| Polizisten brechen Schweigen: Asylanten-Verbrechen werden auf Weisung von oben vertuscht | Unzensuriert.at | 90,373 |
| Merkel will in Afrika für Einwanderung nach Deutschland werben | Deutsche Wirtschafts Nachrichten | 86,028 |
| Merkel hofft auf 12 Millionen Einwanderer | Wochenblick | 68,215 |
| Ja, Asylbewerber bekommen wirklich kostenlosen Zahnersatz | Freie Zeiten | 66,482 |
| Flüchtlinge mit zwei Ehefrauen in Deutschland: Beide können Sozialhilfe bekommen | Epochtimes | 61,942 |
| Auf jeden neugeborenen Deutschen kommen fünf neue Migranten | Freie Zeiten | 57,659 |

Table 1. Misleading stories in the top 100 shared news on Facebook taken from the tweets' URLs.

election. Our data suggest that the larger players in the online communication were propaganda topics on the AfD and critical opinions on the refugee crisis.

A second approach involved focusing on specific cases of fake news that were reported in the German press. We monitored how the stories evolved over time, which tweets had the most relevance and which users were most active in the discussion. We picked two cases: the Käßmann incident and the Von Sahringen story.

The Käßmann incident involves a speech made by Margot Kässmann, a Lutheran theologian, during the evangelical church day in May 2017. During the speech, she compared the program of the AfD — which pursues increasing the quota of German children without an immigrant background — with the Nazi ideologies. Her quote was shortened to make it incorrectly appear as if she had said that all Germans with two German parents and four German grandparents were Nazis. This generated fury on social media among AfD supporters.

Figure 1 shows how the story evolved during the end of May. The black area under the main trend corresponds to the tweets that not only referenced the story but also included the word Nazi. Most of them express rejection to Käßmann's shortened quote. These selected tweets are more prominent at the beginning when the story breaks and then again during the last days of the story. In our dataset the five tweets with the most retweets were:

(1) RT @SteinbachErika: Wenn ich nicht bereits aus der Partei namens EKD ausgetreten wäre, nach den Aussagen von Frau Käßmann wäre das jetzt fällig
(2) RT @FraukePetry: #Käßmann blamiert sich mit ihren Aussagen auf ganzer Linie - Geschichtsrevisionismus in Reinform. #AfD
(3) RT @uebermedien: Eine infame und verleumderische Kampagne gegen Margot Käßmann. @AfD_Bund @kirchentag_de
(4) RT @SteinbachErika: Hervorragend! Imad Karim an Margot Käßmann: "Diesen 'Nazi', der in Ihrem Kopf geistert, habe ich NIE getroffen"
(5) RT @Joerg_Meuthen: Guten Morgen! Kann man allen Ernstes nüchternen Kopfes einen solchen Unsinn verzapfen? #Käßmann #AfD

Only one of the five tweets is debunking the story, while the rest are reacting against Käßmann's statement. In our data, 4,590 users were tweeting or retweeting the story. The users that commented the most were: (1) e_pitzky (2) Rumsucher (3) PeterPa34083139 (4) krippmarie (5) mrstone0856

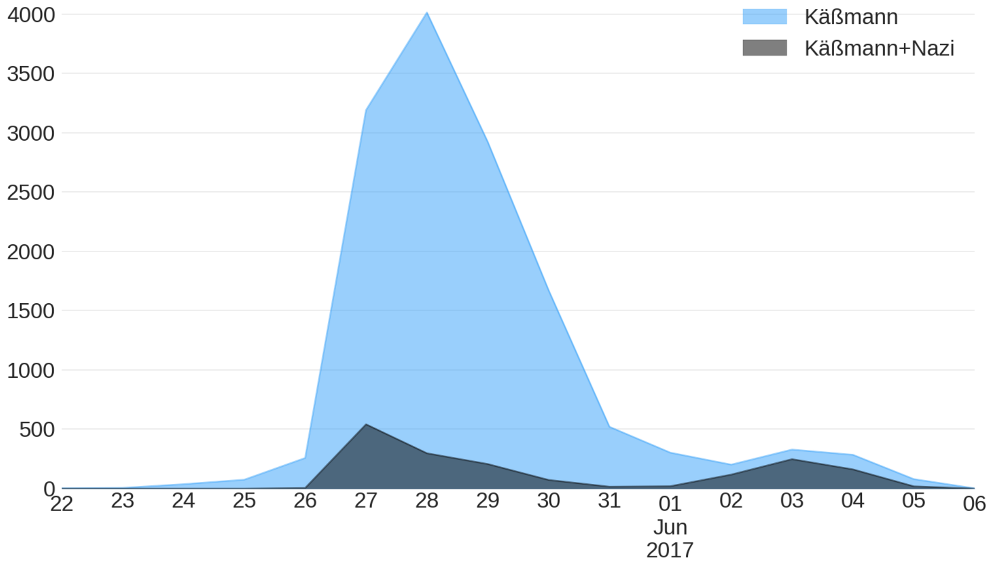


Fig. 1. Development of Käßmann's fake news story: Tweets that only included the word Käßmann and tweets that also included the word Nazi.

All of them sent over 50 tweets related to the case. By looking at their profile, they seem to be AfD supporters. We cannot completely verify whether these accounts are automated accounts or not. However, it is not common that a normal user would send over 50 tweets about one topic. These accounts definitely had the purpose of making the information go viral.

The second story that we followed is about a vote-rigging claim days before the German election. It started with a tweet from the account *@von_Sahringen* that stated: "I was called to be an election helper. On Sunday the votes for the AfD will be made invalid". Hours later, the account for Germany's official election bureau responded saying that this act was prohibited. By then, the hashtag #wahlbetrug ("election fraud") had started spreading on Twitter. The hourly trends of tweets mentioning Von Sahringen and Wahlbetrug can be seen in Figure 2. The largest peak for Von Sahringen occurred after the post from the election bureau. After this peak, the discussion on *Wahlbetrug* began to take off. The trend continued until the election, even though several media sites[1] had already reported that the account was a fake account with a modified picture of a Pakistani actress.

By looking at the most active users in the conversation, we find that bots were part of spreading the election fraud narrative. From the top five accounts mentioning @Von_Sahringen, two were closed by Twitter: A_Flicklgruber and ExilFury. Two other accounts changed their identity: EmperorFawful, a male user, became Julia Bathory, a female user; and Ouando_MdB, a supposedly AfD representative in the German parliament became KasimirQY, a shower curtain

[1]https://medium.com/dfrlab/electionwatch-final-hours-fake-news-hype-in-germany-cc9b8157cfb8

salesman in Estonia. This is typical behavior of bots that change their identity (description, image, location) to serve new purposes. On the other hand, the top five users that tweeted about the election fraud were all closed by Twitter. We assume that since the case was extensively reported on news media, Twitter acted against the accounts that were spreading the story. The next subsection continues the discussion on bots and their impact on election-related events.

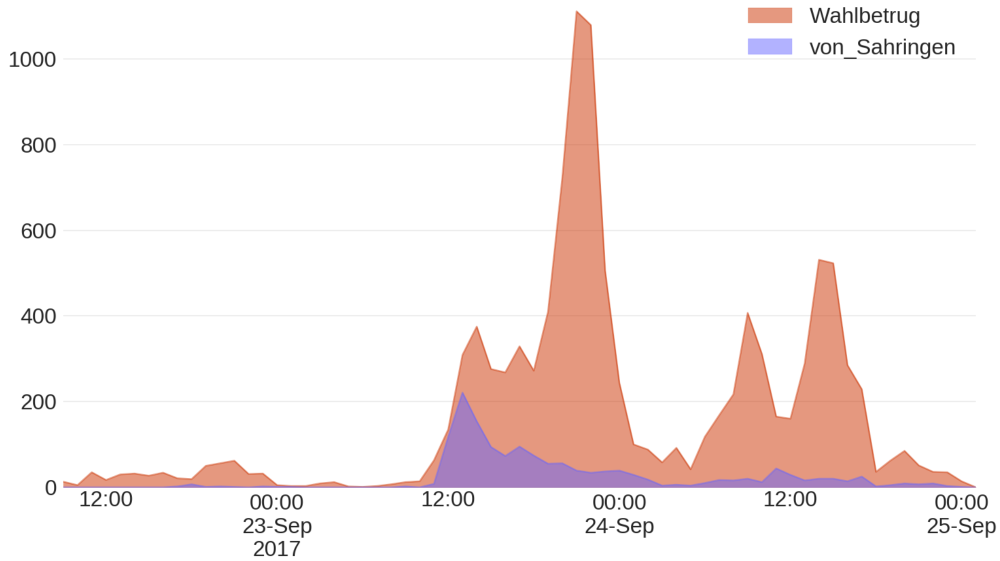


Fig. 2. Development of the Wahlbetrug story: Tweets that mentioned the fake Twitter account and tweets that included the word Wahlbetrug.

### Social Bots

Social bots are automated accounts that try to emulate human behavior to influence normal users. They produce content and interact with humans on social media [4]. It has not been proven, whether social bots have been successful in the past in influencing people, but they are known to have manipulated trends in social networks. This could alter the perception of what is happening online and then influence the platform´s algorithms into showing the falsified content to more users.

During the US election, it was estimated that around 19% of the tweets connected to election topics were created by bots [5]. Different media sources were speculating how much influence bots would have in the German federal election[2]. Unfortunately, a percentage number is not completely reliable, since it is unclear which measure from automatization makes an account a bot. Indeed, the term social bot is loosely defined. Additionally, the existing detecting methods only identify accounts with certain predefined characteristics and bots are constantly evolving to avoid getting "caught" (account closed by the platform). In order to have a better understanding of social bots in Germany, our task was to use a similar study as conducted in [5], while also using other detection methods for comparison.

We used three approaches to identify bots in our dataset: two heuristic-based methods and one machine learning method. Heuristic methods are based on rules deciding whether a tweet comes from an automated account or not.

[2]http://www.sciencemag.org/news/2017/09/social-media-bots-tried-influence-us-election-germany-may-be-next

The rules depend on specific properties from a tweet. For example, a tweet that originates from a verified account is directly treated as non-automated. We selected four different properties that have been used in previous works to identify tweets with a bot-like behavior: tweets coming from a suspicious source, tweets that are text duplicates and not retweets, tweets from users with excessive amount of tweets per day and tweets from users with a ratio of friends and followers close to 1. The two heuristic approaches differ only in the number of rules that have to be true to categorize a tweet. A simple approach includes all tweets that fall into any of the four categories. A second, stricter approach involves categorizing tweets as bots when they comply with at least two of the heuristics.

On the other hand, machine learning methods automatically find patterns with the help of hand-coded data, called training data. These methods are useful to make predictions on data that are similar to the training data. For the machine learning approach, we used the open source tool Botometer, which was developed at the Indiana University [6] and has been used extensively in the literature, including the aforementioned US study. This tool categorizes Twitter accounts depending on 1,150 features that are used in a machine learning model. Botometer gives a score between 0 and 1 per user and all of the users with a score larger than 0.5 are considered to be bots. Botometer specializes in English tweets and thus it is not as accurate for the German language. For our analysis, accounts that were explicitly closed by Twitter cannot be analyzed by Botomer and they are directly considered as bots. Concrete explanations regarding the three selected methods of bot detection are included in the Appendix.

For the analysis on bots, we first focused on the tweets that had mentioned the AfD during the complete period of observation. As already mentioned before, during the month of September the discussion on AfD increased in comparison to the main candidates. For this reason, we decided to use this sample of data and try to understand whether this effect was caused by bots. We selected only the German tweets since the AfD effect is only present in the German tweets. The dataset comprises 2,747,193 tweets from 195,779 users.

The results of the three methods are shown in Figure 3[3]. Each method gives a different percentage average of bot accounts. This originates from the fact that social bots are not a well-defined category and each method detects different kinds of suspicious accounts. The simple heuristic approach gives an average of 14.98 percent, the strict heuristic 1.5 percent and with the Botometer tool — including closed accounts — 9.9 percent. The results of the simple heuristic and Botometer are similar to previous research on bots, where around 10-15 percent of the conversation on Twitter is attributed to automated accounts.

More insights can be obtained from observing the percentage of bots over time. This percentage is based on the total number of tweets, which means that the 21% of the simple heuristic on the first of April corresponds to 1,037 tweets and the 13% on the first of September to 3,235 tweets. For the three methods, the percentage remains constant throughout the first months and then declines around June. During the month of September, it further declines in both heuristic methods. For the Botometer approach, the percentage remains under 10 percent for most of the time during the last two months. The five peaks above 10 percent are attributed to now-closed accounts. Even though there were overall more suspicious accounts tweeting about the AfD during the month of September, their activity was insufficient to strongly influence the overall conversation. By contrast, during the month of September, the normal user accounts generated more content that included the word AfD than bots. Accordingly, we conclude that automated accounts were not the main cause of the increase in conversation on the AfD.

We decided to analyze other datasets and compare with the AfD results to have a better measure of bot effectiveness. There were two specific cases in the election period where the media explicitly reported on bot attacks. The first

[3]We had previously shown in [7] the Botometer results without including the closed accounts.

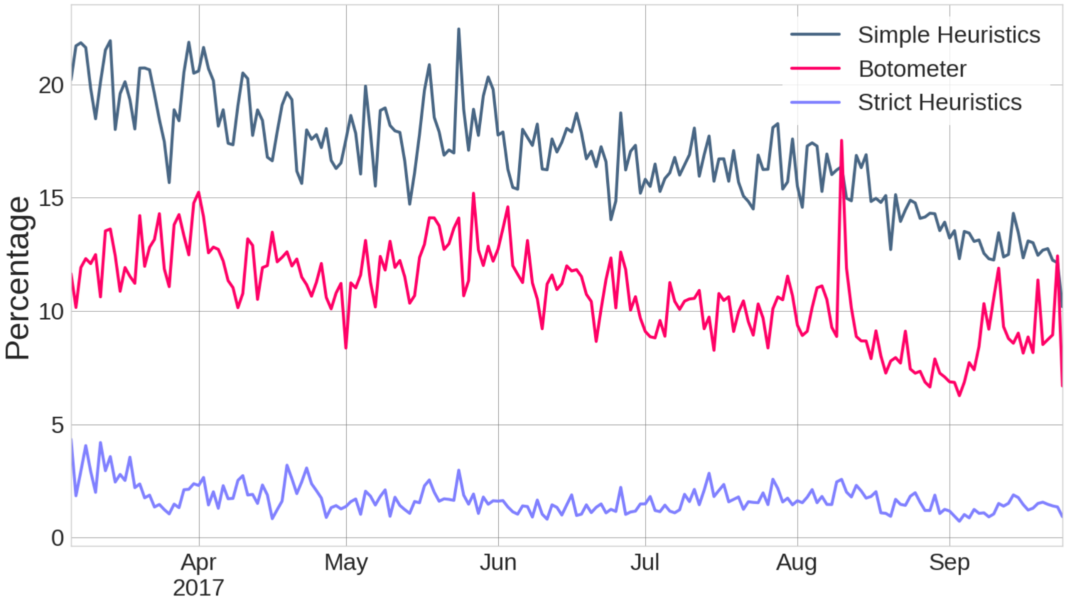


Fig. 3. Percentage of tweets from bots in the collection period according to three methodologies. The evaluated sample comprises all the German tweets that included the word AfD.

one[4] was during the TV debate on September 3 between Merkel and Martin Schulz, the candidates from CDU and the SPD, respectively. During the debate, the hashtag #verräterduell ("traitors' duel") became prominent in the Twitter conversation, although it did not make it to Twitter's trends. Nevertheless, it was considered an attack from far-right groups, as shown by a journalistic investigation[5]. The second case was part of the aforementioned fake news story from the election fraud[6]. After the Von Sahringen story broke, there were many accounts using the hashtag #wahlbetrug ("election fraud"). According to the investigation, a Russian network of bots was responsible for spreading tweets with AfD propaganda that included the election fraud hashtag.

We collected 3,615 tweets that included the hashtag #verräterduell and 8,617 tweets with the hashtag #wahlbetrug from our data. The number of tweets is too small to use the heuristic methods, so we analyze the data only with the Botometer approach, which focuses on users and not on tweets. The results for the different events are presented in Figure 4. The case with the most bots is the *Wahlbetrug* with 58 percent of tweets coming from automated accounts. Among this 58 percent, 51 percent of the tweets come from closed accounts by Twitter. This case was very present in German media and happened in the days leading up to the election, which probably motivated Twitter to be effective in closing suspicious accounts. By comparison, 28 percent of tweets belong to bot accounts in the *Verräterduell* case and 11 percent come from already-closed accounts. For the AfD, among the 2,747,193 tweets, 4.5 percent of the 9.9 percent belong to closed accounts. This is much lower than the events known for having bot attacks.

[4]https://medium.com/dfrlab/botspot-memes-target-der-spiegel-merkel-678a2fc52b05
[5]https://www.buzzfeed.com/karstenschmehl/willkommen-in-der-welt-von-discord-teil1?utm_term=.laYM2EkLKz&bftwdenews#.fiXJ84ayXY
[6]https://www.focus.de/politik/deutschland/bundestagswahl_2017/bundestagswahl-analyse-hunderte-fake-twitter-profile-verbreiten-beitraege-von-afd-unterstuetzern_id_7631486.html

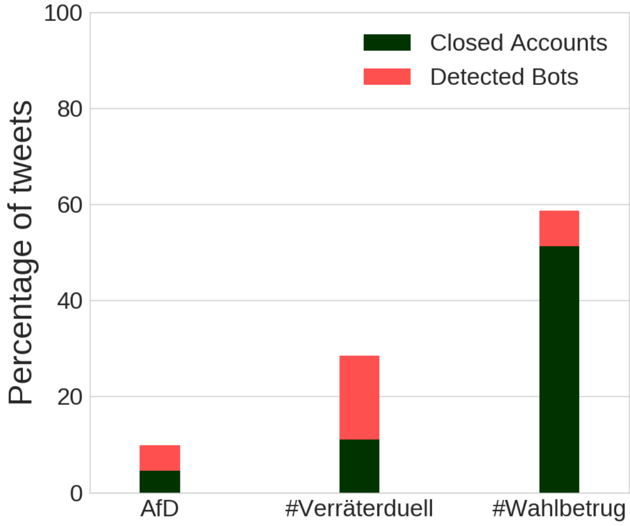


Fig. 4. Percentage of tweets from bots for the three evaluated cases. Each case is divided in detected bot accounts by Botometer and closed accounts by Twitter.

Our analyses confirm that social bots were indeed depleted during the election period as an online manipulation tool. These attacks were able to gain attention from the media, although a real effect on voter intention cannot be quantified from the results. Apart from the coordinated bot attacks, we do not see evidence of social bots having as much effect on the overall political conversation as expected by media experts. However, the phenomenon that we observe is that the right-wing opposition party AfD is dominant on Twitter, and most of the bots we found were working in their favor. Our work is consistent with [8], where 1 million tweets of German political content were collected in a ten-day period.

### Foreign Intervention

The last form of online manipulation that we researched was the intervention of foreign actors in the German elections. This is a broad topic that has to be approached carefully since the plausibility of tracing back the origins of online manipulation is extremely limited. For this subsection, we focus on actors that have been speculated to have attempted to influence the elections through manipulation.

The first actor under investigation is Russia, primarily, since it was in the media limelight during the US elections. The US Congress continuously investigated the allegation of Russian efforts to meddle in their elections and evaluated whether the usage of Russian propaganda had an effect on the outcome. In Germany, it was also speculated whether Russia could have had plans to target Germany[7].

The first way to evaluate intervention is to focus on previously-disclosed efforts. In October 2017, Twitter released a list of 2,752 Twitter accounts that the company identified as being connected to the Russian Internet Research Agency[8]. The identified Russian trolls were analyzed in [9] and [10]. The latter study collected tweets related to the US election for two months and subsequently found 221 Russian trolls in the data. However, until now it has not been shown

[7]http://www.spiegel.de/netzwelt/web/bundestagswahl-2017-debatte-um-moegliche-manipulationen-durch-russland-a-1165520.html
[8]https://www.recode.net/2017/11/2/16598312/russia-twitter-trump-twitter-deactivated-handle-list

whether these accounts were also active in topics outside the US elections. We proceeded to explore whether the already-deactivated accounts appeared in our Twitter dataset. Surprisingly, we found 23,595 tweets from 458 of the accounts tied to the IRA. As explained in the collection methodology, the tweets need to have a connection to a topic in German politics. From the 458 accounts, 98 tweeted in German and generated 13,932 German tweets from the total of 23,595 tweets. This means that the identified Russian trolls were also trying to spread information to the German users.

The activity over time of the accounts separated in German and non-German tweets (Figure 5) shows that before May most of the accounts were tweeting in languages other than German, which changes at the beginning of May. In May and June most of the tweets are German and in July and August there is a similar quantity of German and non-German tweets. In September, the month of the election, almost all the tweets are German, which corresponds with the idea of the troll accounts trying to influence Twitter during the election period.

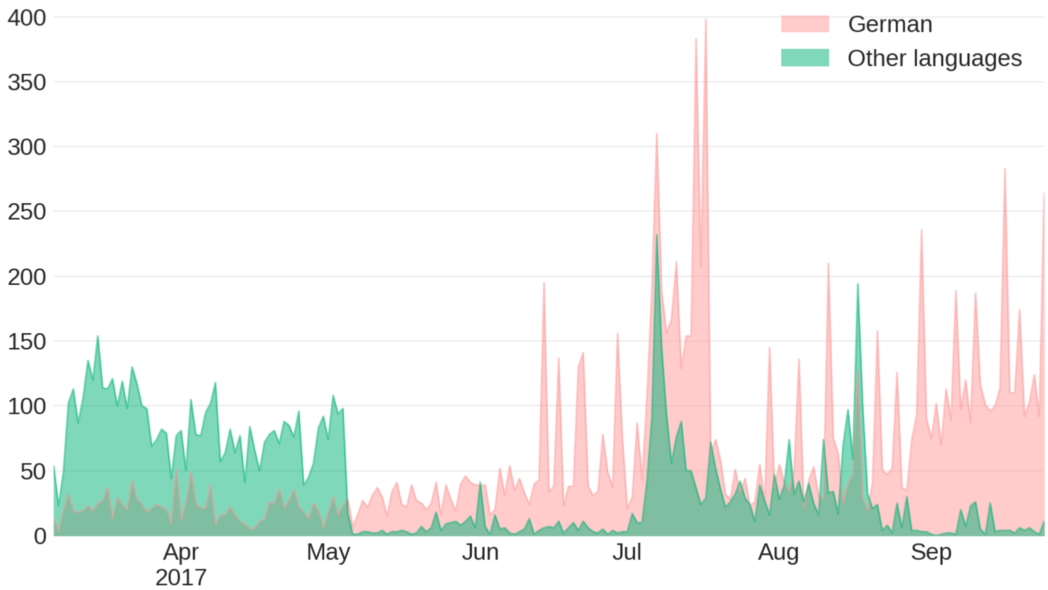


Fig. 5. Number of tweets from accounts belonging to the list of detected Russian trolls. The tweets are separated between German and non-German tweets.

The contents of the tweets offer more insights into how these accounts operated. Interestingly, 11,730 of the German tweets included an URL, which means that the majority of the tweets (~84%) had the intended purpose of amplifying information from other sources. The five media sources that were shared the most by the accounts were the Dresdner Neueste Nachrichten, Bild, Die Welt, Tagesspiegel and the Berliner Zeitung. We find that most of the tweets include links to what can be referred to as mainstream media. From more than 1,000 shared media links, only ninteen refer to sputnik.com, nine to Russia Today's German portal and two to sputnik.de. The supposedly Russian trolls were only amplifying messages from normal media sites, which differs from their activity in the US discourse where sites like Breitbart.com and thegatewaypundit.com had considerable diffusion [10].

The influence of the troll accounts can be traced back by evaluating retweets. From the total tweets, language independent, 11,571 were original tweets and they originated 92,043 retweets in our database from 46,198 users. Only 37 bot accounts retweeted contents between each other. As previously mentioned, the collected data is just a sample from the complete Twitter traffic and it is biased towards political topics. Nevertheless, a replication factor of 8 in our sample suggests that the troll accounts were not isolated in Twitter and their activity was noticed by other Twitter accounts. The top retweeted accounts were media accounts, i.e die Welt, Tagesschau, Stuttgarter Nachrichten, Tagesspiegel and Spiegel.


Fig. 6. Word cloud of the tweets from the detected Russian trolls.

We further analyzed the 13,932 German tweets. In the texts, the top five hashtags were Merkel, AfD, CDU, SPD and G20. The first four are included in the list of hashtags that we are following to collect tweets and the fifth hashtag corresponds to the high activity during the days of the 2017 G20 conference (Figure 5). Regarding the topics discussed in the tweets, we cleaned the data by eliminating stop words, smileys and URLs and created a word cloud (Figure 6) of the most discussed topics. The word Merkel is the most common one, followed by AfD. Also interesting are the words, flüchtlinge (refugees), polizei (police) and the hashtag merkelserfolge (Merkel's successes), which was mainly used in a negative, sarcastic tone on Twitter.

## REFERENCES


[1] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359:1094–1096, 2018.

[2] Gianpietro Mazzoleni, Julianne Stewart, and Bruce Horsfield. *The media and neo-populism: A contemporary comparative analysis*. Greenwood Publishing Group, 2003.

[3] Rob Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. Partisanship, propaganda, and disinformation: Online media and the 2016 U.S. presidential election, 2017.

[4] Emilio Ferrara, Onur Varol, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.

[5] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 U.S. presidential election online discussion. *First Monday*, 21, 2016.

[6] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint*, arXiv:1703.03107, 2017.

[7] Simon Hegelich. *Die digitale Revolution der Meinungsbildung Social Media im Wahlkampf.* Hans Seidel Stiftung, 2018.

[8] Lisa-Maria Neudert, Bence Kollanyi, and Philip N. Howard. Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter?, 2017.

[9] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. *arXiv preprint*, arXiv:1801.09288, 2018.

[10] Adam Badawy, Emilio Ferrara, and Kristina Lerman. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. *arXiv preprint*, arXiv:1802.04291, 2018.

# B Coordinated and Suspended Accounts on Twitter in the run-ups to General Elections


#### Authors

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich



#### Unpublished



#### Abstract

This study investigates coordinated behavior and suspended accounts on Twitter during the fourteen days leading up to the general elections in the United Kingdom, Mexico, Germany, and Greece. We collected over 23 million political tweets using Twitter's Streaming API. We analyzed users' temporal activities and the semantic content of tweets to understand and quantify coordinated behavior. Additionally, we identified accounts suspended by Twitter and compare their impact between countries. Overall, we find low levels of coordinated behavior and few suspended accounts in comparison to the number of active users. However, between 20% and 40% of the accounts have more than 1,000 followers. We find the highest number of detected accounts in the Mexico dataset and the lowest percentage of coordinated activity in the UK dataset. Furthermore, we manually labeled the partisanship from a 10% sample of both coordinated and suspended accounts to understand their political intent. We find that most accounts support or oppose the main political parties, with the exception of Germany, where the focus is only on one party the *Alternative für Deutschland* (AfD). Although each country has a different political spectrum, we find similar patterns of partisan accounts.


#### Contribution of thesis author

Theoretical design, model design and analysis, manuscript writing, revision and editing

# Coordinated and Suspended Accounts on Twitter in the run-ups to General Elections

Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich

Technical University of Munich, Germany

**Abstract.** This study investigates coordinated behavior and suspended accounts on Twitter during the fourteen days leading up to the general elections in the United Kingdom, Mexico, Germany, and Greece. We collected over 23 million political tweets using Twitter's Streaming API. We analyzed users' temporal activities and the semantic content of tweets to understand and quantify coordinated behavior. Additionally, we identified accounts suspended by Twitter and compare their impact between countries. Overall, we find low levels of coordinated behavior and few suspended accounts in comparison to the number of active users. However, between 20% and 40% of the accounts have more than 1,000 followers. We find the highest number of detected accounts in the Mexico dataset and the lowest percentage of coordinated activity in the UK dataset. Furthermore, we manually labeled the partisanship from a 10% sample of both coordinated and suspended accounts to understand their political intent. We find that most accounts support or oppose the main political parties, with the exception of Germany, where the focus is only on one party the *Alternative für Deutschland* (AfD). Although each country has a different political spectrum, we find similar patterns of partisan accounts.

**Keywords:** Twitter, Coordinated Activity, Suspended Accounts, Elections, Germany, Mexico, Greece, UK

## 1 Introduction

Twitter constitutes a major political communication hub in pre-election periods. Politicians use the platform to propagate their political positions, journalists produce and distribute political information, while the electorate uses the service to consume political news, get informed, and contribute to the political discourse [27, 10]. Communication patterns in the above cases are not always unbiased. Researchers have shown that the overall information circulation on Twitter is strongly shaped by specific user types and tactics. Hyperactive and hyper-partisan users, automated and fake accounts, together with tactics such as astroturfing, the distribution of low credibility or hyperpartisan news, spamming and inauthentic or coordinated behavior strongly influence the circulated political content [12, 24, 5, 8].

In this study, we seek to understand coordinated behavior during pre-election periods. Coordinated behavior on Twitter denotes the collective organization of actions between a set of accounts, with the aim to inflate attention metrics and to promote a narrative or item [14]. This coordination takes place regardless of the automated or organic nature of the accounts, or their benevolent or malicious intentions [22]. Prior research studies show that political communication on Twitter was partly shaped by accounts' coordinated behavior during pre-election periods in various countries, such as during the Brexit [15], in the US 2016 elections [4], the 2018 Italian general elections [14] and the 2018 French elections [11]. However, no prior analysis compares coordinated behavior between countries, nor evaluates it as a standard constituent part of political communication. To bridge this gap, we analyze and detect coordinated behavior for four general elections: the 2017 German federal election, the 2018 Mexican general election, the 2019 United Kingdom general election, and the 2019 Greek legislative election. For the four elections, we additionally identify and analyze the "bad actors" according to Twitter. These are the accounts that the company suspended for violating the code of conduct. The Twitter rules stipulate that users are not allowed to post hateful, violent or abusive content, artificially amplify or suppress information, impersonate individuals or organizations, nor interfere with elections [2]. This paper aims to answer the following research questions:

**RQ1:** Overall, how prevalent are coordinated and suspended accounts in the run-ups to general elections?
**RQ2:** What is the political intent of coordinated and suspended accounts in the four selected countries?

Our analysis shows that the percentage of tweets from coordinated accounts correspond to up to 3% of the total tweets, and up to 9% for the suspended accounts. Overall, we identify a low number of coordinated and suspended accounts in comparison to the total number of users. However, a considerable amount of them has a high follower basis (more than 1,000 followers). The political intent of these accounts varies from country to country but there is one party in each country which is mostly favored by them. We believe this study sheds the light on comparing the political communication on Twitter between different countries in periods leading up to general elections and helps to understand the political motivation from possible unauthentic accounts. We hope this study motivate scholars to perform similar quantitative and qualitative research on other countries.

## 2 Related Work

Researchers have long studied social media users in pre-election periods in order to uncover behaviors that influence political discourse. This includes the investigation of social bots, trolls, hyperactive, hyperpartisan, activist, and malicious accounts that aim to shape circulated content [4, 9, 23, 5, 8]. Nevertheless, the explicit classification of users in one of the above categories has been proven troublesome because an account's behavior might hold features that belong to more than one of the above categories [15]. Furthermore, researchers often face limitations regarding available data and methodologies applied that obstruct accurate inferences [1].

An alternative way to dealing with such issues is not to analyze accounts' type, but to investigate accounts' practices. Regardless if a set of accounts are human, semi-automated, or fully automated, hyperactive, activist, or malicious, researchers can analyze them based on their interaction patterns. Understanding interaction patterns has been proven an efficient methodology for tracking if a set of accounts behavior is inauthentic, deviating from the average behavior of an arbitrary normal user, or coordinated, i.e. performing collectively organized actions towards a specific aim [14].

Researchers have applied different techniques to trace inauthentic and/or coordinated behavior. Zhang et al. [30] investigated tweeting intervals to assess if accounts' behavior deviates from the average human one. Pacheco et al. [22] studied the coordination of multiple actors to reveal suspicious behaviors, regardless of their automated/organic nature and malicious/benign intent using network theory and clustering analysis. Keller et al. [16] traced specific coordination patterns between Twitter accounts by the clustering of activities and the processing of textual features. Similarly, Pacheco et al. [21] detected coordinated practices based on automatic retweets and content duplication (URLs/text). Francois et al. [13] generalized the notion of coordinated behavior, by stating that it can appear in three ways: network structure, temporal activities, and semantic content, while Monsted et al. [19] deployed themselves social bots and coordinated their activities in order to measure social influence on Twitter.

Although there is an extensive literature of abuse on social media, few studies have explicitly analyzed the characteristics of suspended accounts. Thomas et al. [28] examined suspended users to identify and characterize spam accounts. While Alorainy et al. [3] used tweets from suspended accounts to train a classifier to detect hateful speech, Volkova et al. [29] developed machine learning techniques to predict if an account would be suspended. Chowdhury et al. [6] collected the 1% percent sample from the overall Twitter conversation for eight months in 2018 and identified 2.4 million suspended accounts. The researchers found that politics was a major conversational topic among these accounts. The only work that investigated suspended users during an election period is the one by Le et al. [17]. They collected tweets mentioning Hillary Clinton or Donald Trump during the 2016 US election and identified clusters of suspended accounts according to their retweet and mention network. However, until now, there is no prior work that quantifies partisanship by manually coding suspended accounts.

## 3 Data Collection and Methodology

**Data Collection** For some social media platforms, researchers can access public user data with the help of application programming interfaces (APIs). We collected the data with the help of Twitter's Streaming API[1]. The API allows

[1] https://developer.twitter.com/en/docs/tweets/filter-realtime/overview

to querying tweets according to specific hashtags or user accounts. For each country, we curated a list of hashtags related to the elections. We included the word election plus the election year (e.g. *Elections2019*, *Elecciones2018*), the political candidates' names with and without the year, the name of political parties, and hashtags coined by the political campaigns. We made an effort to include a balanced set of hashtags that takes into consideration the different political orientations. We further collected the tweets from the accounts of political candidates, political parties, and other relevant political actors. We included the interactions with these accounts, namely mentions and retweets. We also strove towards having a balanced list of users that covered the different political ideologies. The collection period spanned the fourteen days leading up to the general election including election day. In the case of Germany, we additionally collected tweets for fourteen days starting two weeks after election day. With this data, we aim to compare the coordinated Twitter activity between an election and a non-election period. From the overall collected data, we filtered the tweets from other languages other than the official language of the country. The complete list of filtered tweet ids can be found in our Github repository[2] for future study replicability. The repository also includes the list of hashtags and user accounts that we curated for each country.

The collection process presents two major sources of bias. First, Twitter's API only provides a sample of the complete Twitter conversations and the sample has been proven to be biased [20]. Second, the political landscape of each country is different and the number of hashtags and user accounts differ between countries. Therefore, any inference that comes from comparing the countries should be taken with caution.

**Coordinated Behavior** Users interact on Twitter by posting original tweets and sharing tweets from others, the so-called retweets. The first step to find coordinated activity was to define a set of features that can be useful to find users with similar activity. We focused on three features: retweets, URLs, and hashtags. We only considered URLs and hashtags in original tweets. The second step consisted in making a list of the users' retweets, URLs and hashtags and compare them. For this, we filtered out users from which we did not have sufficient information—we only considered users with at least 10 retweets, users with at least 10 unique URLs, and users with at least 10 unique hashtags. In this way, we obtained three filtered lists of users. For each list, we then compared every user with the rest of the users using the Jaccard similarity between sets of features:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

which is defined as the size of the intersection of set A and B divided by the size of the union of the two sets. We selected the pairs of users with a Jaccard similarity higher than 0.33 for further analysis. We selected this threshold given

[2] Will be added after the peer-review process to ensure anonymity

that for any two users with the same (even) number of elements, the Jaccard similarity is 0.33 if they have the half of elements in common. Given $|A| = N$, $|B| = N$ and $|A \cap B| = \frac{N}{2}$, we have:

$$J(A, B) = \frac{\frac{N}{2}}{N + N - \frac{N}{2}} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1}{3}$$

To get a broader perspective on the data, we calculated Jaccard similarities for all pairs of users for the four countries and for the three features. We found that less than 0.05% of the pair similarities between users is above 0.33. This represents a high threshold that only takes into account users with very similar tweeting patterns. Although this highly reduces false positives, it is possible that we miss accounts that have a lower percentage of similar tweets but are nevertheless part of a coordinated campaign (false negatives).

The next step consisted in comparing the timing of activities between similar users. We calculated the time differences between the timestamps of common features. In the case of users with similar retweet activity, the difference was calculated directly from the timestamp of a retweet. URLs and hashtags may have non-unique timestamps as it is possible for a user to include them in more than one tweet. Therefore, we calculated the difference between two users using all possible timestamps where the accounts used the common feature in their tweets. We selected only the minimum time difference for a given URL/hashtag per pair of users.

For each feature, we thus had a new list of pairs of similar users, and each pair had itself a list of time differences from the similar elements. For each pair of users, we selected the median of the absolute time differences as decision criteria and defined coordinated users as those pairs which had a median smaller than ten minutes. We selected this decision threshold after evaluating the time differences between all retweets in the datasets. This means that we took the list of retweets of a given tweet and calculated all pair differences between the retweets. We repeated this procedure for all original tweets that had more than one retweet in the datasets. The cumulative distribution of absolute time differences is shown in Figure 1. We observe that the probability of the differences being less than ten minutes is 0.047. If we treat each retweet as an independent event (which in real life is not the case), we obtain that the probability of two users having three identical retweets[3] in less than ten minutes is equal to $0.046^3 \approx 0.0001$. Only 0.01% of probability.

Finally, we grouped the pairs that shared one similar account. This allowed us to find clusters of accounts with possible coordinated behavior. An additional post-processing step was needed in the case of URLs and hashtags as a normal non-coordinated account may have posted original content, which then was copied by a group of coordinated accounts. To filter out the original content

[3] This corresponds to the extreme case of having two accounts with ten retweets each, five similar retweets and three (the median) of them being posted within a ten minute difference

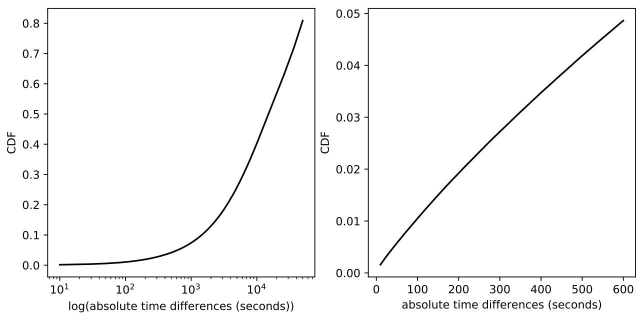


**Fig. 1.** Cumulative distribution of absolute time differences (seconds) between retweets of the same tweet in log scale (Left), and only considering up to 10 minutes (Right)

accounts, we looked for each cluster of coordinated accounts and found if one of them posted the tweets before the rest of the accounts in the group and we excluded them from the coordinated category.

**Suspended Accounts** Apart from detecting coordinated accounts, we searched for the accounts that Twitter suspended. We only focused on the accounts with more than 20 posts in our datasets, including tweets and retweets. We made this decision as there is a higher probability that the active political accounts were suspended in connection to the election. We were able to find which of the accounts had been suspended as Twitter explicitly states if an account was suspended for violating the platform rules. Finally, we calculated the percentage of suspended accounts that we had previously classified as coordinated.

**Partisanship** We manually assign a partisanship to suspended and coordinated accounts according to their political preferences. For each country apart from Greece, we labeled a random sample of 10% of suspended accounts and 10% of coordinated accounts. For Greece, we considered all the detected accounts as they were fewer than for the other countries. The procedure consisted of labeling accounts as either supporters or opponents to a political party. Supporter accounts promoted only one political party in their tweets. Opponent accounts attacked only one political party and did not show any support for another party. We did not label accounts that supported more than two political parties or accounts that attacked more than one party and showed no support for another one. We found few instances of such accounts and omitting them should not make an impact on the results. Accounts without a concrete political stance were not considered for the partisanship analysis. We also refrained from labeling accounts from the second German dataset corresponding to the non-election period.

## 4 Results and Discussion

The collected datasets represent the four countries during the election periods. Table 1 shows the descriptive statistics of the datasets, including the additional dataset for Germany outside the election period. The UK dataset is the largest followed by Mexico, then both datasets from Germany and at last-place Greece. The number of unique users in the data follows the same pattern. This also corresponds to the order of countries when considering the number of Twitter users per country [7]. A majority of the Twitter activity consists of retweets, more so for UK and Mexico (>70%) as for Germany (54%). Apart from Germany, more than the majority of users do not have original political content and were active only through retweeting. In both datasets from Germany the ratio is almost one to one between tweets and retweets. This may indicate an artifact of the data collection procedure and does not show that German users were more active posting original tweets than retweeting. However, we do observe that users in Germany are more politically active during the election as we collected 2 million tweets more from German politics in the election period.

**Table 1.** Descriptive statistics of the five country datasets. Germany* refers to the dataset collected during the non-election period. The percentage of retweets users refer to those users that have no original tweets in the collected data.

| | **UK** | **Mexico** | **Germany** | **Germany*** | **Greece** |
|---|---|---|---|---|---|
| Total tweets | 8,521,963 | 8,150,663 | 4,681,767 | 2,602,488 | 342,918 |
| Retweets | 75.35% | 74.27% | 54.02% | 51.31% | 66.73% |
| Unique users | 1,639,983 | 1,293,368 | 360,716 | 231,979 | 59,476 |
| Retweets users | 57.98% | 54.52% | 40.49% | 38.41% | 55.79% |

For each country, we detected possible groups of coordinated accounts according to three features: retweets, URLs, and hashtags. Table 2 shows the percentage of tweets that were posted by these accounts for each feature. The percentage of coordinated retweets considers only retweets, whereas the percentages of coordinated URLs and hashtags refer only to the original tweets. The UK has the lowest percentages of coordination for the three features. Mexico and Greece have the highest percentages of coordinated retweets and similar percentages for URLs and hashtags. Germany has the highest percentage with coordinated hashtags, but lower levels in the other two features. This higher percentage can be mostly attributed to a highly active group of thirteen coordinated accounts supporting the political association "Freie Wähler". The accounts' user names have the same structure "FWni" plus the name of a German town, and the majority of their tweets were promoting the hashtags *#FREIEWÄHLER* and *#anstaendigealternative*, which means decent alternative. The higher percentage for coordinated hashtags is not observed in the case of the German non-election dataset. When taking into consideration the three features together, coordination behavior was higher for Greece with 2.91% of the collected tweets, followed by Germany, Mexico, and the UK. However, the higher total activity in the Greece dataset represents only the posts of 125 accounts. For Mexico, we found 3,638 coordinated accounts, the highest number from the five datasets, followed by Germany with 706 accounts. In comparison to the non-election period for Germany, we observe there were more coordinated accounts during the election period and these accounts generated a higher volume of coordinated posts.

**Table 2.** Percentage of coordinated activity found for each country. The percentage activity from all the coordinated accounts found is reflected in total coordinated.

| | **UK** | **Mexico** | **Germany** | **Germany*** | **Greece** |
|---|---|---|---|---|---|
| Coord. retweets | 0.14% | 1.82% | 0.75% | 0.75% | 2.36% |
| Coord. URLs | 0.05% | 0.49% | 0.67% | 0.85% | 0.55% |
| Coord. hashtags | 0.11% | 1.08% | 3.21% | 0.29% | 1.24% |
| Total coordinated | 0.34% | 2.16% | 2.77% | 1.04% | 2.91% |
| Coord. accounts | 478 | 3,638 | 706 | 351 | 125 |

Table 3 shows the number of suspended accounts per country and the percentage of tweets posted by them. Given that we only report on accounts with more than 20 posts in our datasets, these percentages would be higher when considering less active users. As with coordinated accounts, Mexico leads with the number of suspended accounts (3,602), followed by the UK (1,759). However, the UK accounts were responsible for only 1.76% of the total amount of posts, which is lower than in the other countries. Similar to the previous result on coordinated accounts, Greece has the lowest amount of suspended accounts (88). Interestingly the highest percentage of activity from suspended accounts corresponds to Germany in non-election periods with 9.37% of the total activity. This suggests that during non-election times the "bad actors" (according to Twitter) have a higher percentage contribution to the total political conversation than during elections. This result follows the lines of previous research [25] that found that outside of election period in Germany bot activity is lower but represents a higher percentage of the overall conversation. Table 3 also includes the percentage of suspended accounts that we detected for coordinated behavior. For Mexico, almost 16% of the suspended accounts are present in the list of coordinated accounts. For the rest of the countries, this percentage is less than 8% and for the UK there is no overlap between the two sets of accounts.

It is important to look at the number of followers from the suspended and coordinated accounts to quantify their possible impact. Some accounts may be prolific in generating tweets but if there are no users following them, their activity would go unnoticed and their tweets could only have an impact on Twitter

**Table 3.** Percentage of tweets posted by suspended accounts, number of suspended accounts and percentage of suspended accounts that we detected as coordinated. We only looked for suspended accounts with more than 20 posts in the collected data.

| | UK | Mexico | Germany | Germany* | Greece |
|---|---|---|---|---|---|
| Suspended tweets | 1.76% | 4.40% | 6.72% | 9.37% | 2.34% |
| Suspended accounts | 1,759 | 3,602 | 1,436 | 1,295 | 88 |
| Suspended accounts (%) in coordinated | 0% | 15.92% | 6.45% | 5.29% | 7.31% |

trends. Figure 2 shows the reachability of suspended accounts (left) and coordinated accounts (right). Each plot represents the cumulative distribution of accounts with a given number of followers. We observe that depending on the country, between 20 and 40 percent of the suspended accounts had less than 10 followers, and between 20 and 30 percent of them had more than 1,000 followers. The same pattern appears for the coordinated accounts, apart from Mexico with more than 50% of the coordinated accounts with less than 10 followers, and Greece and the UK with almost 40% of the accounts having more than 1,000 followers. In both plots, Mexico's cumulative distribution is situated to the left of the rest, which means that on average the reach of the suspended and coordinated accounts was lower in comparison to the other countries. The contrary holds for the UK and Greece, which have the curves on the right. For Greece, this may be an artifact of having fewer accounts in comparison to the other countries.

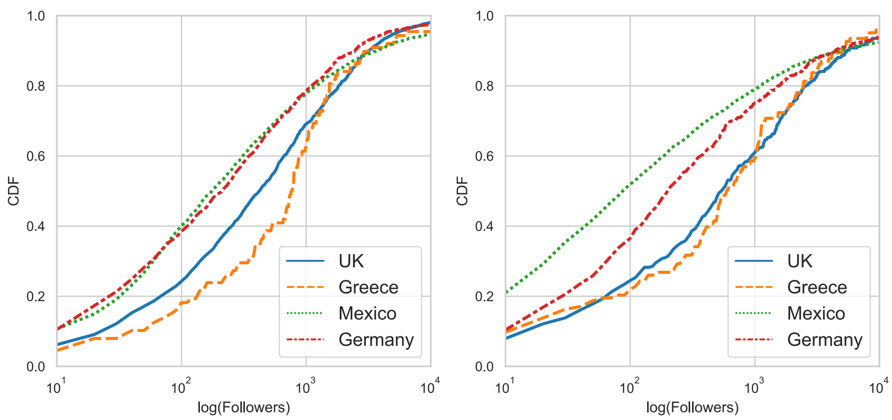


**Fig. 2.** Cumulative distributions of number of followers (in log scale) from suspended accounts (Left) and coordinated accounts (Right)

The last analysis consisted of manually classifying a sample of the suspended and coordinated accounts for each country. We assigned the accounts as supporters or opponents of political parties as explained above. Figure 3 presents the percentage of accounts supporting (in blue) or opposing (in red) a political party divided by country and by account type. We omit the political parties that did not appear in the labeling procedure. The percentages do not sum up to 100 as there were accounts that were neither supporting nor opposing a political party. In the case of the UK, we observe that the suspended accounts were mostly supporting or opposing the two major political parties, Labour and Conservative. The percentages are higher for Labour and less than five percent of the accounts supported the Scottish National Party and the Liberal Democrats. In comparison, while there was no coordinated account in the sample that supported the

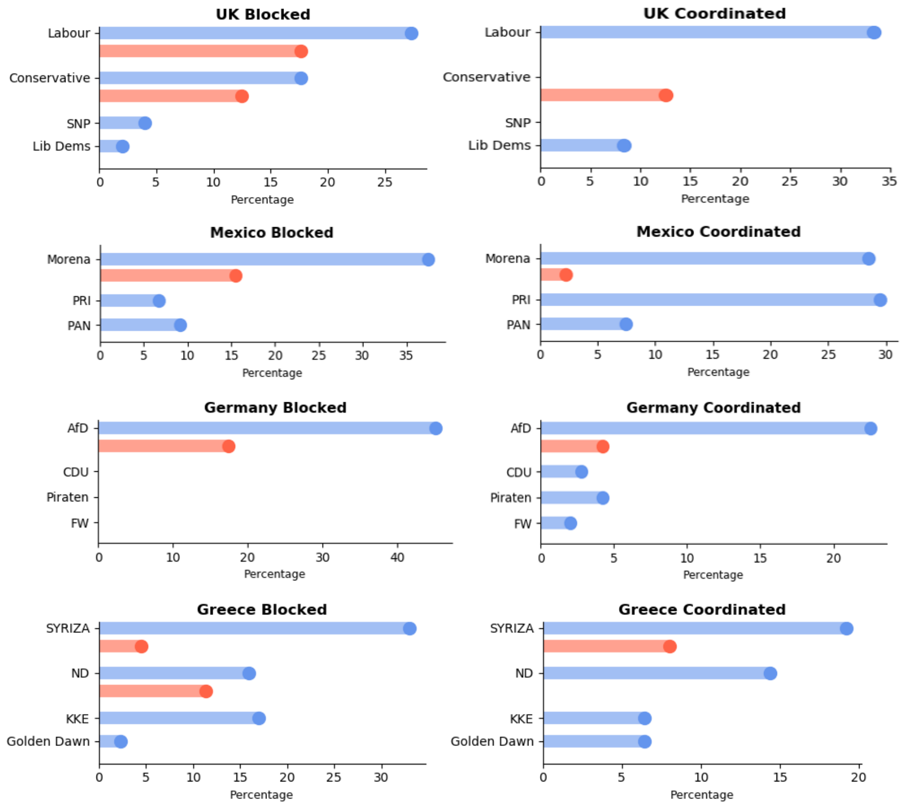


**Fig. 3.** Partisanship percentages of coordinated and suspended accounts per party and per country. For UK, Mexico, and Germany, we consider only a random 10% sample. For Greece, we take all accounts into consideration. The blue bars correspond to supporters of a political party and the red bars correspond to opponents of a party. The political parties that do not appear on the plots had no labeled account linked to them.

Conservative party, 33% of the accounts supported the Labour party. This imbalance could be correlated with the fact that the Labour Party dominated the conversation on Twitter during the weeks before the election [18].

In the 2018 Mexican general election, three political coalitions were formed. However, the accounts we labeled referenced directly Mexico's major political parties and not the coalitions. From the suspended accounts, the highest percentage of support went to the Morena party with 37.5%. The other two major parties, PAN and PRI had support from less than 10% of the accounts. Only for Morena, we find accounts attacking the party and at the same time not showing explicit support for another one. Interestingly, from the coordinated accounts sample, there are slightly more accounts supporting PRI than Morena, whereas for PAN the percentage is similar as in the case of the suspended accounts.

A different pattern appears from the sample of suspended accounts in Germany. All of them were either supporting or opposing the far-right party AfD. 45.1% of accounts supporting this party is the highest percentage of all the considered countries. From the coordinated accounts, the AfD is again the party with the highest support. However, the AfD is not alone, as there were other three parties with less than 5% percentage of accounts supporting them: The CDU, one of the governing parties, the Pirates, a small party not represented in Parliament, and the FW, a political association, which only contest elections in the state of Bavaria. The high skewness of support towards the AfD is consistent with previous research that showed that the AfD dominates the Twitter conversation and has the highest support from automated accounts [26]. The pattern for Greece is more similar to the UK case, with the suspended accounts supporting and attacking the two largest political parties in the Hellenic Parliament, the New Democracy (ND) and the Syriza party. However, the latter had more support from both types of accounts. There is a third party with similar percentage support to the ND, the KKE, the communist party of Greece.

Overall, we observe that although the panorama is different for every country, the patterns are similar between suspended and coordinated accounts of the same country except for the UK's Conservative party and Mexico's PRI. This analysis allowed us to understand the intent of the politically motivated accounts, which serve as "keyboard warriors" by supporting or attacking their opponent. Although we find a relatively small number of coordinated accounts in comparison with the total users in the conversation, we observe that a substantial percentage of them had a high number of followers. This is a contrast to simple spam accounts, which have been found to have lower numbers of followers [28]. However, it may be that the accounts' followers are part of a bot-farm or a coordinated unauthentic campaign. Identifying if these followers are regular, non-spam accounts is outside of the scope of this paper, but should be considered in future research.

**Limitations** Detecting suspicious accounts on Twitter is not a straightforward task. While there are regular users that interact with Twitter in abnormal ways, there are malicious accounts that behave like normal users. The aim of our study was not to separate between accounts with benign or malignant intent but to identify unauthentic coordinated behavior. The introduced methodology is based on heuristics that try to minimize the number of false positives. However, by doing so, we are increasing the threshold for possible false negatives. We also limited the detection method by only taking into consideration three features from accounts from which we had enough information to be categorized as coordinated. Another limitation is that there are two biases in the collection procedure; first, from the Twitter API directly, and second, from having different sets of hashtags and users that we followed per country. A final limitation is that we only labeled 10% of the detected accounts for three countries, which may not represent the complete spectrum of the intent of these accounts. For example, even if we did not find coordination accounts supporting the UK's conservative party, this does not imply that they did not existed. It may be that our methodology did not detect them or did not appear in the 10% sample.

## 5 Conclusion

In this paper, we analyzed over 23 million political tweets collected in two-week periods before the general elections in the UK (2019), Greece (2019), Mexico (2018), and Germany (2017). We investigated coordinated behavior by looking at accounts that posted URLs, hashtags, and retweets in similar time patterns. Additionally, we identified accounts that were suspended by Twitter. We find a similar number of coordinated and suspended accounts for each country. Although the percentage of detected accounts is low in comparison to the total number of active users, between 20% and 40% of these accounts had high reachability with more than 1,000 followers. For Mexico, we find the highest number of suspended and coordinated accounts, but these have on average fewer followers than those form the other countries. For Germany, we find that the suspended and coordinated accounts were mostly supporting or attacking the far-right party, the AfD. For Greece, we observe the lowest level of Twitter engagement with political communication and a similar low level of coordinated activity. However, the UK is the country that has the lowest percentage of tweets from suspended (1.76%) and coordinated accounts (0.34%) when taking into account the complete political conversation.

This study shows an overall comparison between four different political ecosystems. Although the comparisons should be taken with caution, we consider that cross-country analysis is helpful to identify patterns of suspicious accounts. The main contribution of this study relies on the manual classification of suspended and coordinated accounts. Although there is extensive literature about unauthentic accounts of Twitter, there are no studies to the best of our knowledge that manually labelled the tweets from these accounts. They mostly relied on counting hashtags or mentions, which does not completely represent the intent of suspicious accounts. Therefore, we urge the research community to invest their efforts in including qualitative analysis of political tweets from unauthentic accounts.

## References


1. Bot or not? The facts about platform manipulation on Twitter, https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.html
2. The Twitter Rules, https://help.twitter.com/en/rules-and-policies/twitter-rules
3. Alorainy, W., Burnap, P., Liu, H., Javed, A., Williams, M.L.: Suspended accounts: A source of tweets with disgust and anger emotions for augmenting hate speech data sample. In: 2018 International Conference on Machine Learning and Cybernetics (ICMLC). vol. 2, pp. 581–586. IEEE (2018)
4. Bessi, A., Ferrara, E.: Social bots distort the 2016 U.S. Presidential election online discussion **21** (11 2016)
5. Bovet, A., Makse, H.A.: Influence of fake news in Twitter during the 2016 US presidential election. Nature communications **10**(1), 1–14 (2019)
6. Chowdhury, F.A., Allen, L., Yousuf, M., Mueen, A.: On Twitter Purge: A Retrospective Analysis of Suspended Users. In: Companion Proceedings of the Web Conference 2020. pp. 371–378 (2020)
7. Clement, J.: Leading countries based on number of Twitter users as of January 2020. *Statista* (2020), retrieved April 06, 2020 from https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/
8. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In: Proceedings of the 26th international conference on world wide web companion. pp. 963–972 (2017)
9. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botornot: A system to evaluate social bots. In: Proceedings of the 25th international conference companion on world wide web. pp. 273–274 (2016)
10. Enli, G.S., Skogerbø, E.: Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. Information, communication & society **16**(5), 757–774 (2013)
11. Ferrara, E.: Disinformation and social bot operations in the run up to the 2017 French presidential election. arXiv preprint arXiv:1707.00086 (2017)
12. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Communications of the ACM **59**(7), 96–104 (2016)
13. Francois, C., Barash, V., Kelly, J.: Measuring coordinated vs. spontaneous activity in online social movements
14. Giglietto, F., Righetti, N., Rossi, L., Marino, G.: It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. Information, Communication & Society **0**(0), 1–25 (2020)
15. Howard, P.N., Kollanyi, B.: Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum. arXiv preprint arXiv:1606.06356 (2016)
16. Keller, F.B., Schoch, D., Stier, S., Yang, J.: How to manipulate social media: Analyzing political astroturfing using ground truth data from South Korea. In: Eleventh International AAAI Conference on Web and Social Media (2017)
17. Le, H., Boynton, G., Shafiq, Z., Srinivasan, P.: A postmortem of suspended Twitter accounts in the 2016 US presidential election. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 258–265 (2019)
18. Marchal, N., Kollanyi, B., Neudert, L.M., Au, H., Howard, P.N.: Junk News & Information Sharing During the 2019 UK General Election. arXiv preprint arXiv:2002.12069 (2020)

19. Mønsted, B., Sapieżyński, P., Ferrara, E., Lehmann, S.: Evidence of complex contagion of information in social media: An experiment using Twitter bots. PloS one **12**(9) (2017)
20. Morstatter, F., Pfeffer, J., Liu, H., Carley, K.M.: Is the sample good enough? Comparing data from Twitter's Streaming API with Twitter's Firehose. In: ICWSM. Proceedings of the seventh International Conference on Weblogs and Social Media, Cambridge, Massachusetts (2013)
21. Pacheco, D., Flammini, A., Menczer, F.: Unveiling coordinated groups behind white helmets disinformation. arXiv preprint arXiv:2003.01313 (2020)
22. Pacheco, D., Hui, P.M., Torres-Lugo, C., Truong, B.T., Flammini, A., Menczer, F.: Uncovering coordinated networks on social media. arXiv preprint arXiv:2001.05658 (2020)
23. Papakyriakopoulos, O., Serrano, J.C.M., Hegelich, S.: Political communication on social media: A tale of hyperactive users and bias in recommender systems. Online Social Networks and Media **15**, 100058 (2020)
24. Ratkiewicz, J., Conover, M.D., Meiss, M., Gonçalves, B., Flammini, A., Menczer, F.M.: Detecting and tracking political abuse in social media. In: Fifth international AAAI conference on weblogs and social media (2011)
25. Serrano, J.C.M., Hegelich, S., Shahrezaye, M., Papakyriakopoulos, O.: Social Media Report: The 2017 German Federal Elections. TUM University Press, Munich, Germany (2018)
26. Serrano, J.C.M., Shahrezaye, M., Papakyriakopoulos, O., Hegelich, S.: The Rise of Germany's AfD: A Social Media Analysis. In: Proceedings of the 10th International Conference on Social Media and Society. pp. 214–223. SMSociety '19, ACM (2019)
27. Stieglitz, S., Dang-Xuan, L.: Social media and political communication: A social media analytics framework. Social network analysis and mining **3**(4), 1277–1291 (2013)
28. Thomas, K., Grier, C., Song, D., Paxson, V.: Suspended accounts in retrospect: An analysis of Twitter spam. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 243–258 (2011)
29. Volkova, S., Bell, E.: Identifying effective signals to predict deleted and suspended accounts on Twitter across languages. In: Eleventh International AAAI Conference on Web and Social Media (2017)
30. Zhang, C.M., Paxson, V.: Detecting and analyzing automated activity on Twitter. In: International Conference on Passive and Active Network Measurement. pp. 102–111. Springer (2011)