Ph.D. Dissertation in Social Data Science

Political Participation, Responsiveness and Discourse in the Social Media Age

August Lohse
Supervised by Morten Axel Pedersen, Hjalmar Alexander Bang Carlsen and Gregory Eady

July 2024

August Lohse
Political Participation, Responsiveness and Discourse in the Social Media Age
Ph.D. Dissertation in Social Data Science, July 2024
Supervisors: Morten Axel Pedersen, Hjalmar Alexander Bang Carlsen and Gregory Eady

University of Copenhagen
Faculty of Social Science
Copenhagen Center for Social Data Science
Øster Farimagsgade 5
1353 Copenhagen K

Det er fedt det er lys
Lys er rigtig fedt det noget jeg godt kan li'
Det har jeg brug for at jeg kan se
Det er fedt det er lys
Det det det dejligt det et kys
Et kys er rigtig godt det noget jeg godt kan li'
Det dog ikk' altid jeg har brug for det
Men det nu dejligt med et kys
Jeg er liderlig og har lyst
Lyst til dig fordi du den jeg bedst kan li'
For en gangs skyld utvetydigt
Elsker jeg dig inderligt
Er det dig jeg har lyst til
Men får pludselig lyst til at flygte
Flygt' fra dig jeg ved ikk' hvorfor jeg er sådan her
Men jeg føler mig indespærret
Det som om jeg ikk' kan trække vejret
Jeg synes ikk' du har fortjent det her
Et problem jeg ikk' kan løse
Det mit hjerte det gør ondt
Ondt fordi du sagde til mig det er forbi
Det havde jeg frygtet men ikk' forudset
Fire år et øjeblik
Det er mig og jeg er trist
Trist fordi… fordi sådan er det!
Min jasmin har det godt på bordet ved vinduet
Og tårerne trillede men nu er de tørret væk i solens stråler
Og alt det der før gjorde ondt
Er smeltet væk i solen til drømme og håb
Jaer min jasmin har det godt på bordet ved vinduet
Og blodet er størknet og såret er helet
I solens stråler
Og alt det der før gjorde ondt
Er smeltet væk i solen til drømme og håb
Det er livet det er lyst
Lyset smelter isen i mit bistre sind
Til jeg pludselig får lyst igen
Det er fedt det er lys

- Nicolai Rønde-Becker, 2024

Preface

This is an article-based dissertation containing two submission-ready preprint articles, one manuscript which is currently under review at Scientific Reports, and finally an article published in Social Science Computer Review. All the articles have been slightly adapted to fit with the format of this dissertation.

Paper 1: **August Lohse**, Tobias Priesholm Gårdhus, Snorre Ralund, Hjalmar Bang Carlsen (2024). *Mobilizing the Margins - Demographic, Cultural and Political Characteristics of Those Who Mobilize Their Opinion on Social Media.* Preprint.

Paper 2: **August Lohse**, Tobias Priesholm Gårdhus, Snorre Ralund, Hjalmar Bang Carlsen (2024). *The Like Effect - Political Participation and Responsiveness on Social Media.* Preprint.

Paper 3: Thyge Enggaard, **August Lohse**, Morten Axel Pedersen, and Sune Lehmann (2024). *Analyzing Differences between Discursive Communities using Dialectograms.* Preprint under review at Scientific Reports.

Paper 4: Sofie L. Astrupgaard, **August Lohse**, Emilie M. Gregersen, Jonathan H. Salka, Kristoffer Albris and Morten A. Pedersen (2023) *Fixing Fieldnotes: Developing and Testing a Digital Tool for the Collection, Processing and Analysis of Ethnografic Data.* Social Science Computer Review, 2023, doi: 10.1177/08944393231220488

Acknowledgements

This dissertation could not have been realized without the help of my exceptionally great friends, family, and colleagues. First, I would like to extend my thanks to my excellent supervisors, Morten, Hjalmar, and Greg. You have bestowed upon me an unusual level of trust and freedom to pursue my research interests, and you have been nothing but helpful and supportive along the way. It has truly been a privilege to be trained in the art of science by the three of you.

It has been a true pleasure to be part of the DISTRACT group. I have had the best time working in the DISTRACT Lab, and I have a hard time imagining a better team to be a part of. I would also like to extend my thanks to all my excellent colleagues at SODAS. It has been truly amazing being part of this one-of-a-kind research community, and I am going to miss you all. I would like to thank to all the excellent administrative personnel who have helped me over the years, as well as all the truly excellent student assistants at SODAS. Here I need to add extra emphasis on thanking Tobias - I could not have done it without you! An acknowledgment of the excellent Ph.D. groups at SODAS is also in order. You are all awesome, and it has been wonderful and very inspiring to work with you all over the years. I would also like to thank Ingo for creating and facilitating an outstanding collegial community for the Ph.D. students at SODAS.

A lot of people and institutions have welcomed me during my Ph.D., and I would like to extend my warmest thanks for that. Specifically, I would like to thank Josh Tucker and the CSMaP group for welcoming me into their group while I was in New York. Similarly, I would like to thank Maria Castro-Bleda and the VRAIN group for welcoming me at UPV. Finally, a warm thanks to the Santa Fe Institute for allowing me to participate in their summer school; I had an amazing time.

To my parents, I would like to thank you so much for encouraging me to be curious about the world and supporting me in all the ways possible. To the rest of my wonderful family, I will just say that I love you, and I look forward to showing you what I have been working on for the last 3 years. To my amazing friends, I would also like to extend a huge thanks for keeping me sane during these years. I think that you have played a large part in making sure that my Ph.D. time has been more chill than most.

Finally, I would like to thank Katinka for sticking with me through all of my frustrations, existential crises, new career plans every 2 weeks, and long periods of separation while I was abroad. You are the most amazing person I know, and I am continuously inspired by your bravery and ability to take us on new adventures, be it getting married or moving abroad. I cannot wait for what comes next - I love you!

Abstract

This dissertation investigates political participation on social media and introduces innovative methods for studying political engagement both online and offline. The first two studies examine the use of Facebook "likes" as a form of political participation, analyzing who uses likes to interact with politicians and whether these respond to this feedback. These studies reveal that politicians do respond to social media feedback such as likes, but the individuals engaging in this way represent a skewed segment of the public. Specifically, while liking politicians' posts on social media provides a means of political participation for minorities and those generally less interested in politics, it also attracts more ideologically extreme users. The third study proposes a novel approach to studying political participation by employing word embeddings to perform computational discourse analysis, which highlights the main differences between various groups based on their text. Finally, the fourth study explores how researchers can integrate ethnographic data with computational methods to investigate forms of political participation, such as face-to-face meetings, which do not leave digital or physical traces.

Resumé

Denne afhandling undersøger politisk deltagelse på sociale medier og introducerer innovative metoder til at studere politisk engagement både online og offline. De første to studier undersøger brugen af Facebook-"likes" som en form for politisk deltagelse ved at analysere, hvem der bruger likes til at interagere med politikerer, og om disse reagerer på denne feedback. Disse studier viser, at politikere faktisk reagerer på social medie-feedback såsom likes, men de personer, der engagerer sig på denne måde, repræsenterer et skævt segment af befolkningen. Specifikt giver likes på politikernes opslag på sociale medier mulighed for øget politisk deltagelse for minoriteter og dem, der generelt er mindre interesserede i politik, men det tiltrækker også mere ideologisk ekstreme brugere. Det tredje studie introducerer en ny tilgang til at studere politisk deltagelse ved at anvende moderne sprogteknologi til at udføre computerdrevet diskursanalyse, som fremhæver de væsentligste forskelle mellem forskellige grupper baseret på de tekster de producere. Endelig undersøger det fjerde studie, hvordan forskere kan integrere etnografiske data med moderne computerbaserede metoder for at undersøge former for politisk deltagelse, såsom ansigt-til-ansigt møder, der ikke efterlader digitale eller fysiske spor.

Contents

1 Introduction

In this dissertation, I investigate the role of social media in modern political participation and responsiveness and develop new computational tools to analyze political discourse and study offline political participation. My research is situated within the emerging interdisciplinary field of Social Data Science (SDS),[1] which aims at analyzing questions of social science with big datasets and/or advanced computational methods. In my research, I use large social media datasets to examine how millions of individuals engage with politicians through "liking" their content and how politicians respond to such feedback. Additionally, I develop a novel computational approach to summarize and interpret the extensive amounts of political texts on social media, aiming to identify and understand group differences. Finally, I experiment with combining computational tools and ethnographic methods in my study to investigate offline political participation.

The objectives of this dissertation are thus threefold:

1. To deepen our understanding of who uses social media feedback to mobilize their opinions and the societal impacts of this mobilization.

2. To develop a new computational method to identify and explore discursive differences between groups.

3. To demonstrate the application of computational tools in studying offline political participation, using ethnographic fieldnotes as a data source.

The dissertation is composed of four distinct papers focusing on political participation on social media and the development of interdisciplinary computational text analysis methods. Each paper engages with a different scientific literature, but all draw heavily on computational methods, especially methods for text analysis from the field of Natural Language Processing (NLP).

The first paper, *Mobilizing the Margins* (paper 1), presented in chapter 2, examines the role of social media feedback in political participation on social media. The study focuses on describing the type of people who choose to engage with politicians on social media in comparison to those who do not. In the paper, I analyze who "likes" politicians' posts on Facebook using one of the most comprehensive datasets of non-US social

[1] The terms Social Data Science and Computational Social Science (CSS) are often used interchangeably, though scholars in CSS typically have backgrounds in the natural or technical sciences, whereas scholars in SDS usually come from the social sciences. I personally use both CSS and SDS to describe my work, but I will solely use the term SDS in this dissertation for clarity.

media activity to date. The paper contributes to the literature on political participation and mobilization, primarily stemming from communication studies, political science, and sociology, by investigating everyday political participation rather than participation in social movements, which is often the object of study.

In the second paper, the focus shifts from who chooses to engage with politicians on social media to the societal consequences of this type of political participation. In *The Like Effect* (paper 2), presented in chapter 3, I explore to what extent politicians are responsive to this type of social media feedback. Here, I demonstrate that increased prior feedback on a topic is associated with a higher likelihood that the politician will revisit that topic, controlling for various other explanations. The paper contributes to the literature on social media and politics by examining the role of social media feedback in directing politicians' issue attention, and by highlighting the importance of the often overlooked direct feedback mechanism of social media.

In the third paper, called *Analyzing Differences between Discursive Communities using Dialectograms* (paper 3), presented in chapter 4, I study political participation from a very different angle. Here, I develop a new computational method to study why people disagree and what the disagreements are about. The method involves using word embeddings to identify differences, as well as creating discursive maps that show what the differences between groups consist of, facilitating computationally grounded discursive analysis. I show how this novel method can be used to study political participation on social media by investigating the discursive differences between supporters of the Republican and Democratic parties on the social media platform Reddit. This paper makes a methodological contribution to the field of NLP and computational cultural analysis by showing a new way to discover and analyze discursive differences.

Finally, in chapter 5, I present the paper *Fixing Fieldnotes* (paper 4), where I transition from studying political participation on social media to offline political participation. In this paper, I use a case study of a large political festival in Denmark to examine how to combine computational methods and ethnographic data, in order to study forms of political participation that do not leave physical or digital traces behind. I show that collecting ethnografic fieldnotes in a structured manner allows them to be easily combined with computational methods such as topic models and network analyses. This paper contributes to the growing field of computational anthropology and makes the case that ethnographic data is a rich but often overlooked data source, ripe for computational analysis.

In the remainder of this introduction, I outline the literature relevant to the four papers as well as their contributions to existing knowledge. I then introduce some of the main methods and tools utilized in this dissertation, such as how one works with text as data. Finally, I present the three main datasets I work with and discuss how I have addressed questions of GDPR and research ethics throughout this dissertation.

1.1 Online and Offline Political Participation

In this section, I introduce the four different scientific literatures that my papers address. While all four papers concern political participation, either directly as the object of study in papers 1 and 2, or by developing new methods for analyzing online or offline political participation in papers 3 and 4, they contribute to different branches of the scientific literature. Therefore, I devote a subsection to each literature below and discuss how my work contributes to these fields.

I begin by discussing the literature on political participation on social media, focusing on how social media can facilitate political protests and serve as a platform for social movements. In paper 1, I expand on this literature by studying how social media also functions as a platform for everyday political participation through interactions between politicians and the public.

Shifting focus, I examine the extensive literature on social media and politics, which tends to focus on the potential negative consequences of social media, such as polarization, hate speech, and fake news. I discuss how paper 2 contributes to this literature by adopting a more agnostic approach and focusing on how social media functions in modern-day politics by serving as a channel for political responsiveness.

Next, I delve into NLP and computational discourse analysis, explaining recent advances in NLP and their applications in studying culture, meaning, and discourse. I highlight the utility of word embedding models in analyzing massive datasets of text, such as those found on social media, and discuss my development of a new word embedding-based approach to study political discourse in paper 3.

Finally, I explore the literature on computational anthropology. I discuss how I experiment with combining ethnographic data and computational methods to study offline political participation in paper 4, and argue that engaging with ethnographic data from an abductive data science perspective allows researchers to gain the most from this highly contingent data source. Collectively, these subsections demonstrate how my work contributes to the current state of the art across multiple disciplines.

1.1.1 Political Participation on Social Media

The literature on political participation on social media has largely focused on the internet's and social media's transformative capabilities for political participation through social movements. Much work has been devoted to understanding how these platforms can facilitate both offline and online social movements, such as the Arab Spring and #MeToo (see for example: Eltantawy and Wiest, 2011; Suk *et al.*, 2021). However, less attention has been paid to how social media users employ the functionalities of the platforms, such as "posting," "sharing," "liking," etc., to participate politically every day without being part of a social movement. Here, I first discuss the current literature and then argue why a focus on everyday political participation on social media is warranted.

The low cost and speed of communication on social media make it ideal for facilitating offline social mobilization and easing collective action problems (Eltantawy and Wiest, 2011). Social media is especially important in authoritarian regimes where traditional media channels are often highly censored and the risk of collective action is higher (Tucker *et al.*, 2017; Khondker, 2011). The power of social media in this regard was clearly highlighted with the Arab Spring, where social media played a big part in spreading information, organizing protests, and mobilizing people to participate (Eltantawy and Wiest, 2011; Miller and Vaccari, 2020). Social media has since been integral in the organization of somewhat offline social movements such as Los Indignados in Spain and the Occupy Wall Street movement (Tucker *et al.*, 2017). Similarly, social media has been show to increase turnout for protests and to help spread protest movements from place to place (González-Bailón *et al.*, 2011; Larson *et al.*, 2019; Enikolopov *et al.*, 2020). Social media has thus proven to be a powerful mobilization tool for offline political participation.

Social media is not just a coordination tool for offline political participation however; participation in social movements on social media can also be a form of political participation in and of itself (Bennett and Segerberg, 2012). Many social movements either got started or primarily exist on social media. These social movements tend to follow a logic of connective rather than collective action, being more self-organizing and personal, without the need for a central governing body organizing events and protests or determining the strategy of the movement (Bennett and Segerberg, 2012). Instead, these types of social movements tend to rely heavily on forms of participation native to social media, such as posting political content, sharing other people's content, or providing feedback through comments or likes, all of which help spread a specific political message across the social network (Bennett and Segerberg, 2012). In this way, the affordances of social media allow for social movements such as the #MeToo movement (Suk *et al.*, 2021), Black Lives Matter (Shahin *et al.*, 2024), and the Yellow Vests (Morselli *et al.*, 2023) to spring to life, even without having clear leadership and organization from the get-go (Bennett and Segerberg, 2012).

This type of political participation on social media has been demonized in the popular press as "clicktivism" or "slacktivism," described as a lazy form of political participation with no real legitimacy and power in the offline world, which only serves to make the person feel good and look good to their social network (Morozov, 2009). Similarly, it has been argued that this sort of political participation can potentially crowd out other, more legitimate forms of political participation, as people spend their limited resources on participating in this way rather than others (Morozov, 2009).

However, increasingly, scholars are insisting that actions such as using the "like-button" on social media should, in some cases, be considered a legitimate form of political participation (Theocharis, 2015; Earl, 2016; Halupka, 2018; Halupka, 2014). Lonkila and Jokivuori, 2023, for example, choose to define online political action fairly broadly as: "*actions carried out by individuals who strive to raise awareness about or induce, support or oppose a change in social, cultural or political matter*" and argue that things such as Facebook likes serve as the foundational "nano-level" acts of political participation on social media (Lonkila and Jokivuori, 2023, p. 806). To be sure, the act of publicly liking political content is an expressive action, done to maintain "face" and show off among one's network on social media (Lonkila and Jokivuori, 2023), but it is also a political act when used with the intention of expressing political opinions and diverting attention towards a political cause, person, etc. (Lonkila and Jokivuori, 2023; Halupka, 2018).

While scholars thus theoretically recognize that engaging with political content on social media are acts of political participation, much empirical attention is still paid to how political participation on social media functions within the context of a social movement. Yet, engaging with political content on social media is a widespread and everyday activity that transcends social movements and represents individual acts of political participation that do not always follow either collective or connective logics (Margetts *et al.*, 2015; Theocharis, 2015). Every day people write about their opinions on political topics or engage with politicians on social media, without these actions being done in relation to any specific social movement. These everyday interactions, though sometimes overlooked, are significant forms of political participation that merit serious academic attention. This is particularly accute considering the emerging evidence that politicians are responsive to social media feedback such as people liking their posts (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023) and that this feedback is given by a small biased segment of social media users (see for example: Wojcieszak *et al.*, 2022; Hargittai, 2020; Bode, 2017). As such, overly focusing on the role of political participation as part of social movements, risks missing the politically powerful everyday participation on social on social media.

In paper 1 I therefore delve into this by examining the everyday political participation happening on social media. Specifically, I examine the characteristics that differentiate the people who choose to participate in politics in this way. In this research, I explore the demographic, cultural, and political characteristics of those who participate and compare them to the people who do not. Utilizing data from nearly all Danish Facebook users over almost a decade, my study offers one of the most extensive comparative analyses of politically active and inactive social media users in a non-US setting to date, thus significantly enriching the literature on political participation.

1.1.2 Social Media and Politics

In recent years, especially following Donald Trump's election in 2016, the research focus on social media's potential negative impact on politics has intensified, highlighting issues such as polarization, hate speech, and the spread of fake news (Miller and Vaccari, 2020).[2] However, social media also plays a more neutral role in politics by acting as a platform for political campaigns, enabling two-way interactions between politicians and the public, and providing real-time public opinion data to politicians. New evidence indicates that the risks associated with social media for society may have been exaggerated (Guess *et al.*, 2023b; Guess *et al.*, 2023a; Nyhan *et al.*, 2023). This suggests the need for a more balanced or agnostic perspective on social media's societal role and a stronger focus on the way social media actually functions in politics. Here I will explore some of the expansive literature on the negative effects of social media on politics, focusing on polarization, hate speech, and fake news as examples, after which I discuss how my

[2] For a detailed review, see (Zhuravskaya *et al.*, 2020).

work in paper 2 fits into this literature and represents a more agnostic research approach to social media research.

One of the most studied outcomes of social media in politics is polarization, facilitated by online echo chambers or filter bubbles.[3] A common argument has been that the personalization and recommendation algorithms governing the content presented to the users in social media feeds, create ideologically segregated filter bubbles that function as echo chambers where people just hear their own point of view repeated, which in turn polarizes people (Pariser, 2011; Shmargad and Klar, 2020).

While there is abundant evidence that social media is an ideologically polarized environment, the evidence on whether social media affects individual levels of polarization is somewhat mixed (Zhuravskaya *et al.*, 2020). One study finds that having people turn off social media does reduce their level of polarization (Allcott *et al.*, 2020). Another earlier study finds that exposure to opposing views on social media might actually be a cause of polarization, as opposed to the idea of self-enforcing echo chambers (Bail *et al.*, 2018), though a newer study finds the opposite (Levy, 2021). The recent large-scale US 2020 Facebook and Instagram Election Studies use some of the largest datasets to date to study this (Clegg and Nayak, 2020). These studies show that reducing the amount of content from like-minded sources has no effect on polarization on a large scale (Nyhan *et al.*, 2023). Similarly, changing the content of people's social media feeds by removing reshares, which come mostly from ideologically similar users, or changing the recommendation algorithms to work chronologically, does not affect levels of polarization (Guess *et al.*, 2023b; Guess *et al.*, 2023a). The most recent large-scale studies of potential individual polarizing effects of social media thus tend to provide evidence that exposure to content on social media, while highly ideologically segregated, does not *cause* polarization.[4]

A second topic extensively covered by the literature is the prevalence of hate speech on social media. Hate speech on social media is not only harmful to the social media users who are targeted by it, but it also flows into the offline world (Müller and Schwarz, 2021; Bursztyn *et al.*, 2019; Müller and Schwarz, 2023). A causal link between online hate speech and offline hate crimes has been established in the US, Germany, and Russia using exogenous variations in the uptake of social media (Müller and Schwarz, 2021; Bursztyn *et al.*, 2019; Müller and Schwarz, 2023). As such, combating hate speech on social media might be key to reducing hate crimes in the offline world. Political elites seem to play an important role in this, as they can both be the conduit of hate speech, as in the case of Donald Trump as shown by Müller and Schwarz, 2023, but they can also have a dampening effect by speaking out against it as shown by (Siegel and Badaan, 2020).

A final major concern about social media is the prevalence and spread of fake news, misinformation, and disinformation. The three concepts are somewhat contested, but in general, misinformation and fake news

[3]There is a huge literature examining both ideological polarization and affective polarization. Here I focus on the broad strokes of the literature and therefore do not engage with the differences. For a good read on this see (Iyengar *et al.*, 2012).

[4]While recent research tends to show no effect of social media on individuals' polarization, we cannot fully say that social media has not caused polarization to begin with, or that longer and larger treatments may simply be needed to change people's attitudes. For a discussion on this see (Guess *et al.*, 2023b; Guess *et al.*, 2023a; Nyhan *et al.*, 2023).

are used rather interchangeably and are defined as false or misleading news, imitating actual news, whereas disinformation contains the same basic features but also requires an intention to mislead (Golovchenko, 2020). While the public impression may be that fake news is abundant on social media, it turns out that the news diet of most people on social media comes from reputable sources. Studies have shown that for US citizens about 6% of their news diet on social media is fake news (Grinberg *et al.*, 2019), however, when accounting for other news sources such as television, fake news only makes up 0.15% of US citizens' general news diets (Allen *et al.*, 2020). A study of the 2016 presidential race also found that people on average only saw a few fake news stories during the length of the campaign (Allcott and Gentzkow, 2017), and another found that only about 1 in 4 Americans visited a fake news website at least once (Guess *et al.*, 2018), indicating that the problem is not as widespread as popular belief would have it.

However, the engagement with fake news is very unevenly distributed, with roughly 1% of users sharing fake news being responsible for 80% of the shares (Grinberg *et al.*, 2019). Fake news thus seems to be driven to a large extent by super users, sharing and engaging with large amounts of it. Another feature of the spread of fake news is that it tends to be shared and read by right-wing users (Grinberg *et al.*, 2019; Allcott and Gentzkow, 2017; Guess *et al.*, 2018). So much so that (Guess *et al.*, 2018) find that 6 out of 10 visits to fake news websites during the 2016 US presidential race came from the 10% most right-wing users (Guess *et al.*, 2018). Fake news thus does not seem like a general problem, but a huge issue in a very specific subset of the social media public.

While social media has mostly been studied with a focus on these and other potential negative consequences for society, another smaller share of the literature has focused more on how social media functions in society in general. One function of social media in politics that has been studied, is how it provides a two-way information channel between politicians and the public, allowing politicians to learn about the preferences of the public and vice-versa (Barberá *et al.*, 2019; Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023). Specifically, the public make their opinions known on social media by posting messages, which makes it easy and cheap for politicians to gather information on public opinion in real-time (Barberá *et al.*, 2019). Similarly, politicians also write about their views on policy on social media, content which the public interacts with by liking, sharing, commenting, etc. (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023). Politicians use this information to inform what issues they address in the future (Barberá *et al.*, 2019; Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023). In this way, social media functions as a channel for political responsiveness in society, a channel that we need to understand if we want to understand how public opinion and policy outcomes are related. In paper 2, I therefore build on these prior studies and seek to understand how this channel of responsiveness varies across topics, periods, and types of politicians. In this way, I attempt to engage agnostically with social media research, by analyzing some of the more fundamental functions of social media in politics, rather than directly studying its potentially negative effects.

1.1.3 NLP and Computational Discourse Analysis

The research within the field of NLP has exploded in the last decade. What is considered state-of-the-art now is so far from what was achievable 10 years ago that it is almost unfathomable. Much of the work in NLP relies on creating dense vector representations of words, sentences, documents, etc., that numerically represent the information of these text objects (Turney and Pantel, 2010). We call these vectors *text embeddings* because they are representations of text *embedded* in a latent semantic space, where distances between vectors are semantically meaningful. Over the last decade, two literatures have evolved side by side. One strand, which we could call the model-focused strand from within the pure NLP community, has largely been working on how to make embedding models that better capture the richness of human language and thus perform better when used to, for example, classify and/or produce text. This approach typically focuses on making models that improve performance on public benchmarks such as GLUE (Wang *et al.*, 2019).

The second strand, which we could call the corpora-specific strand of research, focuses on how we can use these embeddings to make structural analyses of language in a specific set of texts and what this can tell us about the nature of language, culture, society, and more. There are no benchmarks for such analyses, but better embedding models are obviously able to capture more minute relationships in the text data.

Here, I will discuss each in turn, first focusing on the literature on how one turns text into meaningful vector representations. Then I will discuss how researchers use these representations to analyze a specific corpus of texts and how I contribute to this field by developing a new method for computational discourse analysis, comparing two different corpora using text embeddings.

Embedding Text

Embedding text into a latent semantic space relies on *embedding models* that transform the text into a vector representation. Early "models" for embedding text relied heavily on the co-occurrence between words, documents, authors, etc. In one of the simplest possible cases, one could imagine a word-document matrix, with each row representing a word, each column a document, and each cell having the value of the count of the word in that document as seen in table 1.1. In theory, each row in this matrix is a vector representation of a word and thus a word embedding, and similarly, each column is a document embedding. However, there are a number of issues with such simple embeddings, such as the dependency on the marginal frequency of the words and length of the documents, the very large dimensionality of the matrix, and the dependency of the ordering of the documents and words in the matrix.

Word	Document 1	Document 2	Document 3
word1	2	0	3
word2	1	4	0
word3	0	1	2
word4	3	2	1

Table 1.1: Word-Document Co-occurrence Matrix

These issues can of course be dealt with, for example, by using various other measures than co-occurrence count such as Positive Pointwise Mutual Information (PPMI) (Niwa and Nitta, 1995). Similarly, the very high dimensionality can be reduced using dimensionality reduction methods, such as Principal Component Analysis (PCA) (Hotelling, 1933), t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008), or Uniform Manifold Approximation (UMAP) (McInnes *et al.*, 2020). These methods reduce the vectors to more manageable dimensions, which one can then work with.[5] As such, embedding text can be as simple as counting words across documents.

While these early methods yielded some good results on basic semantic tests (Bullinaria and Levy, 2007) and could capture relational similarities such as the difference between *apples* and *apple* being roughly similar to the difference between *cars* and *car* (Levy and Goldberg, 2014), the NLP community has largely moved on from using co-occurrence statistics to embed text, with the advent of deep learning models for text embedding.

All modern text-embedding models rely on deep learning. In its simplest form the seminal Word2Vec model, which I will describe in greater detail in section 1.2, uses a single-layer shallow neural network to create word embeddings. Specifically, one trains a neural network to predict a focal word in a sentence given the other words in the piece of text, the so-called context words, or the context words given a focal word. Afterwards one can then essentially use the values of the weights of the hidden layer in the neural network as the embedding vector for words (Mikolov, Chen, *et al.*, 2013; Mikolov, Sutskever, *et al.*, 2013). Other extremely successful early models such as Doc2Vec (Le and Mikolov, 2014) and the GloVe model (Pennington *et al.*, 2014), which I employ in paper 3, also build on this shallow neural network approach.

These models traditionally generate what are termed *static* embeddings, where each object (word, sentence, document, etc.) is represented as a singular, multi-dimensional point in latent space. However, the cutting-edge models now employ *dynamic* embeddings. This means that the models represent the objects based on their context. For instance, in these models, a word would receive different embeddings depending on the sentence it appears in, whereas a model like Word2Vec would produce one embedding for each word globally. Currently, transformer models, such as the slightly older BERT and RoBERTa models along with the new iterations like GPT-3.5 and GPT-4, dominate this space (Devlin *et al.*, 2019; Liu *et al.*, 2019; OpenAI *et al.*, 2024). These advanced transformer models, sometimes called Large Language Models (LLMs), produce dynamic embeddings that offer more nuanced text representations. Unlike simpler neural networks, these models are far from shallow. For comparison, a Word2Vec model contains just one hidden layer typically containing 300 parameters (Mikolov, Chen, *et al.*, 2013), whereas the original BERT model features 12 layers, each with 768 weights, and 12 self-attention heads, totaling approximately 110 million trainable parameters (Devlin *et al.*, 2019). The GPT-4 model, though not officially detailed, is rumored to contain 120 layers with an impressive 1.8 trillion trainable parameters (katerinaptrv, 2023).[6] Each of these layers can be used as a text embedding, meaning a BERT model, for example, produces 12 embeddings of

[5] For an excellent review of the early literature, please see Turney and Pantel, 2010.
[6] The detailed architecture of GPT-4 is not public and is based on speculative information.

each text piece. However, in practice, researchers often employ various techniques such as averaging or selecting specific layers in order to reduce this to a single representation (Ma *et al.*, 2019).

Another important difference between the early models and modern LLMs, is that LLMs are typically pre-trained on massive text-corpora through self-supervised methods such as masked language modeling. The idea behind this is that by pre-training a model to a general task such as next-word-prediction or masked-word-prediction, one can very effectively fine-tune the model to a more specific downstream task later (Devlin *et al.*, 2019). I make use of this approach, by fine-tuning LLMs to to predict a persons ethnic heritage given their name in paper 1 and to determine the topic of a social media post in paper 2.

While the power of pre-trained LLMs and dynamic embedding models is undeniable from a model-focused point of view, they are a bit more difficult to work with from a corpora-specific point of view. Specifically, when applying such a model to your textual corpus, it is difficult to discern what relations between words are derived from the specific corpora. versus what the models have learned from the massive pre-training dataset. Similarly, the embeddings themselves are more complex to work with, as each word is now represented by a distribution of embeddings rather than a single point in the latent space.

Computational Discourse Analysis

Researchers interested in studying culture and the meanings that cultures assign to concepts such as class, race, and gender have often resorted to text analysis (Lévi-Strauss, 1963). The texts a culture produces serve as a lens through which we can look at how that culture thinks and what meaning it assigns to different things. Studying culture in this way has traditionally been the domain of qualitative researchers utilizing methods such as deep hermeneutic reading (Ricoeur, 1981) and/or qualitative coding of texts, where researchers partition texts based on theoretical categories (Glaser and Strauss, 1967; Kozlowski *et al.*, 2019).

However, researchers have long argued that with the development of modern computational text analysis tools, this sort of cultural and discourse analysis has entered a new era (Mohr *et al.*, 2015; Lee and Martin, 2015). Besides allowing researchers to study large amounts of texts, computational methods also have unique advantages over traditional deep readings and qualitative coding (Lee and Martin, 2015). Specifically, Lee and Martin, 2015 argue that many computational discourse analysis tools allow for a more *scientific* approach to text analysis, which, at the same time, extracts patterns from the data in a reproducible way, yet still preserves the attribute of qualitative analysis where the researcher remains the *interpreter* of such patterns (Lee and Martin, 2015). Tools such as semantic networks, which structurally map the relations between words, documents, and authors, present the researchers with condensations of the data but still allow the researcher to make interpretative analyses of the data (Evans and Aceves, 2016). Much in the same way a map is used to summarize geographical information, culture can be summarized computationally (Hoffman *et al.*, 2017; Lee and Martin, 2015).

Word embeddings, as described earlier, represent the relations between words as positions (vectors) in a semantic space (Kozlowski *et al.*, 2019). As such, by examining the geometry of the vector space these embeddings exist in, we can examine the *geometry of culture*, as Kozlowski *et al.*, 2019 poetically puts it. This approach is what we could call corpus-specific as it aims at analyzing the structure of the language in the corpus one is using. One obvious application of this is language itself. Hamilton *et al.*, 2018, for example, uses this to study so-called statistical laws of semantic change in language, by looking at which words move around most in the latent space over time in a dataset of roughly 6% of all books ever published (Hamilton *et al.*, 2018). Using this, they are able to show that more frequently used words are more semantically stable over time and that more polysemous words are less stable in their semantic meaning, e.g., their latent position (Hamilton *et al.*, 2018).

Studying the geometry of a latent word embedding space can seem somewhat arcane at first glance, but this approach has been shown to map onto real-life phenomena. In their work, Caliskan *et al.*, 2017, for example, found a very strong relationship between the distance between gender words such as *woman* and occupation words such as *nurse* and the real-life gender biases in occupation numbers (Caliskan *et al.*, 2017). As such, they show that the way words relate to one another in these latent semantic spaces tends to mirror our cultural biases and can thus be used to study these (Caliskan *et al.*, 2017).

Word embeddings also provide us with a very powerful way to study culture intersectionally. They have, for example, been used to examine the biases and stereotypes in culture by examining the distances between the vectors representing focal words such as *man* and *woman*, and those of occupations, class, fatness, and more (Bolukbasi *et al.*, 2016; Caliskan *et al.*, 2017; Arseniev-Koehler and Foster, 2020; Kozlowski *et al.*, 2019). Similarly, a method largely popularized by Kozlowski et al. (2019) uses antonyms such as *rich* and *poor* to span a class vector, onto which all other words related to gender or race can be projected, in order to understand intersectional relations (Kozlowski *et al.*, 2019).

These methods have also been used to track the meaning of words and concepts over time (Azarbonyad *et al.*, 2017). Researchers have, for example, looked at how the word *gay* has changed meaning during the 20th century by examining its position in the latent space in relation to other words (Kulkarni *et al.*, 2014). Some have taken this idea to the extreme, mapping cultural patterns of gender and race over 100 years (Garg *et al.*, 2018a), and even 200 years in the case of (Jones *et al.*, 2020).

A final emerging use of word embeddings as a tool for computational discourse analysis, which I will touch upon here, is the use of word embeddings for comparing discursive communities. The idea, which was originally developed for machine translation between languages (Mikolov, Le, *et al.*, 2013), has since been applied to many other comparisons between discursive communities. The idea is that by comparing the words that are structurally similar or dissimilar between discursive communities, we can gain an understanding of what separates them. Researchers have, for example, used this approach to examine polarization in the US by examining word pairs that do not translate to themselves between corpora of Republican and Democratic supporters (KhudaBukhsh *et al.*, 2020). Some of the more recent attempts aim to formally measure the entire worldview of a discursive community, to then compare groups of people by comparing their total worldview structurally (Milbauer *et al.*, 2021). In paper 3, I develop a new way of approaching the task of examining differences between communities. Instead of only identifying specific words that are of interest, I develop a method that allows us to analyze what the difference *itself* consists of. In this way, paper 3 contributes to the literature by developing a new method for identifying and explaining group differences in latent semantic space. This method is specifically designed to help researchers understand what two groups of people disagree about using the text they produce, and as such it is excellent for analyzing political differences between groups on social media.

1.1.4 Computational Anthropology

Not all political participation leaves behind digital or even physical traces to study. In my work with paper 4, I study a large political festival in Denmark, focused on face-to-face interactions between the public and political elites (Folkemødet, 2022). To study this, I use ethnographic data, which I analyze using computational methods from NLP and network science. As such, my work places itself in the burgeoning field of computational anthropology.[7] However, there is still a lively scholarly debate over what exactly computational anthropology *is* (See for example Pedersen, 2023; Peponakis *et al.*, 2024). Therefore, I will briefly explain the different branches of computational anthropology described in the taxonomy of the field developed by Pedersen, 2023, and discuss how my work in paper 4 fits into this literature by engaging ethnographic data *as* a data scientist.

In his introduction to computational anthropology, Pedersen, 2023 argues that it mainly comes in three somewhat overlapping varieties. The first is the anthropology *of* computational methods, the second is the anthropology *with* computational methods, and finally the anthropology *as* computational methods (Pedersen, 2023). The first strand of research is the most straightforward, as this is more or less traditional anthropology that engages with computational methods as a field of study, for example, investigating data scientists themselves (Boellstorff *et al.*, 2015) or the algorithms they produce (Seaver, 2017) as the object of study. This is in many ways similar to the field of digital anthropology, which studies online communities such as online multiplayer games like World of Warcraft (Nardi, 2010) and online activist groups (Barassi, 2015; Peponakis *et al.*, 2024), though these studies tend to use digital platforms as fields of study rather than the object of them. Both digital anthropology and the anthropology of computational methods thus typically employ traditional anthropological methods and approaches from anthropology, such as ethnography, and simply engage with computational methods and the digital as the object or field of study (Pedersen, 2023).

The second approach of doing anthropology *with* computational methods typically refers to combining traditional anthropological approaches and computational methods in some way, though it can mean a lot of different things in practice (Bornakke and Due, 2018). One key way in which anthropological research can be combined with computational methods is through *complementarity*, meaning researchers analyze

[7] The terms computational anthropology and computational ethnography are often used interchangeably. I prefer to use computational anthropology as I consider anthropology a field and ethnography a specific method that has traditionally been employed a lot in anthropology but is not a necessary component of it.

some things via ethnographic methods and others using computational methods. Researchers then use the two analyses to triangulate a conclusion about the phenomenon they investigate (Munk, 2019). Examples of this approach are studies like the Copenhagen Network Study that employed network data on social interactions, surveys, and ethnographic data to analyze friendship formation (Stopczynski *et al.*, 2014; Blok and Pedersen, 2014). Researchers have also used ethnography to "thicken up" their big data from social media by combining data on diplomats' tweeting behavior with ethnographic and interview data on the same diplomats, in order to understand the interplay between the "front-" and "backstage" behavior of diplomats (Adler-Nissen *et al.*, 2021). Similarly, using ethnographic data like this can also help researchers access data on phenomena that do not leave a digital trail (Bornakke and Due, 2018). One example of such work is Albris *et al.*, 2021, who study the rise of a new party in Denmark, both by following their online presence and by shadowing and interviewing their communications employees, to understand *why* they were communicating the way they were (Albris *et al.*, 2021). In all these examples, the two approaches neatly complement one another and provide different insights about the phenomenon under question, but the two approaches don't interfere or affect each other much (Munk, 2019).

A similar but slightly different approach to combining anthropological methods and computational methods is to use either to validate the findings of the other. Ethnography can, for example, be a very effective tool for generating theories of behavior and meaning. These can then be tested or measured at scale using big data (Charles and Gherman, 2019). Conversely, researchers can find patterns in large datasets using computational methods, which can then be examined ethnographically in order to verify that the theories generated from the large-scale data actually map onto real-life behavior (Charles and Gherman, 2019). In this way, ethnographic data can thus be combined with big data as a means to find "ground truth," as data scientists typically only see trace data of behavior, and as such need to infer meaning from a few indicators of behavior, which could easily be confirmed by an ethnographer in the field (Bjerre-Nielsen and Glavind, 2022).

One can also combine anthropology and computational methods by applying computational methods to ethnographic data. Examples of such work include Santucci *et al.*, 2020's work using discrete-event models to study the general structure of Native American and Corsican myths, Albris *et al.*, 2021's work on students' digital device use during COVID-19, and Munk and Winthereik, 2022's study on the digitization of everyday life during the COVID-19 lockdowns (Santucci *et al.*, 2020; Albris *et al.*, 2021; Munk and Winthereik, 2022).

The final variant, anthropology *as* computational methods, proposes an even more integrated and radical approach, which is basically to do data science work as a form of anthropology (Pedersen, 2023; Pedersen, 2021). The main idea stems from the observation that anthropology and data science share a relatively abductive approach to science, in that they both tend to iteratively form theories from data, revisit data, and revise the theory (Paff, 2022). As such, researchers can use the pattern recognition abilities of computational methods to find patterns in the data, and the human theoretical knowledge to interpret them (Pedersen, 2023; Pedersen, 2021; Munk, 2019). This approach can be completely devoid of any traditional anthropological methods such as ethnography but remain within the confines of anthropology by approaching the computational methods as interpretative tools used to make sense of the lived human experience (Pedersen, 2023; Pedersen, 2021). Examples of such work include Hsu, 2014, who use various computational methods such as visualizing data on maps and computational analysis of music to study Asian American indie rock musicians, or Breslin *et al.*, 2022, who study the performance of trust on Twitter during COVID-19 lockdowns (Hsu, 2014; Breslin *et al.*, 2022).

I aim to contribute to computational anthropology both *with* and *as* computational methods. In paper 4, I employ computational methods on a corpus of ethnographic fieldnotes, thus combining computational and ethnographic methods. However, I use these methods to create maps, networks, topic models, etc., of the data, meant for qualitative interpretation by the researchers on the team. As such, I would argue that I approach the ethnographic data from a computationally exploratory or abductive point of view, attempting to construct various empirically founded interpretable aggregations of the data, in order to find general patterns for interpretation. In this way, my study provides an example of how researchers could approach *any* type of data in a way that is anthropological and interpretative, but driven by computational methods. My approach also provides an example of how one can combine ethnographic data and computational methods to study forms of political participation, such as face-to-face conversations, that do not leave digital or physical traces behind while still utilizing the power of modern computational methods.

1.2 Methods

In this section, I will discuss three key methodological themes of this dissertation. The first is how one can measure social science concepts using thin digital trace data such as Facebook likes. Using the examples of ideology and cultural preferences, I will describe how I and others have worked with this challenge. The second methodological theme I will discuss is how I work with text data. I will delve into how I use deep learning to embed and label text data, as well as how I utilize linear algebra and geometric intuition to perform computational discourse analysis with word embeddings. Finally, I will discuss my methodological considerations related to developing methods for collecting, digitizing, and computationally analyzing ethnographic data.

1.2.1 Measuring Concepts with Digital Trace Data

In the course of making papers 1 and 2, I worked extensively on how to measure concepts of theoretical interest from relatively thin big data. In these papers, I worked with a Facebook dataset, which I will describe in more detail in section 1.3, containing posts made by Danish Facebook pages, including the text, page name, and time. The dataset also contains all the likes made to these posts, including the username of the user who liked the post. Although I lack direct demographic and opinion data on the users, the pages and posts they endorsed by liking them, provide valuable insights about them (Bond and Messing, 2015). Using these digital traces of endorsement, I am able to measure various concepts of interest to social scientists such as political ideology and cultural preferences. Here I discuss each in turn, describing how I and others have tackled the issues of measuring them.

Measuring Ideology

Ideology is an important concept in the social sciences, and in papers 1 and 2, I examine both ideology among social media users and politicians. Measuring the ideology of politicians, voters, social media users, etc., has a long tradition. Traditionally, the so-called "latent-space models" have been the theoretical underpinning of most measures of ideology (see Duncan MacRae, 1958; Poole and Rosenthal, 1985). At a high level, latent space models conceive of people as having a so-called ideological "ideal point" in a latent ideological space of either one or a few ideological dimensions (Duncan MacRae, 1958; Poole and Rosenthal, 1985). Measuring ideology thus becomes a matter of spanning a valid ideological space and placing people within that space such that they are placed as close as possible to their unobservable ideal point, a process often known as ideological scaling or ideal point estimation.

The most famous measure for ideological scaling of politicians is undoubtedly the DW-NOMINATE scores, where the authors use votes on bills in the US Congress to estimate the legislators' ideal points (Poole and Rosenthal, 1985). Here the authors use legislators' votes of yea or nay on the available bills to estimate the ideological ideal points of each legislator using maximum likelihood estimation to identify the ideological position that best predicts their vote choice (Poole and Rosenthal, 1985). Other approaches rely on the text that politicians produce to place them ideologically. Rather simple approaches, for example, rely on counting ideologically salient words in political texts to place politicians in an ideological space (e.g., Slapin and Proksch, 2008; Temporão *et al.*, 2018). Some more modern approaches have relied on estimating the latent positions of politicians by creating text embeddings of their speeches, and then performing singular value decompositions on these speeches to identify the most important (semantic) dimensions of difference between politicians (Rheault and Cochrane, 2020). One of the most recent text-based approaches has been to use the generative power of GPT-3.5 (ChatGPT) to create ideology scores of politicians on custom policy scales, using the embedded ideological knowledge of the model (Wu *et al.*, 2023). All of these approaches tend to produce scores that correlate fairly well with what experts of politics judge to be the ideological position of the politicians as well as with the canonical DW-NOMINATE scores (Slapin and Proksch, 2008; Temporão *et al.*, 2018; Rheault and Cochrane, 2020; Wu *et al.*, 2023).

Scaling the ideology of ordinary citizens has traditionally been a more difficult task because they don't tend to vote on laws or write political texts (Grimmer and Stewart, 2013a). Social media data has proven to be an extremely valuable source of data in this regard, as social media provides researchers with extremely fine-grained data on the preferences of individuals, including their political preferences. Using the network structure of social media, researchers have scaled users ideologically by what they interact with using the like or follow button (Bond and Messing, 2015; Barbera, 2015; Barberá *et al.*, 2015).[8] Conceptually, these scaling models rely on an assumption of social homophily, meaning that they assume that social media users who like or follow the same political content tend to be similarly ideologically minded (Barbera,

[8] Some text-based methods for scaling ordinary citizens on social media have also been proposed, using the overlap of various keywords between users' posts and that of political elites (Temporão *et al.*, 2018).

2015; Bond and Messing, 2015). This is a fairly reasonable assumption, especially in the case of Facebook likes, which serve as endorsements of content (Bond and Messing, 2015). It is also a powerful assumption because it allows us to ideologically scale millions of social media users using thin observational data.

In Barbera, 2015 the author use the follower network of users and political elites on Twitter to estimate ideology. The author begins by defining a bipartite network G consisting of two sets of nodes, a set $U \in \{1, ..., i\}$ of social media users and a set $V \in \{1, ..., j\}$ of political elite users. Edges are then defined such that $E_{i,j} = 1$ if user i follows political elite user j and $E_{i,j} = 0$ otherwise. The ideological scores are estimated using a predictive model for $E_{i,j}$ based on the euclidian distance $||\theta_i - \theta_j||$ between the ideological position θ_i of user i and the ideological position θ_j of political elite j as:

$$Pr(E_{i,j} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \theta_j) = \text{Logit}(\alpha_j + \beta_i - \gamma||\theta_i - \theta_j||) \tag{1.1}$$

where γ is a parameter for how much ideological distance influences following behavior, and α_j and β_i are normalizing factors accounting for the base level of political elites followed by i and followers for j (Barberá *et al.*, 2015). In theory, one could solve this using maximum likelihood estimation or Markov Chain Monte Carlo as in Barbera, 2015, but this quickly becomes intractable and computationally impossible due to the number of parameters we need to estimate (Barberá *et al.*, 2015). Luckily, Barberá *et al.*, 2015 find that performing correspondence analysis on the normalized adjacency matrix of G provides an extremely efficient approximation method for estimating the optimal θ_i and θ_j compared to the more computationally heavy methods (Barberá *et al.*, 2015). The conceptual idea of this approach has since been expanded on by Eady *et al.*, 2023, who uses the sharing of news stories on Twitter to jointly estimate the ideological positions of politicians, news outlets, and social media users (Eady *et al.*, 2023).

In paper 1, I use the approach of Barberá *et al.*, 2015 to estimate the ideological positions of the Danish social media users who engage in politics, though I use Facebook likes rather than Twitter follows, and as such do not define $E_{i,j}$ as either 0 or 1, but rather the count of posts by politician j that user i has liked. Though due to normalization, the resulting adjacency matrix is similar. Using just likes as data, I am thus able to label the ideology of the social media users in my data.

Measuring Cultural Preferences

The idea of categorizing and quantifying people's cultural resources as latent positions was introduced in the seminal work of Bourdieu with the concept of *cultural capital* (Bourdieu, 1973; Bourdieu, 1986a). This concept has typically been described in three forms: embodied cultural capital (cultural knowledge, preferences, way of speech, etc.), objectified cultural capital (works of art, musical instruments, etc., that one owns), and institutionalized cultural capital (the titles and credentials one has) (Bourdieu, 1986b; Feng and Tan, 2024).

While the concepts are pretty clearly defined, finding good operationalizations of the various forms of cultural capital has typically been difficult (Vryonides, 2007). Institutionalized capital seems the easiest to measure, as it is documented in diplomas, ranks, etc., but the other two are more challenging. Researchers have, for example, been forced to rely on very crude proxies such as the number of books in a person's home in order to measure objectified cultural capital (Sieben and Lechner, 2019), or questionnaires about people's preferences in music, literature, and art to measure embodied cultural capital (Bourdieu, 1986a; DiMaggio, 1982). Many studies have also simply collapsed the three sub-categories and instead relied on parents' educational level as a crude proxy for total cultural capital (Vryonides, 2007).

In paper 1, I argue that social media is an invaluable tool for analyzing people's cultural preferences, thereby offering insights into their embodied cultural capital. Analogous to the way likes to political content can approximate a user's political ideology, likes to cultural content can similarly provide a window into their cultural tastes. Using their likes to non-political pages, I therefore partition all users into what I call *cultural preference clusters*, which I define as distinct groups of people characterized by a shared preference for a specific cultural domain.

I estimate these clusters using techniques from network science. Specifically, I create a bipartite network G, consisting of two sets of nodes: $U \in \{1, ..., i\}$ representing all the social media users, and $V \in \{1, ..., j\}$ representing all the non-political Facebook pages. I then add an edge $E_{i,j}$ between U_i and V_j, with a weight equal to the number of times the social media user U_i has liked a post made by V_j. I then project everything onto the page set, meaning I create graph G_v which only contains the nodes in V, but where the edges E_v are based on the nodes in U. In simpler terms, this means that I connect pages j_1 and j_2 based on the number of social media users who have liked a post by both j_1 and j_2. Now G_v is obviously going to be extremely densely connected, as most pages will have at least two users co-liking them, leading to the much-dreaded "hairball" graph (Coscia and Neffke, 2017). To deal with this, I then pick out the essential part of the network, which I call B_v, by performing network backboning with noise correction, which extracts the edges that are most likely to be signal rather than noise (Coscia and Neffke, 2017).

Using the backbone B_v, I proceed to cluster non-political Facebook pages based on the co-liking behavior of social media users. I employ the heavily used Louvain modularity optimization method (Blondel *et al.*, 2008), which is very fast compared to other clustering methods; however, conceptually, any community detection algorithm could be used. I then interpret these clusters, by examining the pages contained within them and assign meaningful labels to them, in a process similar to what one might know from topic modeling with text. Using this approach, I am able to examine the cultural profiles of each user based on their likes. Unlike the more traditional methods, which typically scale people's levels of cultural capital on a scale from low to high (Bourdieu, 1973; Bourdieu, 1986a), my approach results in a number of cultural preference clusters that defy simple ranking, such as preferences for content related to animals, cars, or politics.

Using relatively thin, but very big data, thus allows me to measure the ideology and cultural preferences of millions of social media users. Throughout Papers 1 and 2, I apply a similar approach in order to measure various other concepts, such as gender, partisanship, and political responsiveness.

1.2.2 Working with Text as Data

Throughout this dissertation, I employ text data in two primary ways. The first is the use of text as a tool for measuring important social scientific concepts, such as classifying the topic a politician is writing about in their social media post or determining the ethnic heritage of a social media user based on their name. In this context, the actual content of the text is not the primary focus; instead, it serves as a means to quantify the concepts under study. Over the past two decades, text classification methods have advanced dramatically due to the success of neural network-based deep learning methods in the field of NLP.[9] Consequently, the next subsection will outline the functioning of neural networks and the application of modern deep learning techniques for text classification. This includes a discussion on how contemporary models are pre-trained using self-supervision and how they can be fine-tuned for various downstream classification tasks.

The second approach focuses on analyzing the text itself as the core objective of the research. There is a longstanding tradition of utilizing text data for discourse analysis through qualitative methods, including hermeneutic readings and structural analyses of socially and politically relevant texts (e.g., Lévi-Strauss, 1963; Ricoeur, 1981; Glaser and Strauss, 1967). As discussed in section 1.1, recent scientific advancements have also introduced computationally based or assisted discourse analysis, predominantly using embeddings generated by deep learning models (e.g., Kozlowski *et al.*, 2019). The second subsection will delve into how I work with text embeddings as vectors, employing linear algebra to explore the culture and meaning of language within text corpora by analyzing the structure of the vector space.

Classifying Text with Deep Learning

Classifying text involves assigning labels to text pieces, such as hate speech/non-hate speech, noun/verb/adjective, or topic categories. Text classification typically comes in two forms: supervised and unsupervised, which refer to whether the models have a pre-labeled set of observations to work with or are supposed to classify or cluster the texts based on the content of the text itself. In the early days, researchers commonly used count and co-occurrence-based methods, such as sentiment analysis based on counting positive and negative words (Young and Soroka, 2012) or ideological scaling using a dictionary of politically salient words (Slapin and Proksch, 2008). Slightly more sophisticated machine learning-based methods, such as naïve Bayes and latent Dirichlet allocation, were also widely used for text classification and clustering (Blei, 2012; Grimmer and Stewart, 2013b). However, in recent years, deep learning methods based on neural networks have largely superseded these earlier approaches. While traditional methods remain useful, I primarily employ deep learning-based methods when working with text classification and thus focus on those here. Similarly, I primarily focus on the supervised approach, as this is what I mainly employ throughout my work in this dissertation, though I also work with unsupervised methods such as topic modeling in papers 2 and 4.

[9] I use the term *deep learning* to refer broadly to methods using neural networks, even though some, such as Word2Vec, are not very "deep" and are better described as "shallow" neural networks.

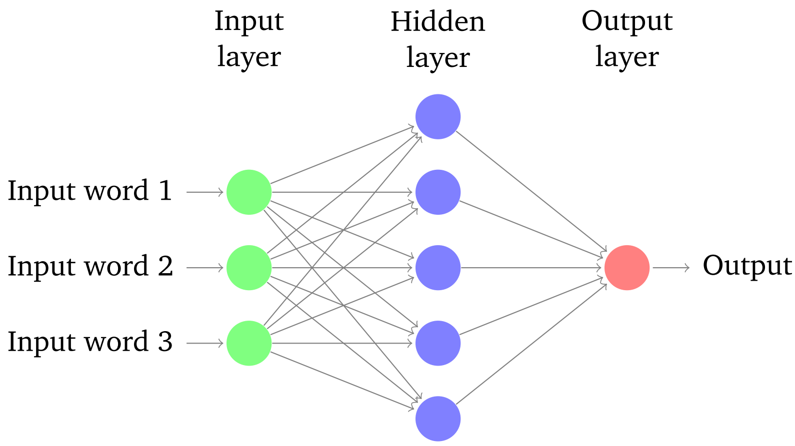


Figure 1.1: Schematic of an Artificial Neural Network

At a high level, neural networks are complex functions that take inputs, pass them through an interconnected network of activation functions, and predict an output (Bishop, 2006). Figure 1.1 shows a schematic of a simple neural network.[10] This network comprises an input layer, a hidden layer, and an output layer. Neurons (the nodes in the hidden layer) consist of a weighted sum of all the inputs and some biases (Bishop, 2006). Each arrow in the network represents a weight $v_{i,j}$, denoting the strength of the relationship between the input i and the neuron j, along with a bias of this connection $b_{i,j}$. In a network like the one in Figure 1.1, each neurons value would be calculated as:

$$neuron_j = \sum_{i}^{I} v_{i,j} \cdot input_{i,j} + b_{i,j} \tag{1.2}$$

Where the value of $neuron_j$ is thus a weighted sum of all the inputs I. After each hidden layer, a non-linear differentiable activation function is applied to all neuron values. Each neuron then passes this value to the next layer, continuing this process until the output layer is reached. The output values are translated into predictions using a final function, such as a softmax function, to map the neuron values to probabilities, indicating the likelihood of an input belonging to a specific class (Bishop, 2006). While the network depicted in Figure 1.1 is shallow and small, modern networks have many more hidden layers with numerous nodes in each, as well as more complex structures like attention layers that account for the sequence of words in a text (Vaswani *et al.*, 2023). In essence, a neural network is a highly complex and non-linear function that maps inputs to outputs.

A unique feature of deep learning, and machine learning more generally, is that these models can be *trained* from data to make better predictions (Bishop, 2006). This is typically done by calculating the gradient with respect to each weight using backpropagation and a chosen loss function, here denoted as J. This function measures how incorrect the model's predictions are compared to the true values (Bishop, 2006). For

[10] This is a basic feed-forward neural network for explanation purposes. More commonly used networks for text data include recurrent neural networks and the modern transformer architecture (Vaswani *et al.*, 2023).

example, using the common Cross-entropy Loss (CL) as the loss function, the gradient of the loss function with respect to the weight $v_{i,j}$ that connects neurons i and j for a dataset of K values can be written as:

$$\frac{\partial J}{\partial v_{ij}} = \frac{\partial}{\partial v_{ij}} CL \tag{1.3}$$

$$CL = -\frac{1}{K} \sum_{k=1}^{N} y_k \cdot \log(\hat{y}_k) \tag{1.4}$$

where y_k is the true value of observation k and $\hat{y}_k$ is the predicted value of observation k in the set of n observations. By combining this with an algorithm known as gradient descent, we can take "steps" towards better predictions by updating the weights in the network as follows:

$$v_{i,t+1} = v_{i,t} - \eta \cdot \frac{\partial J}{\partial v_{i,t}} \tag{1.5}$$

Here, η is the learning rate, which determines the extent to which we adjust the weights each time, and t is a measure of time steps, meaning $v_{i,t+1}$ represents the updated weights. In this way, the network is trained to minimize prediction errors iteratively until it stops improving. This process requires access to y_k, the *true* value of the observations, making it a supervised approach because the model is guided by y_k and learns to optimize towards these true values.

While this approach works well, obtaining ground truth labels when working with text typically involves manually labeling large amounts of text, which is a labor-intensive task. Additionally, supervised learning is task-specific, requiring a new model to be trained from scratch each time a new concept needs classification or when switching domains.

To address these challenges, researchers developed the idea of pre-training models using self-supervised training to create so-called foundational models that can easily be adapted and retrained for various tasks (Devlin *et al.*, 2019). The idea behind this pre-training approach is that for most tasks, the models need to learn general features of language such as structure, vocabulary, etc. (Devlin *et al.*, 2019). By pre-training models on large datasets of language, one can allow the foundational models to learn these features, and then fine-tune the model for specific tasks later. In this way we can significantly reduce the training time and data required to train a model for a specific task such as labeling some social media posts. Pre-training models in this way typically involves self-supervised training, which means training models without explicit labels by creating a loss function that focuses on features of the data itself.

One of the early models making use of this idea of self-supervised pre-training was the famous Word2Vec model. This model consists of a simple neural network with an input layer, a single hidden layer, and an output layer. The primary goal of Word2Vec is to generate vector representations of words that capture their relationships based on their usage in the corpus (Mikolov, Sutskever, *et al.*, 2013). Once trained, these models can then convert text into vector representations, which are useful for various downstream tasks. Word2Vec can be trained using two self-supervised methods: Continuous Bag of Words (CBOW) and Continuous Skip-Gram (SG) (Mikolov, Sutskever, *et al.*, 2013). In the CBOW approach, the model aims to predict a target word w_t from a surrounding context window of words w_{t-n} to w_{t+n}. The objective is to maximize the probability $Pr(w_t|w_{t-n}, \ldots, w_{t+n})$ over the entire corpus. This approach ensures that the vector representation of the target word w_t is similar to those of the context words, thereby positioning w_t close to its context words in the embedding space.

Conversely, the Skip-Gram approach reverses this logic. It seeks to predict the context words from a given target word w_t, maximizing the probability $Pr(w_{t+n}|w_t)$ for each context word w_{t+n} within a window around w_t. This training objective results in the context words being placed near the target word w_t in the embedding space (Mikolov, Chen, *et al.*, 2013; Mikolov, Sutskever, *et al.*, 2013). The SG approach was later further developed to include negative subsampling of non-context words which are then supposed to be placed far from the focal word (Mikolov, Chen, *et al.*, 2013; Mikolov, Sutskever, *et al.*, 2013). In this way, the models are trained by using the text itself at the prediction target, thus voiding any need for explicit labelling of the text.

One issue with these early models such as Word2Vec is that they represent words as singular positions in the latent semantic space. As a result, these models cannot distinguish between different meanings of the same word. For example, the word *fast* could refer to speed or the act of refraining from eating. Modern transformer models like BERT and GPT address this issue by using dynamic embeddings that consider the meaning of each word within its unique context. These models are trained in a conceptually similar way to earlier models, using an approach called masked language modeling. In this approach, a random word in each text piece is hidden, and the model must predict the missing word based on the context (Devlin *et al.*, 2019; Radford *et al.*, 2018). Thus, the masked word becomes the target y_k of the prediction, and the models optimize the weights to make $\hat{y}_k$ as close to y_k as possible.

Once these models have been pre-trained, they can be fine-tuned for specific tasks by, for example, adding a new output layer to the neural network and training that layer (Devlin *et al.*, 2019). This method allows researchers to adapt an already optimized language model to a particular task, which is more efficient and yields better results than training a model from scratch. Although this process still requires some labeled data, fine-tuning a model requires significantly fewer labeled examples and typically generalizes better to unseen data (Devlin *et al.*, 2019; Radford *et al.*, 2018). With sufficiently well-pretrained models, one can even use zero- or few-shot learning to label data with no or very few labeled examples. This approach is increasingly popular with the advent of extremely large and generalized pre-trained models like GPT-4, which can handle various tasks with minimal fine-tuning.

In both papers 1 and 2, I fine-tune pre-trained transformer models to classify large amounts of text. In paper 1, I fine-tune a BERT variant model to predict the ethnic heritage of a social media user based on their name. To do this, I manually label 10,000 names and fine-tuned the model for ethnic heritage prediction, allowing me to label the ethnic heritage of all social media users in my data. In paper 2, I use a similar technique for a more complex task: labeling topics in social media posts by politicians, while allowing for multiple topics to exist within each post. Here I used a semi-supervised approach, incorporating initial unsupervised topic models as well as extensive human reading and validation as recommended by Nelson, 2020 and Carlsen and Ralund, 2022. Using LLMs for these tasks allows me to analyze the participation patterns of millions of social media users and the responsiveness of politicians across hundreds of thousand of posts, something that would be impossible without the aid of these developments in NLP.

The Linear Algebra of Culture

As described in Section 1.1, there are many ways to create text embeddings, but they all share the feature that they are vector representations of text in a latent semantic space. As vectors are essentially positions in space, such vector representations of text allow researchers to study the meaning of a piece of text by its *relative* position to other pieces of text in this latent space. Here I will go through some of the arithmetic one can do with word vectors and describe how I use this to develop a new method for studying group differences.

As an illustrative example, imagine we have a word embedding of the word *woman*. Let's call this vector w_{woman} and assume that it is of the shape 1×300, meaning it is a vector in a 300-dimensional semantic space. The most simple thing that we can do might be to add and subtract vectors from one another. For example, we might add the vector for w_{tall} to the vector w_{woman} to obtain the vector $w_{\text{tall woman}}$ representing the meaning of the bi-gram *tall woman*. Using this, we can also examine some simple relational properties that should be true, such as

$$w_{\text{king}} + w_{\text{woman}} - w_{\text{man}} \approx w_{\text{queen}} \tag{1.6}$$

Which we can rewrite as

$$w_{\text{king}} + w_{\text{woman}} \approx w_{\text{queen}} + w_{\text{man}} \tag{1.7}$$

Here suggesting that the meaning of the word *queen* is equivalent to the meaning of the word *king*, minus the meaning of *man* and plus the meaning of *woman*, or said in another way: What *woman* is to *king* is what *man* is to a *queen* (Mikolov, Chen, *et al.*, 2013; Kozlowski *et al.*, 2019).

However, we are not limited to simple addition and subtraction; we can also calculate the similarity between word vectors using for example cosine similarity between two vectors such as between w_{woman} and w_{man}

$$S_c(w_{\text{woman}} \cdot w_{\text{man}}) = \frac{w_{\text{woman}} \cdot w_{\text{man}}}{||w_{\text{woman}}|| \cdot ||w_{\text{man}}||} \tag{1.8}$$

Thus calculating the similarity between the two vectors w_{woman} and w_{man} based on the angle between them (Kozlowski *et al.*, 2019). Garg *et al.*, 2018b for example uses this to analyze the cultural gender biases in our understanding of occupations, by comparing the distances between words related to gender and various occupations. By calculating and comparing relations between gender and occupation words, such as $S_c(w_{\text{woman}} \cdot w_{\text{nurse}})$ to $S_c(w_{\text{man}} \cdot w_{\text{nurse}})$, they are able to quantify the cultural gender biases in our understanding of various occupations (Garg *et al.*, 2018b).

We can also span new vectors in the latent space by taking the *offset* between two vectors, such as $w_{\text{woman}} - w_{\text{man}}$. This offset, which we could call $O_{\text{woman,man}}$ and which would also be of size 1×300, is what we could consider a "gender dimension" in the latent space, as it spans the space from man to woman and thus represents the *difference* in meaning between the two words (Kozlowski *et al.*, 2019). Once $O_{\text{woman,man}}$ has been established, we can examine other words' relation to this dimension by obtaining the scalar projection of those words onto $O_{\text{woman,man}}$ as

$$\alpha_i = \frac{w_i \cdot O_{\text{woman,man}}}{||O_{\text{woman,man}}||^2} \tag{1.9}$$

This intuitively captures how a word relates to the gender dimension from *man* to *woman*. This approach has, for example, been used by Arseniev-Koehler and Foster, 2020 to examine how gendered different concepts related to fatness are, showing, for example, that *anorexia* projects close to the *woman* end of the gender dimension while *burly* projects towards *man* (Arseniev-Koehler and Foster, 2020). In paper 3, I develop this idea even further and apply it to examine the differences between the meaning of words across two different corpora. While I introduce the method in more detail in Chapter 4, I will briefly repeat myself here to provide an intuition of how I work with word embeddings.

To explain, assume that we have two corpora of texts k_1 and k_2. To examine the difference in how a word w_i is used in k_1 and k_2, I take the vector representations of the word from each of the corpora $w_{1,i}$ and $w_{2,i}$. Using these word vectors as endpoints, I then take the offset $O_i = w_{1,i} - w_{2,i}$. This intuitively captures the differences between the way word i is used in the two corpora. To interpret the difference, I then project the dual set of embeddings of all other words onto this offset as

$$\alpha_i^k = \frac{E_k \cdot O_i}{||O_i||^2} \tag{1.10}$$

for $k = \{1, 2\}$ where E_k is the set of embeddings for all words in corpus k. This approach gives *two* scalar projections of each word onto the offset, which we can then plot against one another to create a type of discursive map that I call a *dialectogram*. This discursive map can then be used to interpret what the discursive difference between the two corpora's use of word i consists of.

The word w_i is what we call the focal word, and can be any word of prior interest to the study, or can be found inductively using the data. In paper 3, I also propose a new way to identify interesting focal words, introducing a distance measure I call *sense separation*. Sense separation is a novel way to measure the distance in meaning attributed to a word, using the distances in the embedding space. Intuitively, it measures to what extent the words that highly and uniquely co-occur with a focal word in one corpus also carry the same meaning as the focal word in that corpus.

To measure sense separation for a focal word i, I identify a set of context words that co-occur more with i than I would expect given the general co-occurrence distribution in that corpus as

$$EC_{i,j}^{k} = \frac{C_{i,j}^{k} \cdot N_c^k}{\sum_{h=1}^{N_w} C_{i,k}^{k} \cdot \sum_{h=1}^{N_w} C_{j,k}^{k}} > 1 \tag{1.11}$$

Where $C_{i,j}^{k}$ is the co-occurrence count between the focal word i and the context word j in corpus k, N_c^k is the total co-occurrence count, and N_w is the number of words in corpus k. As such, $EC_{i,j}^{k}$ is the set of context word embeddings that co-occur more with the focal word than we would expect given the relative frequency of the focal word and the context word. I obtain $EC_{i,j}^{k}$ for both corpora and then take the disunion of the two sets to obtain the sets of context words that uniquely co-occur with the focal word in each corpus, which I call HC_i^k. This set of context words intuitively captures the way one discursive community uses the focal word *differently* than the other community.

To examine how different HC_i^1 is from HC_i^2, I then project all the context words in these sets onto the offset O_i and take the average of their two scalar projections α^1 and α^2, such as in the case of HC_i^1

$$\overline{HC_i^1} = \frac{1}{|HC_i^1|} \sum_{j \in HC_i^1} \frac{\alpha_{i,j}^1 + \alpha_{i,j}^2}{2} \tag{1.12}$$

Finally, to obtain the measure of sense separation for focal word i, I simply take the difference between the average projections of the unique context words from each corpus by subtracting them from one another

$$S_i = \overline{HC_i^1} - \overline{HC_i^2} \tag{1.13}$$

In this way, I use the positions of words in the latent semantic space combined with relatively simple linear algebra and geometric intuition to construct a measure that intuitively captures the main difference in how a word is used across two discoursive groups. In paper 3, I will go into much greater detail with how I do this and show how I use this for the analysis of ideological polarization on social media. However, I hope to have provided here a relatively brief technical introduction to how I use deep learning models to work with text data, not only by labeling it, but also by creating vector representations of text that I then study in order to understand the meaning of the text itself.

1.2.3 Collecting Structured Ethnographic Data

The field of computational anthropology, as outlined in Section 1.1, is a relatively new discipline that employs computational methods in diverse ways. In paper 4, I contribute to this field by developing a protocol and tool for collecting structured ethnographic fieldnotes amenable to computational analysis and by highlighting the kinds of computational analyses this enables. Here, I will discuss some methodological considerations related to collecting structured fieldnotes, which I could not fit into the paper. Specifically, I will present some of the work I have done on developing and designing the data collection tool known as "Ethnote" used to gather data for paper 4 (*Ethnote* 2024). The data collection, detailed in Section 1.3, took place at a large political festival in Denmark in both 2021 and 2022, resulting in 144 fieldnotes collected by 16 ethnographers. Many of the design choices have their root in the methodological challenges arising from working as a large team of ethnographers and social data scientists, collecting data in the field.

The first thing to note about ethnographic fieldnotes is that they are an almost radically unstructured data format (Sanjek, 1990). In Figure 1.2, we can see an example of a real-life fieldnote related to flood forecasting. This note includes text, hierarchical and non-hierarchical lists, arrows indicating causal or temporal ordering, and finally also a drawing at the end. This does not even touch on the fact that the different pieces of text represent different things, such as empirical descriptions of rain gauge systems alongside notes to the author about an earlier IT diagram they need to remember. To make matters worse, there is typically no inter- or intra-ethnographer consistency in this structure, meaning that not only will two ethnographers examining the same phenomenon produce fieldnotes that are structured very differently, the same ethnographer will not even use the same structure across their own notes.

Computational analysis, however, requires that there is some level of *shared* structure across observations or data points to make the observations comparable. Therefore, processing fieldnotes computationally necessitates first making them conform to some form of structure. A key methodological challenge in paper 4 was thus how to create a protocol for fieldnote collection that allowed a large team of ethnographers to collect data in a similar computationally friendly manner while preserving as much of the flexibility of the fieldnote as possible.

The solution I propose in paper 4 is to structure the fieldnotes in a tabular data format based on predefined content categories. In this way, the dataset is structured such that each row represents a fieldnote and each column represents various types of text and metadata, as illustrated in Table 1.2. In paper 4, I call this a *tagging approach*, as each note is tagged with metadata variables that describe the fieldnote's content, such

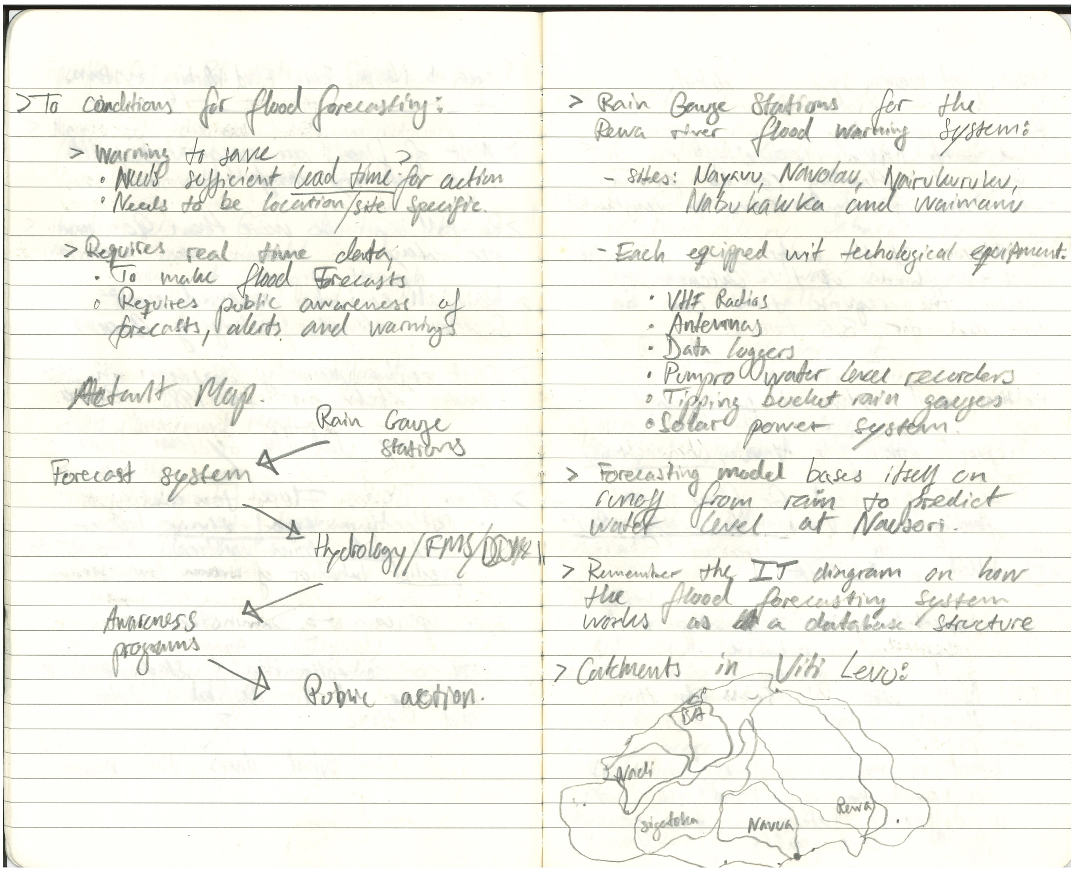


Figure 1.2: Example of a Fieldnote. The fieldnote contains ethnographic data related to flood forecasting (Albris, 2011)

as geographical position, time of day, author, etc. These tags were mandatory for the ethnographer to fill out, to ensure that each fieldnote contained the same information.

Organizing fieldnotes in a tabular format is excellent for computational analysis, but taking notes directly in a spreadsheet is not user-friendly or flexible. To address these issues, the team iteratively developed a note-taking application for smartphones and desktops that allows ethnographers to write their notes in a user-friendly interface while saving the data in a tabular format in the background. The app underwent several development stages during the course of this dissertation, starting as a simple survey tool and evolving into a professionally developed app, as seen in Figure 1.3.

The initial version of the app used in 2021 functioned as a survey tool with predefined questions serving as input fields. I used open-text fields for the observational notes and closed-form fields for metadata. For the 2022 data collection, I developed a web app version, which made it easier to incorporate custom features such as automatic time-stamping, the ability to edit prior notes, and the ability to read the notes of the other team members. Despite these enhancements, the user experience remained largely similar between the 2021 and 2022 versions.

Fieldnote	Name	Date	Time	Weather	Observations	Methodological Reflections
FN001	John Doe	2024-05-01	10:00 AM	Sunny	Observed high foot traffic in the main square.	Need to schedule observations at different times to capture variations.
FN002	Jane Smith	2024-05-02	02:00 PM	Cloudy	Noted a significant number of people using bicycles.	Consider including a survey about transportation methods.
...	...	...	...	...	...	...
FN010	Mark Brown	2024-05-10	05:00 PM	Windy	Evening rush hour showed a mix of pedestrians and vehicles.	Data collection during peak hours provides insights into commuter patterns.

Table 1.2: Conceptual Illustration of Fieldnotes in Tabular Format

After the 2022 data collection, the team identified several features that could be improved to enhance usability in the field. Both the 2021 and 2022 versions allowed only one fieldnote template or set of input fields. Consequently, the team had to spend considerable time designing and testing templates to account for all possible scenarios ethnographers might encounter in the field. However, having such a stringent predefined data structure makes it very difficult to collect data on the unexpected, which is one of the key strengths of ethnography as a method.

To facilitate more flexibility in the data collection process, the app was developed further into the app now known as Ethnote (*Ethnote* 2024).[11] Ethnote addresses many of the previous versions' inconveniences by allowing users to create multiple templates for various situations and modify them on the fly. The app also includes new features such as handling images and supporting in-text tagging rather than only tagging at the fieldnote level (*Ethnote* 2024). Similarly, the app also utilizes a project-based structure, meaning it allows teams of ethnographers to collaborate on creating templates, collecting data, doing analysis, etc. In this way, Ethnote provides a user experience that is much closer to the traditional fieldnote, while natively structuring the data and making it easy to work as an interdisciplinary team. As such, it is my hope that Ethnote will facilitate more interdisciplinary research leveraging the power of ethnographic data to study phenomena that do not leave physical or digital trace data behind.

1.3 Data, GDPR, and Ethics

All four projects presented in this dissertation utilize potentially sensitive data related to political opinion, gender, ethnic heritage, and many other sensitive topics. Furthermore, some of the data could potentially be used to identify individuals. Therefore, it is necessary to discuss how I handle the data in compliance with the General Data Protection Regulation (GDPR) as well as my ethical considerations for working with

[11] Ethnote is available through an application process on `https://ethnote.org/`

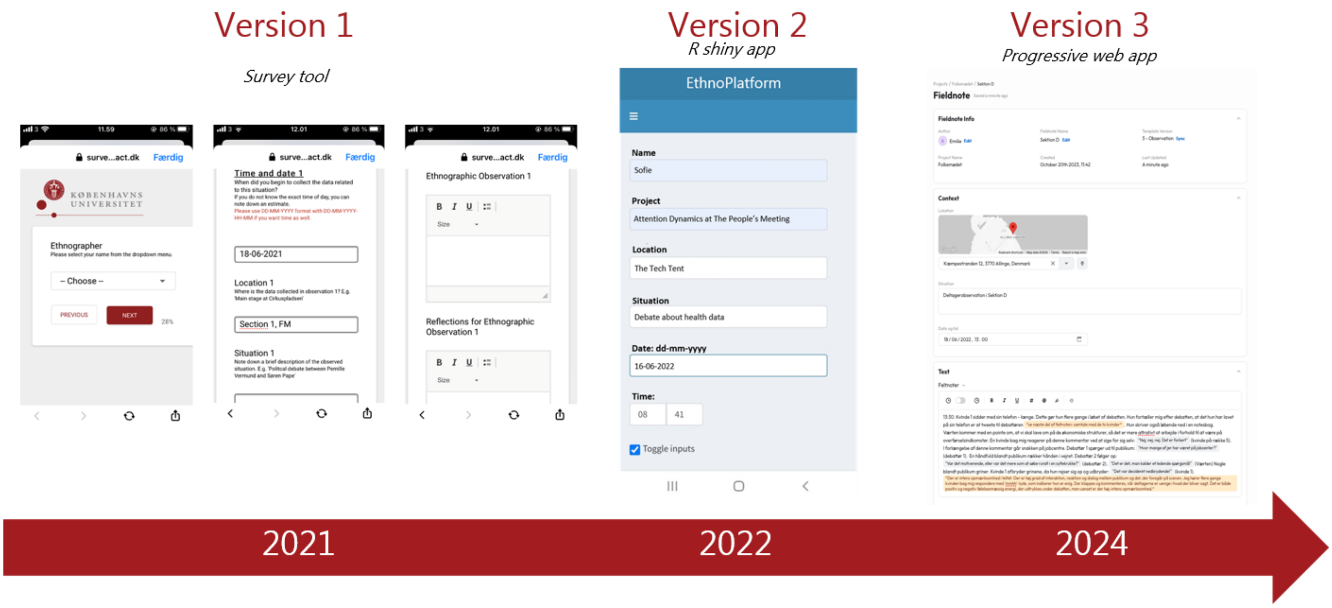


Figure 1.3: Evolution of the Note-Taking App

this data. In this section, I will introduce the three main datasets I use and then discuss the measures I have taken to ensure their legal, secure, and ethical handling.

1.3.1 The Data

In this dissertation, I work with three main datasets. The first dataset, employed in papers 1 and 2, is a comprehensive collection of Danish Facebook pages, their posts, and the associated likes and comments. The portion of the dataset used in paper 1 includes 55.360.264 posts from 375.105 pages and 1.395.933.289 likes from 4.239.090 users. In paper 2, I focus on 128,204 posts by 123 Danish politicians and the likes, shares, and comments these posts received. This dataset includes text from posts and comments, along with usernames which on Facebook are often close to people's real names. Using likes data, one can infer users' interests, potentially revealing sensitive information like sexuality, religious beliefs, and political opinions. In fact, I specifically use the usernames and likes of the users to infer gender, ethnic heritage, ideology, ideological extremity, partisanship, and cultural preferences of the users, making this highly sensitive data.

The second dataset originates from Reddit, focusing on the subreddits r/Democrats and r/Republican. This dataset consists of 1,088,778 posts and comments. The data contains usernames from the users and text content that is highly political, reflecting the vigorous discussions typical of these forums. Unlike Facebook, where real names often accompany user activity, Reddit's usernames (e.g., my own: themaplewarrior) are usually pseudonymous, enhancing user anonymity. This characteristic makes the data less directly attributable to real-world identities, thus offering a greater degree of privacy, though the nature of the content remains sensitive.

The final dataset contains 144 ethnographic field notes compiled from 16 ethnographers' notes from a political festival in Denmark. These notes include observations, researcher reflections, and paraphrased interview data, focusing on attendees' behaviors during political talks. The data is generally not sensitive, as it mainly contains descriptions that cannot be attributed to individuals such as "one person left because he looked bored." However, it still pertains to political behavior, making it somewhat sensitive, though it is anonymized with no identifiable personal information recorded.

1.3.2 GDPR and Ethics

For this section, I will focus on three aspects of the data as it pertains to GDPR and research ethics, namely; informed consent, the risk to the research participants, and the benefits of the research. I will argue that while informed consent is almost impossible in the case of this dissertation, the relatively minor risk due to the lack of an intervention and the security of data storage, coupled with the importance of the insights for society, places this dissertation on the right side of GDPR and ethics.

A key concern when working with data from humans is the idea of informed consent (Salganik, 2018). When collecting the ethnographic data, the team mainly worked in public spaces collecting non-personal data, and as such, consent was not actually a requirement under GDPR for most of this data. However, we did ensure that the data was collected with the full consent of the organizers of the political festival, as well as with consent from any people interviewed or shadowed as part of the data collection. Similarly, the ethnographers wore a sign on their clothes stating "I am collecting data" to signal their presence.

The issue of consent is a lot more complex when working with social media data (Salganik, 2018). When considering consent purely as a question of GDPR compliance, collecting social media data like mine for research purposes is not an issue. Users typically give consent to the collection and processing of their data when creating an account and accepting the terms of service, and at the same time, most social media data is considered part of the public sphere, as long as it is not in very small closed groups, a criterion which all of my data meets.

However, a more interesting question concerns the ethics of collecting social media data for research when it is impossible to get informed consent. Informed consent is a bedrock of the research ethical principle of respecting the participants in your research (Salganik, 2018). Yet, to get consent in my case I would have to contact the millions of users, whom I do not have any contact information for. It is thus not logistically possible for me to get consent from everyone. I have, therefore, to the best of my ability tried to respect the participants' (presumed) wishes in what other limited ways I could (Salganik, 2018). I have, for example, deleted roughly 10% of the Reddit data, as the users had chosen to either delete the posts/comments or make their user profiles private. I could have gathered this data using various internet archives, but I refrained from doing this, in order to respect the wishes of the people involved. Another way I have worked with this principle is by considering the context-relative information norms on these platforms, e.g., what the users of social media expect to happen with their information (Nissenbaum, 2009). It has been common knowledge for a long time that social media platforms use data such as likes to infer users' interests and characteristics, in order to serve personalized advertisements and conduct consumer research. As such, I would argue that doing non-invasive academic research where I utilize the same data to infer some of the same characteristics is within the informational norm on social media.

Respecting the participants is not the only ethical principle one must consider (Salganik, 2018). In my work, I have also put great consideration into the risk/benefit trade-off for the individual social media users and for society as a whole (Salganik, 2018). In terms of the risk to participants, the data itself is very sensitive and personal, which leads to a risk to people's privacy should the data be made publicly available. Both out of concern for GDPR and to minimize the risk to participants, I work to minimize the risk that the data can be used for anything malicious, by storing the data on a secure drive at the University of Copenhagen (UCPH), and not sharing it with other researchers outside the team or other entities outside of the project. This is true for both the social media datasets as well as the ethnographic dataset.

The specific way I use the data also poses very minimal risk to the participants. The social media studies are purely observational, meaning I never poke or prod at them in any way, and the data collection is completely unintrusive. As such, there is no risk of adverse effects to the participants themselves as a result of the studies done. The ethnographic data collection is similarly observational, as we do not attempt to affect the participants in any way. However, as the observers are more visible to the participants than in the social media case, there is a risk that the participants might have felt surveilled and had their experience at the political festival diminished as a result. However, as we only shadowed people with consent, participants should be able to simply move away if they did not want their behavior to be recorded.

On the benefit side of the equation, the two papers using Facebook data have the most clear contribution to the participants themselves. They help reveal that mobilizing your opinion on social media can affect what politicians pay attention to and that there are many inequalities in who mobilizes their opinion. Social media users can use this knowledge to make their voices heard in politics. Though, of course, it is unlikely that any but a few of the actual social media users end up reading the resulting papers, the potential is there. Paper 3 and 4 are more methodologically minded, and as such, provide less directly useful knowledge to the participants. However, as methodological contributions, they can be of use to future research and as such to society more broadly. Finally, I have made strides to communicate the results of paper 4 to a wider non-academic audience by presenting our work at the political festival I studied, at a board meeting for the festival as well as an annual gala hosted by the festival, where many stakeholders of the festival were invited.

During my work with the three datasets in this dissertation, I have thus not only strived to follow the rules laid out by GDPR but also attempted to conduct ethical research. While I have not been able to uphold all ethical standards such as informed consent, I have attempted to balance the ethical scales by minimizing the risk to participants and conducting low-risk observational studies. At the same time, I believe that the benefits of doing such research as mine greatly outweigh the potential risks of collecting data, contaning sensitive information like gender, ideology, and ethnic heritage. Similarly, I firmly believe that it is not only ethical but imperative that SDS scholars measure these sensitive concepts in order to understand how they function in society at a large scale. If we refrained from measuring things such as gender and ethnic heritage, we would be unable to understand how these categories influence the lives of people. That being said, we should of course be careful with how we handle the data and respect participants' wishes as best as possible. In sum, I believe that the way I have worked with my sensitive data in this dissertation has not only been within the confines of the GDPR but also within the research ethical norms of the field of SDS.

2

Mobilizing the Margins - Demographic, Cultural and Political Characteristics of Those Who Mobilize Their Opinion on Social Media

By August Lohse, Tobias Priesholm Gårdhus, Snorre Ralund, and Hjalmar Bang Carlsen

Abstract

In this article, we examine the demographic, cultural, and political characteristics of social media users who express their opinions by liking politicians' posts. We analyze the gender, ethnic heritage, cultural preferences, ideology, extremism, and partisanship of these users using a dataset of 1.395.933.289 likes and 4.239.090 users over an eight-year period on Danish Facebook, comparing politically active users with non-politically active users and the general population. Our findings reveal that minorities frequently use social media feedback to mobilize their opinions, highlighting the potential of social media as a democratizing tool. However, the study also indicates that ideologically extreme users exploit the same channels to mobilize their opinions, raising questions about whose public opinion politicians are listening to on social media.

2.1 Introduction

Social media platforms have fundamentally transformed how individuals express their views and engage in politics. Unlike traditional media, social media provides an immediate and interactive space for public discourse. On platforms such as Facebook and Twitter, users can instantly and unsolicitedly express their opinions by posting and tweeting, or giving feedback to content produced by political elites through likes, comments, shares, and other interactions.

Social media feedback, such as likes, shares, and comments, has become so integral to modern political participation, and many influential social movements of our era, including #MeToo, the Yellow Vests, and Black Lives Matter, rely heavily on these affordances to draw attention to their cause (Suk *et al.*, 2021; Morselli *et al.*, 2023; Shahin *et al.*, 2024). As such, some scholars argue these actions constitute the nano-level units of political participation in the modern age (Lonkila and Jokivuori, 2023, p. 806).

It is not uncommon for journalists, politicians, and academics alike to count up things such as likes and comments from social media and use these as a measure of public opinion (McGregor, 2019; Klanja *et al.*, 2018). Recent research also shows that politicians monitor these kinds of social media activities and use them to prioritize which topics to pay attention to (Barberá *et al.*, 2019; Schöll *et al.*, 2023; Ennser-Jedenastik *et al.*, 2022). However, the social media feedback political elites receive does not represent pure public opinion. The feedback comes from a highly biased subset of users who choose to voice their opinions, who are themselves a biased subset of the population who choose to use social media (McGregor, 2019; Hargittai, 2020; Barberá and Rivero, 2015). Consequently, the public opinion measured on social media is conceptually different from that obtained through traditional surveys, as it consists of unsolicited and directed acts of political participation from a specific population that targets specific political elites, causes, goals, etc. As such, public opinion on social media is what we call *mobilized* public opinion.

Thinking about social media feedback as mobilized public opinion raises a critical question: Who mobilizes their opinion on social media? As the people who choose to mobilize their opinion are a self-selected group of people, and they can exert influence over political elites in this way, it seems paramount to understand who these people are (Schöll *et al.*, 2023; Ennser-Jedenastik *et al.*, 2022). Some studies suggest that more privileged groups tend to use social media more frequently, which would lead us to exacerbate social inequalities by over-relying on social media for public opinion (Blank, 2017; Hargittai, 2020). Previous studies have also found that those tweeting about politics tend to be men and from the ideological right, introducing a different set of potential biases (Conover *et al.*, 2012; Barberá and Rivero, 2015). However, little is known about the individuals who choose to mobilize their opinions on social media by providing feedback such as likes to political elites' content.

In this study, we conduct one of the most comprehensive analyses to date on this topic, focusing on understanding who participates in politics by liking politicians' posts. We utilize a dataset of Danish Facebook likes that includes over a billion likes and millions of users over an eight-year period. By examining digital trace data, we analyze the demographic, cultural, and political characteristics of social media users who actively mobilize their opinions through likes, comparing them to those who do not and to the general public. This study provides essential insights into who participates in politics on social media by mobilizing their opinions through feedback. Notably, our research examines political participation beyond the usual Twitter and US contexts, offering a broader understanding of this global phenomenon.

We begin by introducing the literature on political participation both on and off social media to establish our expectations for who participates using Facebook likes. We then discuss our data and methods for measuring concepts such as gender, cultural preferences, and ideology using digital trace data. Finally, we present our results and discuss their implications for understanding how social media facilitates political participation.

2.2 Social Media Feedback as Mobilized Public Opinion

Social media allows people to participate in politics and voice their opinions in various ways. The *social* aspect of most social media platforms facilitates communication and organization for traditional offline political participation, such as protests, where it has been shown to help spread information and increase turnout (González-Bailón *et al.*, 2011; Larson *et al.*, 2019; Enikolopov *et al.*, 2020).

People also use social media to engage in social movements that primarily exist or originated online, such as #MeToo and Black Lives Matter (Suk *et al.*, 2021; Shahin *et al.*, 2024). Famously, Bennett and Segerberg, 2012 argued that these movements tend to follow a logic of connective rather than collective action, leveraging the social affordances of social media to form more personal, self-organizing movements with less central control (Bennett and Segerberg, 2012).

This form of primarily online political participation has sometimes been labeled clicktivism or slacktivism and described as a lazy form of participation that merely serves to show off within one's social network, potentially crowding out more legitimate forms of political participation (Morozov, 2009). However, most contemporary scholarship contends that this type of political participation is not merely performative but can be a legitimate form of engagement when directed at political elites or causes (See, for example, Theocharis, 2015; Earl, 2016; Halupka, 2018; Halupka, 2014; Lonkila and Jokivuori, 2023). Specifically, Lonkila and Jokivuori, 2023 argue that in these cases, social media feedback serves as nano-level political participation, raising awareness or communicating support/opposition towards political elites or ideas.

If we consider social media likes as nano-level acts of political participation, interpreting aggregate feedback as public opinion might initially seem reasonable. However, those engaging in these nano-level acts are not representative of the general public. Not only are the people interacting with politics on social media a selective subset of users (McGregor, 2019; Barberá and Rivero, 2015), but the active social media users themselves are also a selective subset of the population (Hargittai, 2020).

Thus, the public opinion expressed by liking political content on social media is an expression of a self-selected group of people who actively choose to mobilize their opinion on specific issues, persons, causes, etc. Consequently, the public opinion measurable on social media should more reasonably be considered a *mobilized* public opinion. While traditional notions of public opinion posit it as a general but passive sentiment existing within the public (McGregor, 2019), the public opinion on social media is actively articulated and mobilized towards specific political causes or elites with a distinct purpose. As such we need to ask questions not only of what the public opinion is, but also who chooses to mobilize their opinion.

Here we examine who mobilizes their opinions using Facebook likes. Likes are a very particular way to express one's political opinion, which in social movement and political participation literature, would be considered a low-cost, low-risk type of political participation. The costs refer to the resources and efforts required to participate, while risks refer to potential repercussions from one's participation (Wiltfang and McAdam, 1991; Verba *et al.*, 1995). A like is both cheaper and less risky compared to other forms of political participation, such as attending a demonstration or writing a comment on a political post. Demonstrating or commenting both requires time and effort and carries potential risks, such as being exposed to violence while demonstrating or public ridicule and hate on social media. In contrast, liking a post is less time-consuming and carries fewer visibility risks since Facebook users cannot comment on or react to others' likes. Therefore, Facebook likes may represent one of the lowest-risk and lowest-cost forms of political participation possible. This could have significant implications for who choose to use the like button as a mode of political participation.

2.3 Who Mobilizes Their Opinion

In this paper, we examine who chooses to mobilize their opinion on social media using likes, focusing on the demographic, cultural, and political characteristics of these individuals. Specifically, we investigate gender, ethnicity, cultural preferences, ideology, ideological extremity, and partisanship, as these concepts are central to explanations of political participation and core representational issues in many democratic societies. We review the current literature on each of these characteristics in turn, and develop our expectations based on this.

2.3.1 Demographic Characteristics

Research on gender differences in political participation has yielded varying results depending on the type of activity. Studies on voter turnout in established democracies have generally found few gender differences (Paxton *et al.*, 2007; Verba *et al.*, 1997), although some research complicates this finding (Kostelka *et al.*, 2019). Conversely, research on protest behavior, deliberative discussions, political talk, and political donations has consistently found that men participate more than women (Karpowitz and Mendelberg, 2014; Paxton *et al.*, 2007; Verba *et al.*, 1997; Dalton, 2002; McAdam, 1982; McAdam, 1992).

On social media, there are also gender differences in the types of participation. Some studies find that men are more likely to express themselves politically on social media (Lutz *et al.*, 2014; Strandberg, 2013; Vochocova *et al.*, 2016; Barberá and Rivero, 2015), while others do not find significant gender differences (Gil de Zúñiga *et al.*, 2014; Vesnic-Alujevic, 2012). More recently, Bode, 2017 demonstrated that for political content, there are no gender differences in acts with low visibility, such as likes, whereas gender differences persist in more visible acts such as posting (for a recent study supporting this finding, see Lilleker *et al.*, 2023). Thus, online participation appears to mirror offline participation, with no gender differences in more private acts like voting, but gender differences in more public acts such as demonstrating and engaging in deliberative discussions.

Research on the political participation of ethnic minorities, on the other hand, has found that ethnic minorities tend to participate less in both private acts such as voting (Cho, 1999) and public acts such as protests (De Rooij, 2012). One dominant explanation is the individual-level resource explanation (Verba *et al.*, 1995), which posits that ethnic minorities typically have fewer resources in terms of time and money, resulting in lower participation rates. Another explanation points to unequal access to political and civic organizations as barriers to participation, with research indicating that ethnic minorities are less likely to be invited to participate (Gomez-Gonzalez *et al.*, 2021) and face discrimination from party gatekeepers (Dancygier *et al.*, 2015).

In this context, social media provides a low-cost mode of political participation that does not require organizational embedding, potentially making it an ideal mode of participation for groups with fewer resources. However, reports have also documented the prevalence of hate speech targeted at ethnic minorities on social media (for example: Al-Rawi, 2022; Fonseca *et al.*, 2024; Analyse og Tal, 2024). While the cost of participation is lower on social media, the risk associated with political participation might not be as low for ethnic minorities as for other groups. Nevertheless, given that we are examining Facebook likes, where the user does not become natively visible and interactable, the risks of direct attacks are greatly reduced.

Given the low cost and risks associated with liking as a political behavior, we expect relatively small differences in participation based on demographic characteristics. Specifically, we do not anticipate that the gender differences observed in political tweeting behavior will manifest in the level of political liking, as liking behavior is relatively invisible. For ethnic minorities, we similarly expect smaller differences between ethnic minorities and majorities in social media feedback, due to the low cost, barrier to entry, and reduced visibility.

2.3.2 Cultural Characteristics

Research has long suggested that political participation is primarily driven by an interest in politics and broader societal issues (Campbell *et al.*, 1960; Verba *et al.*, 1997). Individuals who are engaged with current affairs and have a keen interest in social matters are more likely to be politically active. Cultural sociologists have extended this notion, arguing that a preference for political engagement is also influenced by one's social status and cultural capital (Bourdieu, 1984). To be politically active, one must not only understand and interpret complex political issues but also do so in a manner that aligns with social expectations.

In traditional media contexts, the relationship between interest in politics and political participation is quite logical. Information about political events and leaders is disseminated through newspapers, television, and other established media outlets, necessitating a high level of informational awareness for meaningful participation (Prior, 2005). Individuals focused more on consumerism and entertainment rather than current events often lack the necessary information to engage politically (Prior, 2005).

However, the dynamics may be different on social media. Emerging research indicates that while traditional measures of cultural capital predict political engagement in offline contexts (De Vreese and Boomgaarden, 2006), they are less predictive of informal civic and political activities on social media (Earl and Kimport, 2011; Elliott and Earl, 2018; Carlsen and Toubøl, 2024). Social media interactions might reduce the participation gap by allowing individuals to express support or opposition to political statements without the traditional gatekeeping of established media. Thus, it is plausible that a significant portion of the likes politicians receive come from individuals who have cultural interests than politics and news.

2.3.3 Political Characteristics

The relationship between political participation and characteristics such as ideology, extremism, and partisanship has long been a focal point in the social sciences (e.g., Verba *et al.*, 1995; Rogowski, 2014; Meer *et al.*, 2009; Whitford *et al.*, 2006). Research into traditional forms of political engagement, such as protesting, letter writing, and campaigning, often finds that individuals with left-leaning ideologies are more active participants (Meer *et al.*, 2009). Recent studies, however, suggest that this trend is not solely due to their leftist views but also to historical contexts where leftist parties opposed authoritarian regimes, thus fostering a stronger tradition of protest (Kostelka and Rovny, 2019).

Individuals with extreme ideological positions also tend to show higher levels of political engagement. For example, Meer *et al.*, 2009 found that those with extreme ideologies were more involved in all types of traditional political activities they measured. Other research corroborates these findings, indicating that ideologically extreme individuals are more likely to vote (Asano, 2022) and engage in political volunteering (Whitford *et al.*, 2006). Similarly, more partisan individuals also tend to be more willing to participate in politics through traditional modes such as voting (De Vreese and Boomgaarden, 2006; Rau, 2021; Jöst *et al.*, 2024).

On social media, research has found that users writing about politics politically tend to be more ideologically right-leaning (Conover *et al.*, 2012; Barberá and Rivero, 2015). Additionally, these right-leaning users are more likely to openly display their ideological beliefs on social media both by writing more about it and following political elites (Conover *et al.*, 2012). Research also shows that ideologically extreme and partisan users tend to engage more with politics on social media (Kim, 2016; Barberá *et al.*, 2019; Wojcieszak *et al.*, 2022). We expect that many of the dynamics of political participation on social media will carry over to liking political elites' social media content. Specifically, we anticipate that social media users who participate in this way will be more ideologically right-leaning, extreme, and partisan.

2.4 Research Design and Data

This paper employs a descriptive research design, focusing on comparing politically active social media users to their non-politically active counterparts, as well as to the general population. Our goal is to understand who mobilizes their opinion on social media by comparing these groups across the previously discussed demographic, cultural, and political characteristics.

To analyze the two groups of social media users, we utilize data from the most used social media platform in Denmark: Facebook (DST, 2023). Our dataset consists of 1.395.933.289[1] likes to 375.105 public Danish Facebook pages from 2010 to 2018. The dataset was constructed by gathering numerous Danish local pages and using snowball sampling to find other pages that users also followed. To ensure that we only included Danish pages, we used a standard language detection tool from the Gensim Python package (Rehurek and Sojka, 2010) to label all the posts in the data and applied a heuristic to remove pages where less than half of the posts were in Danish.

We created a set of active users who have liked at least 10 non-political pages, and then split these users into politically active and non-politically active, by assigning users who liked at least five posts made by Danish members of parliament (MPs) to the politically active group.

In Figure 2.1, we can see the number of political and non-political posts,[2] users, and likes over time. As we can see, the amount of activity has been continuously growing both terms of posts, likes, and users. Interestingly, the relationship between the amount of political content, users, and feedback seems to be relatively constant over time.

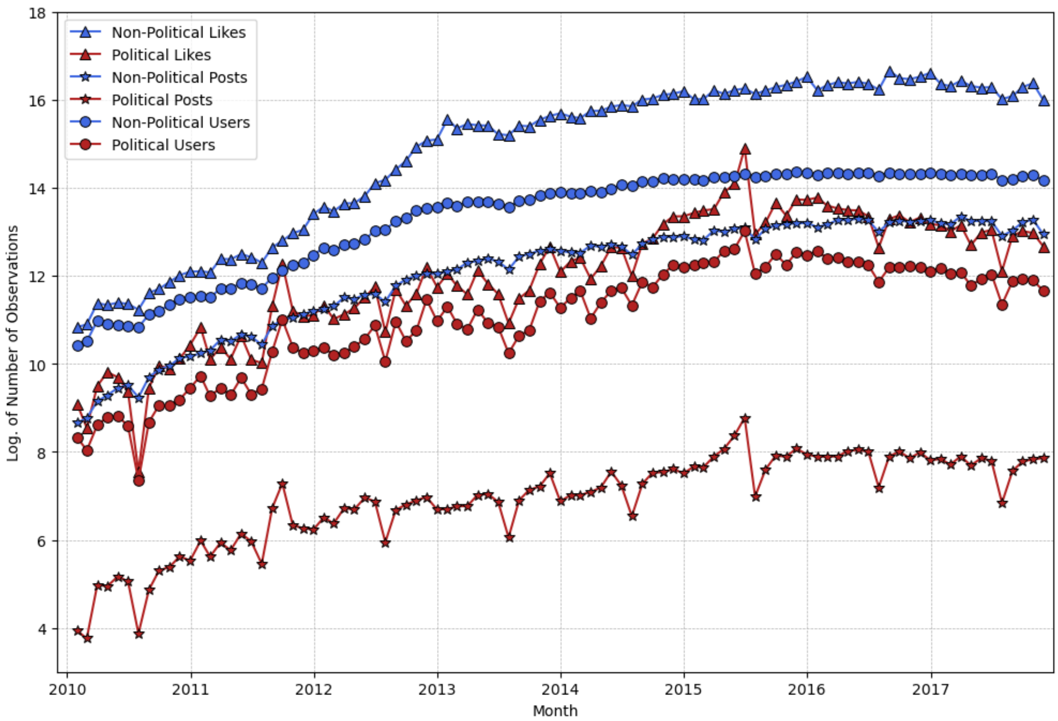


Figure 2.1: Number of Observations Over Time

In order to compare the social media users to the general Danish population, we used demographic data from Statistics Denmark and political opinion data from the most reliable Danish survey of politics, the so-called election survey (DST, 2024b; Hansen and Stubager, 2017). As the election survey is only

[1]This number counts the number of 612.433.604 likes made by non-politically active users to non-MP pages, 751.359.569 likes by politically active users to non-MP pages and 32.140.116 likes made by politically active users to MP pages.
[2]A political post is a post made by one of the MPs.

administered after each general election, we use the election survey collected after the 2015 general election and demographic data from 2015 as well (DST, 2024b; Hansen and Stubager, 2017). In the main paper, we compare the full social media dataset to the population data from 2015, but the results are similar if we use only the 2015 social media data, as seen in Appendix 6.1.6.

2.5 Methods

This paper is purely descriptive, and as such, the methods we employ are relatively straightforward, primarily involving counting and testing differences for statistical significance. A key aspect of our study is how we measure the demographic, cultural, and political characteristics of the three groups of people we compare. In the following, we will therefore describe how we measure these using digital trace data and survey data. For the non-politically active social media users, we do not estimate political characteristics, as we have very limited traces of their political preferences.[3] Similarly, we do not have any comparable data on the cultural preferences of the general population.

2.5.1 Measuring Gender and Ethnic Heritage

To estimate the gender of social media users, we employ a straightforward method based on their usernames. Denmark has relatively strict naming laws, which allows us to use a reference table containing all legal names associated with women, men, and both genders in Denmark (Familieretshuset, 2024). We cross-reference each social media user's username with this list of legal names and assign the corresponding gender if there is a match.

Measuring the ethnic heritage of social media users is more challenging than gender, as there are no official statistics or lists relating people's names to their ethnic heritage. Therefore, we fine-tune a large language model, specifically a BERT transformer model, to classify ethnic heritage based on the name of each social media user (Devlin *et al.*, 2019). We fine-tune it on a randomly selected subset of 10,000 usernames, which we manually label. Our model performs satisfactorily on our holdout test set of names, correctly labeling 94.6% of the users, with good performance on both Danish and non-Danish names, as indicated by an F1 score of 0.945. For a more comprehensive description of our approach to estimating gender and ethnic heritage, see Appendix 6.1.2.

2.5.2 Measuring Cultural Preferences

To measure the cultural preferences of social media users, we use their likes on non-political pages to cluster them into distinct cultural preference clusters. To do this, we create a bipartite graph of the 10,000

[3] We attempted the approach done by Barberá *et al.*, 2015 to estimate ideology scores for non-politicially active users, but were unable to create reliable scores. We suspect this may be due to the multiparty nature of the Danish political system. Future work will determine if this is indeed the case.

most liked non-political pages and all the social media users. We then project the graph onto the page graph, adding edges with a weight equal to the number of user who co-liked two pages. We extract the network backbone using noise-corrected backboning (Coscia and Neffke, 2017), and cluster this graph using the Louvain modularity optimization method (Blondel *et al.*, 2008).

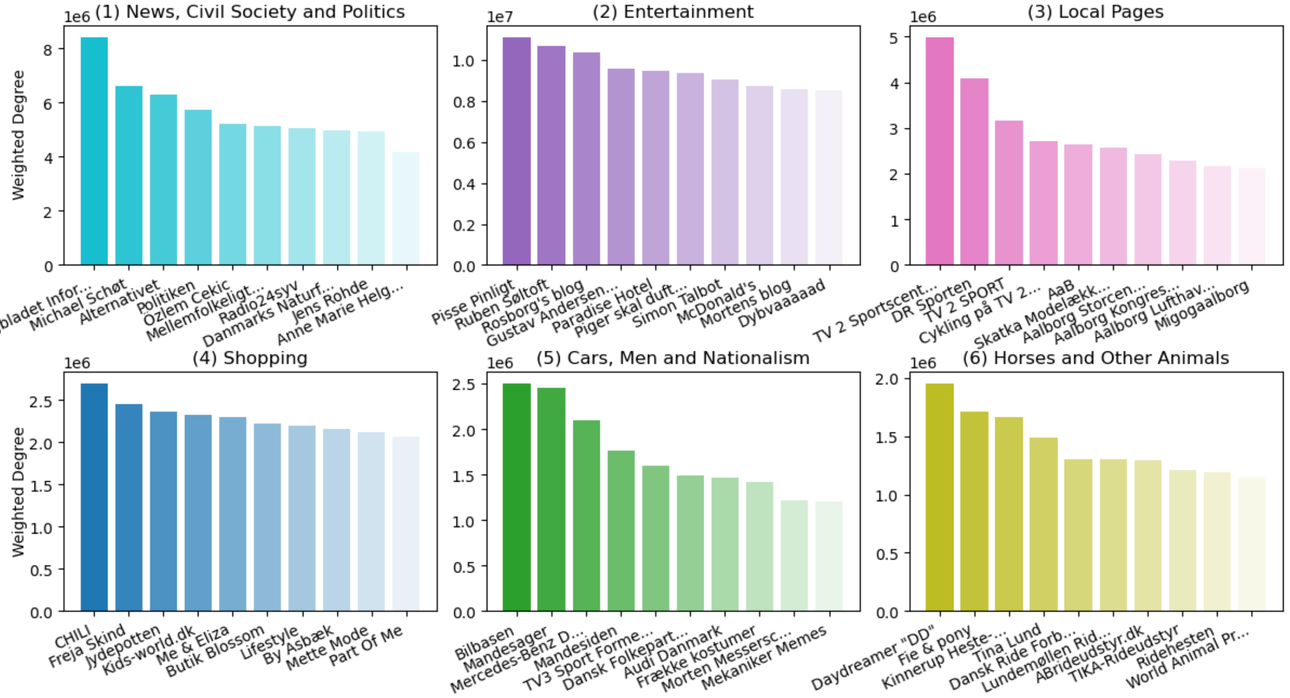


Figure 2.2: Cultural Preference Clusters

Using this partition, we combine the found clusters of pages to construct six coherent cultural preference clusters, as seen in Figure 2.2. The clusters center on News, Civil Society and Politics (1), Entertainment (2), Local Pages (3), Shopping (4), Cars, Men, and Nationalism (5), and Horses and other Animals (6). We then assign each user to a cultural preference cluster by assigning them to the cluster with the pages whose posts they have liked the most. Users assigned to cluster 1 are thus the users who are primarily interacting with political and civics-related content, outside of the MPs' pages.

For a full overview of the initial clusters and the labeling process, see Appendix 6.1.3. For robustness, we also assign users to clusters by weighing the pages by their inverse degree to limit the influence of large pages. This makes very little difference, however, as seen in Appendix 6.1.4.

2.5.3 Measuring Ideology and Extremity

To measure the ideology of the politically active social media public, we utilize the widely used methodology developed by Barbera, 2015. This method estimates ideology scores by assuming that social media users who like or follow the same political social media content are similarly ideologically leaning. We use the approximation method developed by Barberá *et al.*, 2015, which substitutes the original computationally heavy Bayesian estimation method with what is essentially a correspondence analysis of the bipartite interaction network structure between users and politicians (Barberá *et al.*, 2015; Greenacre, 2017).

To ensure the latent space spanned by the correspondence analysis is ideologically meaningful, we first map the space using highly politically active users ($N \geq 10$) and then project all other users into this ideological space, in line with the original paper (Barberá *et al.*, 2015). The final scores are then normalized to a 0-1 scale.[4]

For the general public, we measure ideology using the aforementioned election survey, where respondents self-identify their ideological leaning on a scale from 1-10 (left-right). To facilitate comparison with the politically active social media users, these values are normalized to a 0-1 scale.

We measure ideological extremity for both the politically active social media users and the general population by simply taking the distance to the middle of the normalized ideological scale. This yields a scale from 0-0.5, where a higher value means a more extreme ideological position.

For robustness, we also calculate the ideology and extremity of the politically active social media users based on party identification. To do this, we calculate the average ideological placement of the political parties using the election survey respondents and then assign the politically active users ideology scores based on the parties they have liked. The results show more ideologically extreme users on the right side of the spectrum than our primary method as seen in Appendix 6.1.5.

2.5.4 Measuring Partisanship

To measure the partisanship of the politically active social media users, we use the distribution of their Facebook likes among the various Danish parties. Specifically, we use the ratio of likes each user devotes to their most liked party:

$$P_{\text{SoMe}} = \frac{PL_i}{\sum_j^J PL_j} \tag{2.1}$$

where P_{SoMe} is the partisan score, PL_i is the number of likes for the user's most liked party, and i is the user's most liked party out of the set of J parties. Intuitively, this measure captures the extent to which someone exclusively likes posts from one party (resulting in a value of 1) or spreads their likes equally among the parties (resulting in a minimal score of $\frac{1}{J}$).

[4] As correspondence analysis does not ensure directionality, we inspect the top and bottom 1,000 users and orient the scale based on their likes to politicians' pages, by flipping it if the top 1,000 users have liked more left-leaning politicians than the bottom 1,000. For a detailed explanation of this ideology measure and the approximation function, we refer readers to Barberá *et al.*, 2015.

For the general population, we utilize a battery of questions in the election survey, where respondents rate all parties on a feelings thermometer scale from 1-10 (Hansen and Stubager, 2017). To calculate partisanship, we compare the evaluation each person gives to the party they voted for to the ratings of all the other parties, capturing the average distance from their own party evaluation to the others:

$$P_{\text{General public}} = \frac{\sum_{j \neq i}^{J}(PT_i - PT_j)}{J - 1} \tag{2.2}$$

where $P_{\text{General public}}$ is partisanship, PT is the party thermometer score for each party, and i subscripts the party the person voted for out of the set of J parties.

2.6 Results

In this section, we present our findings by comparing politically active users to non-politically active users, as well as to the public at large. For each analysis, we examine the differences between the groups both overall and over time by aggregating the number of unique users and the number of likes each month for each group. All comparisons mentioned in the text are statistically significant at the 0.05 level unless otherwise specified, as seen in Appendix 6.1.1.

2.6.1 Gender

When comparing the share of women in the population of Denmark to their representation on social media, we can see that among the non-politically active users, there are significantly more women than we would expect given their representation in the population. In terms of users, women are overrepresented on social media by 8.3 percentage points, as shown in Table 2.1. Even more stark is the difference when comparing the share of likes made by women to what we would expect given the population. Here we see that women are overrepresented among the active users by an astonishing 19.83 percentage points. Women are thus highly overrepresented on non-political social media in general.

When examining political participation, we see a similar pattern. Women are overrepresented by 5.77 percentage points in terms of users and 1.21 percentage points in terms of likes as seen in 2.1. Women are thus significantly overrepresented on social media, both in the political and the non-political domain.

However, if we consider the representation of women users on non-political social media, we see that there are fewer women users active in politics than we would expect. Specifically, we see that there are 2.53 percentage points fewer women users among the politically active users than we would expect given the share of women users in non-political Facebook. Similarly, if we examine the share of likes made by women, the difference between politically active women and non-politically active women is remarkable.

	Political Users	Non-Political Users
Share of users (women)	56.10%	58.63%
Share of users (men)	43.90%	41.37%
Share of likes (women)	51.53%	70.26%
Share of likes (men)	48.47%	29.74%
Avg. likes per person (women)	40.91 (pol. likes)	226.98 (non-pol. likes)
Avg. likes per person (men)	49.15 (pol. likes)	136.83 (non-pol. likes)

Table 2.1: **Representation of gender on social media.** The population distribution in 2015 was 50.33% women and 49.67% men (DST, 2024b)

In Table 2.1, we can see that while women are responsible for 70.16% of the non-political likes, they only make up 51.54% of the political likes, a difference of 18.62 percentage points.

This suggests that there is a strong opt-out/in effect in the political domain on social media, meaning that while women are very active on social media in general, they choose to participate in the political domain less often than men. On the user level, the tendency is weaker, but at the nano-level of the like itself, there is a truly massive gender bias, where women users mobilize their opinion way less than men. This is also evident when comparing the average likes in Table 2.1, where we see that the average politically active man has almost 10 more political likes than the average woman, while the average non-political woman has roughly 90 more likes than the average man.

If we now turn to examine these dynamics over time in Figure 2.3, we can see a number of interesting trends in gendered political participation on social media over the years. The first thing we see is that the share of women users among the non-political users has remained relatively stable over the years, whereas the share of politically active users who are women has varied more, from around 60% in 2010-2012, and then hovering around and above 50% in the rest of the periods.

If we examine the political and non-political likes over time, we can see that the share of likes from non-political women has slowly but surely increased over time. For the politically active users, the share falls over time, in a similar fashion to the shares of users.

We also see that the difference between the share of women, users and likes from women becomes larger among the non-politically active users over time. This indicates that non-politically active women become more active over time compared to the men. However, when looking at the politically active users, we see the opposite trend, with the share of likes by politically active women consistently falling below the share of users after 2012. This means that while women have in general been increasingly dominating the social media space in terms of likes, this trend does not manifest itself in the political sphere.

As such, we find that social media is disproportionately used by women users, who disproportionately like more than men. However, while women users are still overrepresented among the politically active users, women participate in politics on social media much less than we would expect given their activity on social media in general. While women have over time become more and more dominant in terms of the share

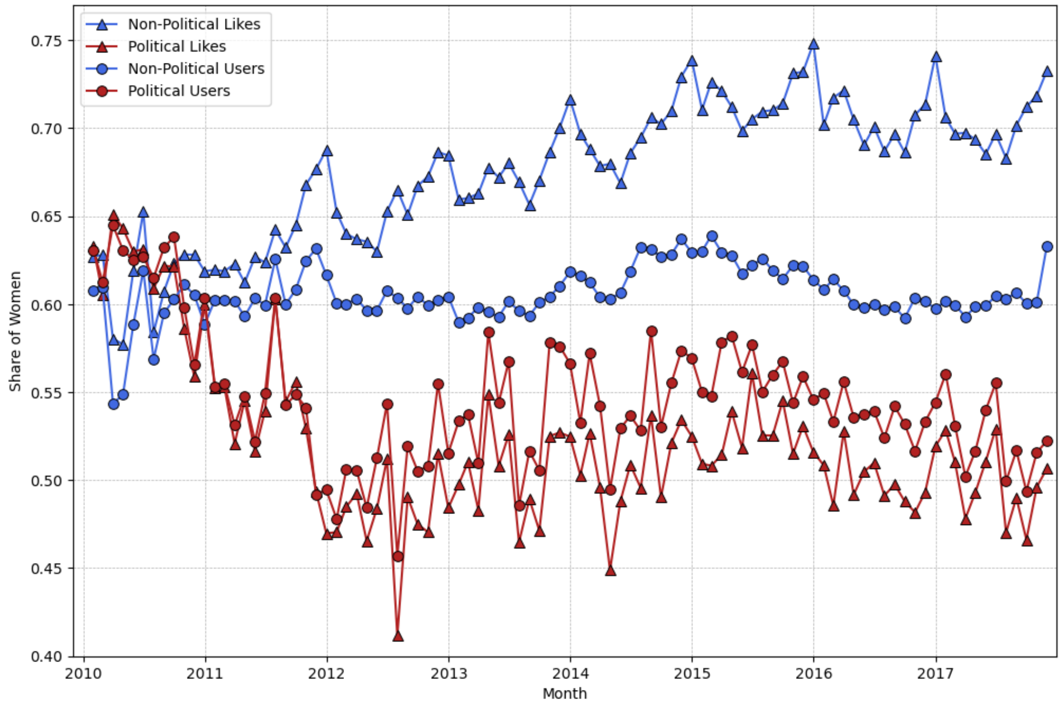


Figure 2.3: Gendered participation over time

of likes they contribute in the non-political realm, there has not been an equivalent trend in the political realm.

At a base level, our findings are thus quite different from Barberá and Rivero, 2015, who find that men tend to dominate political participation in terms of tweeting about politics. However, our results are in many ways in line with our theoretical expectations, that for political participation with low visibility, we see less gender inequality in participation (Bode, 2017). However, when considering the gender distribution of social media in general, it is clear that a lot fewer women choose to participate less in politics by liking than men.

2.6.2 Ethnic Minorities

Ethnic minorities are in general extremely well represented on non-political social media, with almost half of the active users being non-Danish. In the population, only 11.7% are ethnic minorities, so it is surprising that 47.87% of the users are non-Danish, as seen in Table 2.2. Looking at the likes on non-political Danish pages in general, we see that ethnic minorities are not as active as the Danish users but are still overrepresented, with a share of 32.41% of the likes compared to the population. Similarly, we can see that the average Danish active user has liked about 110 more non-political posts than the average non-Danish user.

In the political domain, ethnic minorities are also fairly well represented. The share of non-Danish users is close to double that of the population, with a positive difference of 18.98 percentage points more users

	Political Users	Non-Political Users
Share of users (Danish)	69.32%	52.13%
Share of users (non-Danish)	30.68%	47.87%
Share of likes (Danish)	70.41%	67.68%
Share of likes (non-Danish)	29.59%	32.32%
Avg. likes per user (Danish)	45.10 (pol. likes)	229.71 (non-pol. likes)
Avg. likes per user (non-Danish)	42.82 (pol. likes)	119.94 (non-pol. likes)

Table 2.2: **Representation of ethnic heritage on social media.** The distribution of people with Danish and non-Danish heritage in Denmark in 2015 was 88.3% Danish and 11.7% non-Danish (DST, 2024a)

than we would expect given the population. The same is true if we examine the share of political likes made by non-Danish users, where we see a similar difference. Additionally, we can see that the average number of political likes per user among the Danish and non-Danish politically active users is very similar, with the Danish users on average having liked roughly 2 more posts. This is a stark difference when comparing this to the non-politically active users where the Danish users on average are much more active than the non-Danish users. As such, ethnic minorities are both very active and well-represented among the politically active users on social media.

However, similarly to the case with women, we see that ethnic minorities are participating politically at a disproportionate rate compared to ethnic majority users, when considering their representation on non-political social media. In Table 2.2, we see that compared to participation on social media in general, ethnic minorities have a 17.19 percentage point smaller share of the politically active users. The difference is less stark in terms of likes, where the difference is only 2.82 percentage points, but the direction remains the same. As such, there are a lot fewer politically active users and likes from ethnic minorities than we would expect given their representation on non-political social media.

If we now turn to examine political participation and ethnicity on social media over time in Figure 2.4, we see some very interesting patterns. In the non-politically active set of users, we see that the share of ethnic minority users rises slowly until mid-2014, where the share rises quite dramatically, after which the share falls slowly again. The pattern is the same for the non-political likes.

Looking at the share of politically active users and likes coming from ethnic minorities, we see that they are very slowly rising over time. However, while Both the share of political likes and users rise over time, they are consistently lower than the share of ethnic minority users in the non-politically active set of users and their likes. As such, we see that while ethnic minorities are overrepresented on social media, including in politics, compared to the population, they are underrepresented in politics when considering their presence on social media in the first place. While we theoretically expected less of a difference gap between ethnic majorities and minorities in social media feedback, compared to traditional modes of participation, the fact that ethnic minorities are overrepresented in political participation on social media is very surprising. However, it is still clear that fewer ethnic minorities opt to participate in nano-level political participation than we would expect given the social media activity of ethnic minorities on non-political social media.

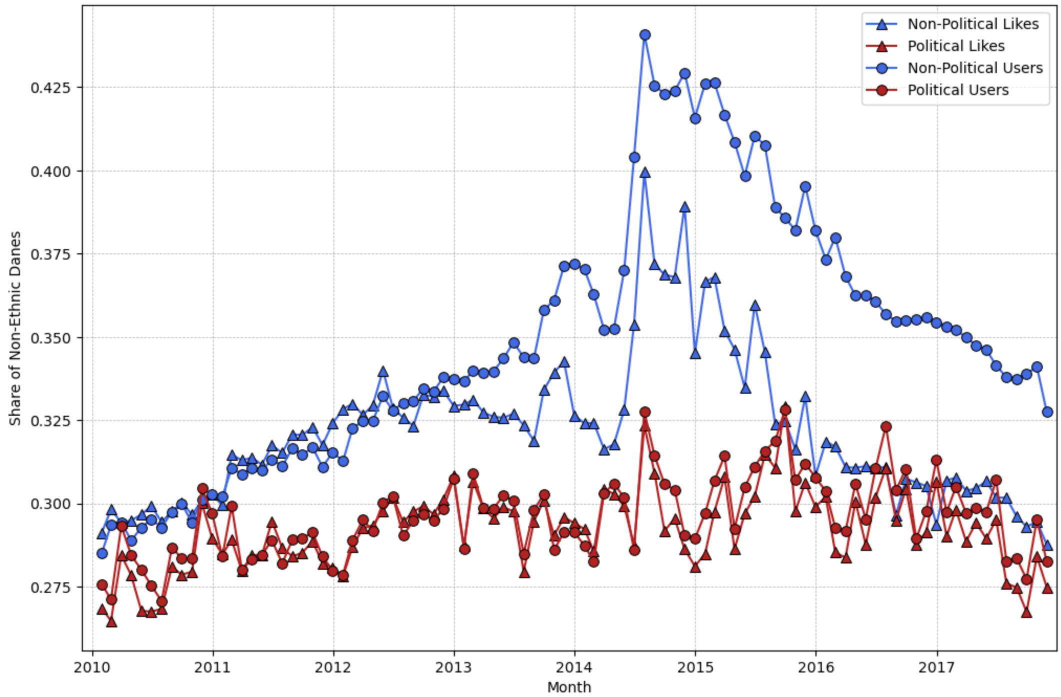


Figure 2.4: **Heritage and participation over time**

2.6.3 Cultural Preference Clusters

The politically active social media users tend to also have a preference for political and civic cultural content. In Figure 2.5, we can see that about 45% of the politically active users are mainly culturally interested in politics and civic content. About a quarter of the users mainly prefer entertainment content, and roughly 10% of the politically active users primarily interact with content related to the local community or shopping. Finally, about 5% have a preference for content related to cars, men, and nationalism, and a negligible share prefer content related to horses and other animals.

The non-politically active users mainly prefer social media content related to entertainment, but are otherwise similar to the politically active users. Around 45% of those users prefer entertainment content, and just above 20% of them mainly prefer political and civic content. Hereafter, the pattern is very similar to the politically active users, with local pages and shopping taking up a little more than 10% each, cars, men, and nationalism about 5%, and finally a negligible share of users primarily devoted to content related to horses and other animals.

If we compare the politically active users to the non-politically active users, the main difference is the share of users preferring political content and entertainment. Users generally interested in politics and civil society are much more likely to also be politically active, whereas users who prefer entertainment content are much less likely to be politically active. As such, it seems evident that even on social media there is a clear relation between political interest and political participation.

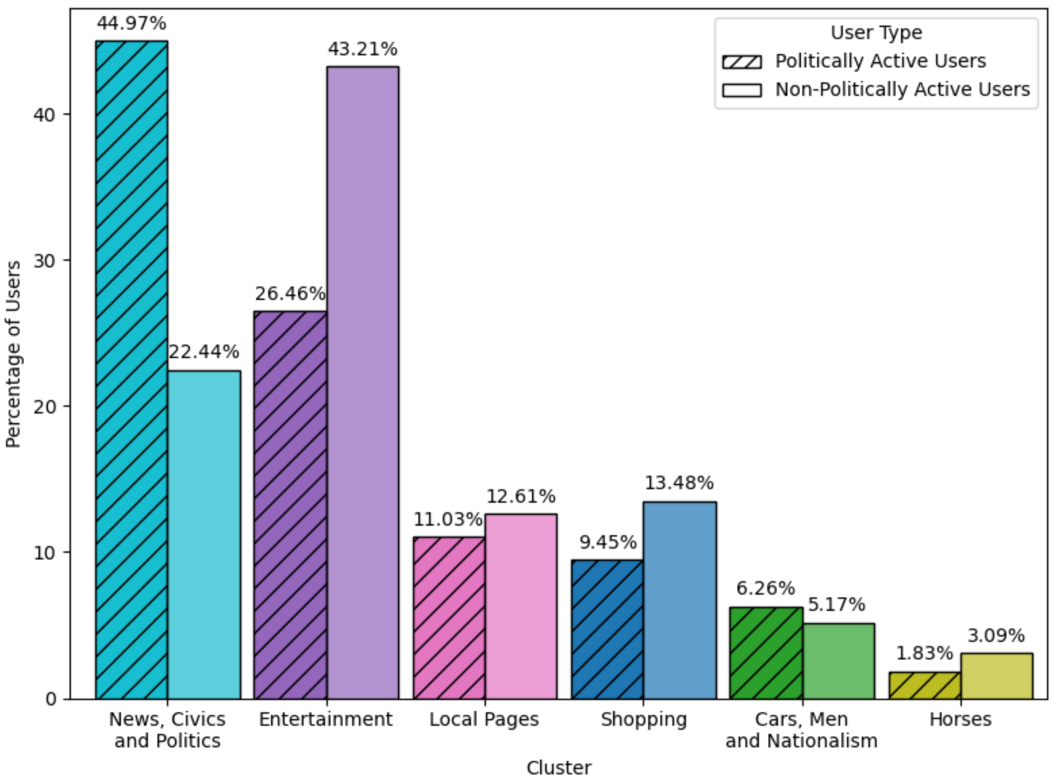


Figure 2.5: Distribution of Users in Cultural Preference Clusters. Each user is assigned the cluster where they have made the most likes.

However, we also see that only about half of the politically active users have a strong cultural preference for political and civic content. Similarly we also see that there are many users who prefer entertainment content, who still engage with politicians. We should also note that almost a quarter of the non-politically active users primarily use social media to engage with political and civic content but do not actually participate in politics by liking MPs' posts. This indicates that a primary interest in politics and civic content is neither sufficient nor necessary for political participation on social media. In other words, while a lot of the social media feedback to politicians comes from a relatively attentive or politically interested public, more than half comes from users who are primarily interested in other things than politics.

If we examine the distribution of users from each cluster over time in Figure 2.6, we see some very interesting patterns. For both the politically active and the non-politically active users, we see that the four smallest clusters have been remarkably stable in terms of the share of users over time. Examining the politically and entertainment-interested users, however, we see more interesting patterns. For the politically active users, we see that over time the share of users who come from the politically interested cluster has fallen from roughly 70% of the unique monthly users in 2010 to hovering around 50% from 2012 to 2018. Similarly, the share of politically active users from the entertainment cluster rises from 10% in the early years to hovering around 25%, and then finally trending downward in the later years.

Among the non-politically active users, we see that the users who prefer entertainment content rose from around 40% to around 50% during the early years of social media, before slowly declining to 30% from

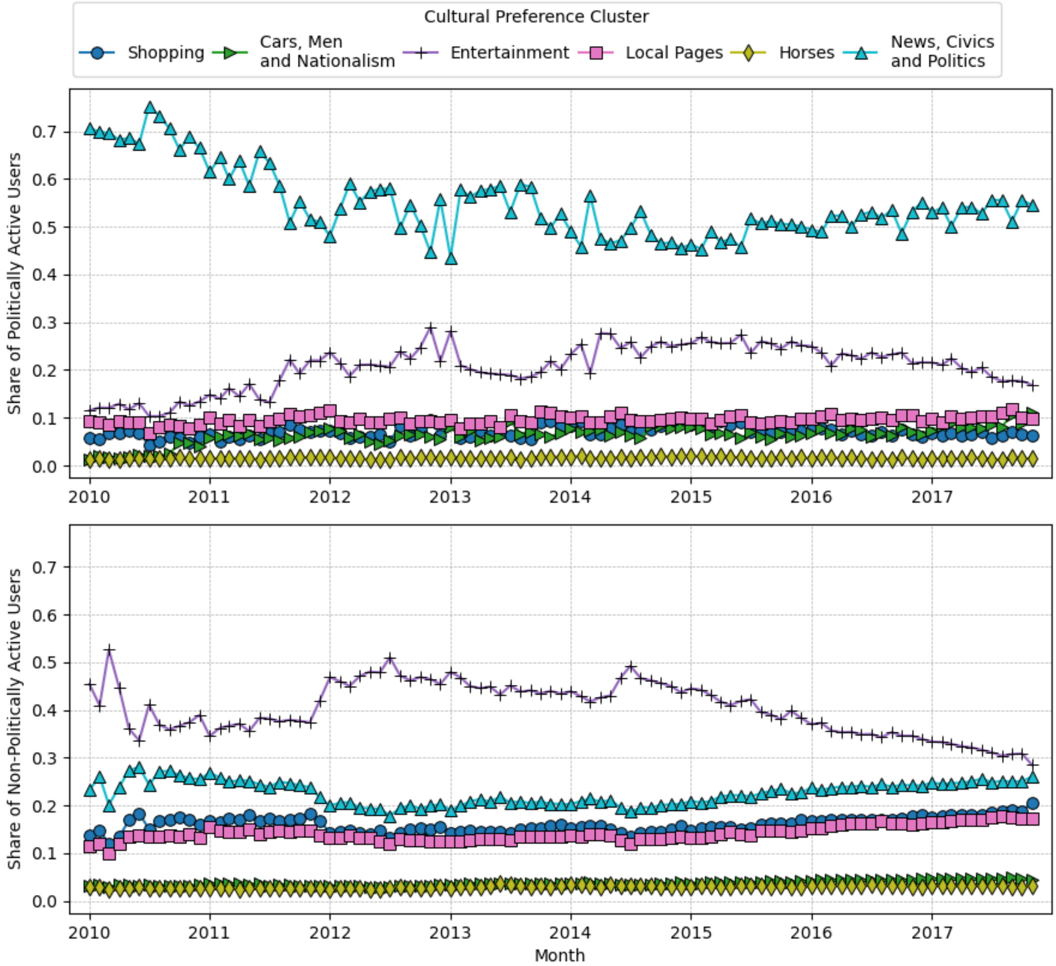


Figure 2.6: Cultural Preference Clusters Over Time

mid-2014 onward. At the same time, we see the share of politically interested non-politically active users trended slightly downward during the first few years, but from mid-2014 started to slowly rise towards 25% of the non-politically active users.

These patterns are very interesting as they seem to indicate that political participation on social media started out being more for the politically interested users but slowly became a domain where users with other cultural preferences felt comfortable. As such, social media seems to enable a fairly large population of non-politically interested people to participate in politics, and more so over time. However, without any direct comparisons to the general population, it is difficult to determine if social media feedback is better at this than other modes of participation.

2.6.4 Ideology, Ideological Extremity, and Partisanship

The politically active social media users are on average more left-leaning than the general public. In Table 2.3, we can see politically active social media users is on average 0.034 or 7% more left-leaning than the population on an ideological scale from 0-1. Looking at ideological extremity, we see that the average politically active social media user is 0.061 or 28% further from the middle of the ideology scale and thus more ideologically extreme. In other words, the politically active social media users are much more ideologically extreme than the population. Finally, we also see that on average the politically active social media users and the population are roughly equally partisan. The population is 0.004 more partisan than the politically active social media users on a 0-1 scale, a difference that is not statistically significant. As such, it seems that the politically active social media users are on average more left-leaning, more ideologically extreme, and similarly partisan to the population as a whole.

	Political Users	Population
Ideology	0.488	0.516
Extremism	0.284	0.220
Partisanship	0.606	0.609

Table 2.3: Ideological extremity on social media and in the population. The left-right scale is normalized to span 0-1, and the ideological extremity measure is the distance on this scale from the midpoint of 0.5. The left-right scale goes from left (0) to right (1) meaning a higher score indicates more right-leaning.

Comparing the distributions of ideology, extremity, and partisanship in Figure 2.7 reveals interesting patterns. In the left-most histogram, we see some stark differences between the distribution of ideology of the politically active social media users and the population. Here we see that while the population appears to be more or less normally distributed around the center of the scale, the politically active social media users resemble a bimodal distribution centered around 0.1 and 0.9, respectively.

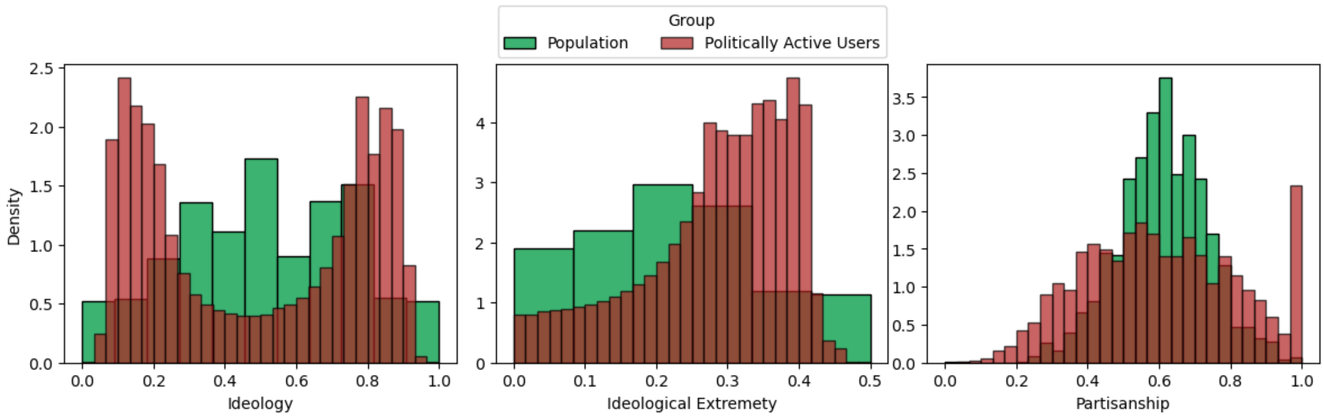


Figure 2.7: Ideology, Extremism, and Partisanship

As such, while the politically active social media users are fairly similar to the population in terms of average ideology score as seen in Table 2.3, there are very few politically active users around the average on the ideology scale. We also see this when examining the ideological extremity of the politically active social media users and the population in the central histogram. Here we see that while the most common extremity value in the population is 0.2, it is closer to 0.4 among the politically active social media users.

Examining the distribution of partisanship in the right-most histogram, we see that while the average partisan levels are similar between the population and the politically active social media users, there are interesting differences in the distributions. Specifically, the population appears to be roughly normally distributed around 0.6, while the politically active social media users are more spread out, with a very large group of highly partisan users who only like one party.

Turning to the relationship between the users' political characteristics and the frequency at which they mobilize their opinion, we see in Figure 2.8 that ideologically extreme users like politicians' posts more often. For ideology and partisanship, we see negligible correlations between like frequency and the scores, but with ideological extremity, we see a weak positive correlation. As such, we see that users who are more ideologically extreme also tend to be more frequent likers. In this way, the people who mobilize their opinion on social media are not just more ideologically extreme, the more ideologically extreme users are also more active than other politically active users.

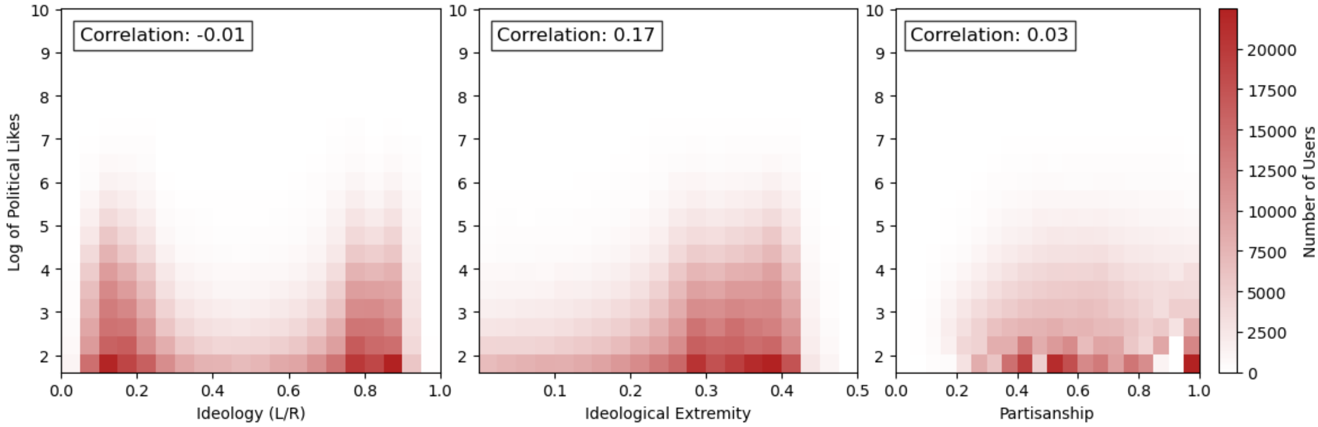


Figure 2.8: Like Frequency, Ideology, Extremism, and Partisanship

If we examine how these dynamics of the politically active users' ideology, extremity and partisanship change over time, we can see some interesting patterns. In Figure 2.9, we see dramatic changes in the average ideology of the politically active social media users over time. Specifically, we can see that in the first years, social media was dominated by the ideological left. However, over the first two years, the scales slowly balance and around 2012 the average ideology of the users and their likes is around the middle of the scale. However, there are large fluctuations from month to month.

Interestingly, we do not see the same behavior for the ideological extremity of the users and their likes. Here we instead see a very slow but stable decreasing trend over time, showing that on average the users liking political posts have become less and less ideologically extreme and partisan, though the change is small.

In sum, we can see that the politically active social media users are a little bit more left-leaning than the general public, but a lot more ideologically extreme, on both ends of the ideological scale. The politically active social media users are similarly partisan to the population, though there are more extremely partisan people on social media than in the public. We can see that the users are equally active on both ends of the political spectrum, but that there is a weak positive correlation between ideological extremity and the tendency to mobilize one's opinion on social media. Over time, the levels of partisanship and ideological

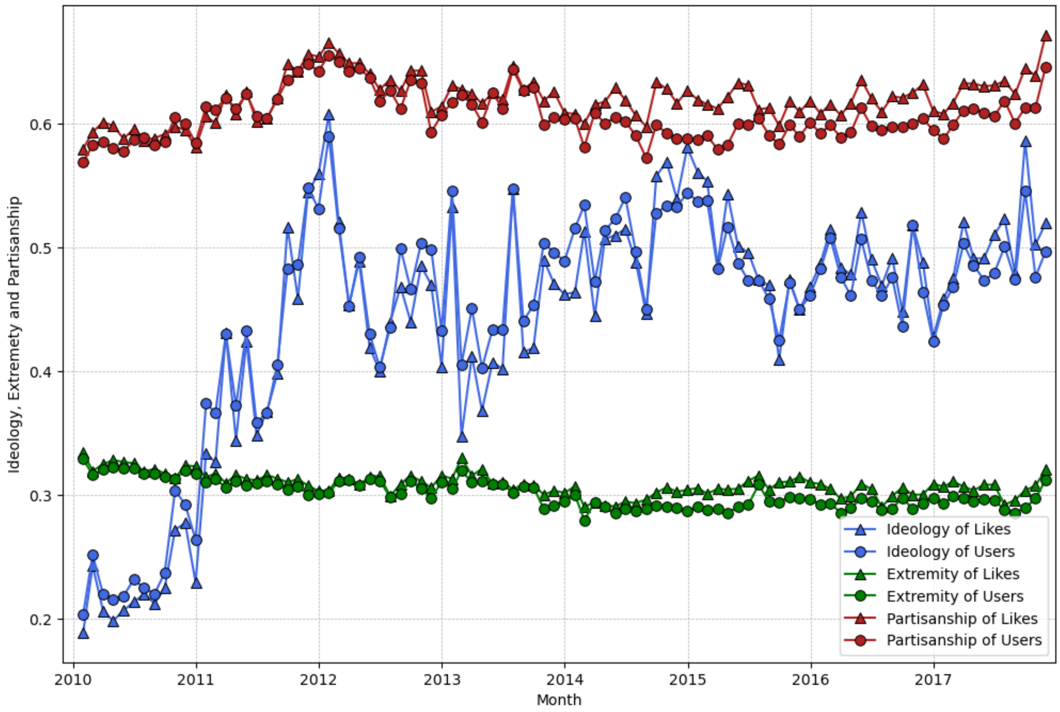


Figure 2.9: Ideology, Extremity, and Partisanship over time

extremity have been remarkably stable, while the average ideology of the politically active users on social media has fluctuated wildly over the years.

2.7 Discussion and Conclusions

In this article, we argue that social media feedback such as Facebook likes should be considered and analyzed as a legitimate form of political participation, and that in its aggregated form we should think of it as a form of mobilized public opinion. We show that social media feedback as a mode of political participation enables groups of people to participate who typically participate less in traditional modes of political participation such as protesting and letter-writing, and even other modes of participation on social media such as tweeting and commenting.

Our empirical results show that women and ethnic minorities make substantial use of social media feedback to participate in politics. However, there is still a strong tendency for minorities to opt out of political participation more frequently. Similarly, we see that less than half of the likes given to politicians come from people who are generally interested in politics. Taken together, this indicates that social media feedback can serve as a channel of political participation for groups of people who, for various individual and structural reasons, do not participate as much through other modes of participation.

Interestingly, the politically active users on social media are more left-leaning ideologically than the general population. This contrasts with prior research, which tends to show that users voicing their opinions on social media are more right-leaning (Barberá and Rivero, 2015; Conover *et al.*, 2012). It is unclear whether this is specific to Danish Facebook or whether examining likes rather than tweets accounts for the difference. One tentative speculation that warrants future examination is that if social media is indeed dominated by right-leaning users in general, then left-leaning users might face greater risk in terms of harassment when voicing their opinions publicly in a tweet or post and therefore might prefer to use the less-visible like-button more often.

The politically active users on social media are significantly more ideologically extreme but similarly partisan to the general population. Additionally, more ideologically extreme users also tend to be slightly more active. While the more extreme users was what we expected, we also anticipated more right-leaning and partisan users, which is not what we found. This suggests that while some dynamics of traditional political participation both on and off social media are also playing out in social media feedback, there are also key differences. Contrary to popular belief, social media has also not become an *increasingly* ideological and partisan space, at least when measured using social media feedback. Instead, it appears to be a fairly stable or even decreasingly ideologically extreme environment.

While we know that politicians are responsive to social media feedback such as likes (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023), examining the population who provide that feedback makes us question whether they *should* be. The fact that social media feedback is both disproportionately made by otherwise marginalized groups in political participation and by people who are ideologically extreme raises an interesting dilemma. Giving political representatives access to opinions from marginalized groups might be beneficial, but overemphasizing ideologically extreme voices could lead politicians toward more extreme ideological positions that do not reflect the general public. This raises numerous questions about how politicians and researchers should work with and interpret social media feedback, questions which should be the focus of future research.

This study presents one of the most comprehensive analyses to date of who chooses to use the like button to mobilize their opinion toward political elites. However, there are a few key limitations to address. The first key limitation is the generalizability of the results in both time and space. While we analyze an extremely large and comprehensive dataset, we are still only examining one national context, Denmark. Denmark has a multiparty system and typically low political polarization (Dinesen *et al.*, 2020), and as such it is unclear how the results relate to more commonly studied countries like the US. Similarly, we study a long period, but our data ends at the start of 2018, and it is unclear how the results hold up in the current day and age.

Another important caveat to our comparison of the general public with the social media users is that the social media users are a nested sample from the general public. Therefore, when we compare the social media users with the general public, we are essentially comparing the social media users to a combination of these social media users and all other individuals in the general population, rather than to a distinct, separate population. This does not necessarily pose an issue, but it requires us to bear in mind that we are not making a direct comparison between the social media users and an offline public. Instead, we are comparing them to the broader societal landscape.

Finally, we acknowledge that our work is purely descriptive, and more research is needed to causally determine what makes likes a more appealing option for some groups of people. Future research should also investigate how these dynamics might differ in other national contexts and in more recent years. Additionally, exploring the potential policy implications of these findings, especially regarding how political representatives might balance the input from social media feedback with other forms of public opinion, could be a valuable avenue for future study.

Overall, our study highlights the complex interplay between social media feedback and political participation, suggesting that while social media can democratize political engagement by including marginalized voices, it also has the potential to amplify more extreme viewpoints. This duality underscores the need for a nuanced approach to understanding and leveraging social media feedback in the political arena.

Bibliography

Analyse og Tal, Consultancy (2024). *Angreb i den offentlige debat*. Accessed: 2024-07-10.

Asano, Taka-aki (2022). „Ideological Extremism and Political Participation in Japan". In: *Social Science Japan Journal* 25.1, pp. 125–140.

Barbera, Pablo (2015). „Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". eng. In: *Political analysis* 23.1, pp. 76–91.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker (2019). „Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data". In: *American Political Science Review* 113.4, pp. 883–901.

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau (2015). „Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" In: *Psychological Science* 26, pp. 1531–1542.

Barberá, Pablo and Gonzalo Rivero (2015). „Understanding the Political Representativeness of Twitter Users". In: *Social Science Computer Review* 33.6, pp. 712–729.

Bennett, W. Lance and Alexandra Segerberg (2012). „The Logic of Connective Action: The Personalization of Contentious Politics". In: *Information, Communication & Society* 15.5, pp. 739–776.

Blank, Grant (2017). „The Digital Divide Among Twitter Users and Its Implications for Social Research". In: *Social Science Computer Review* 35.6, pp. 679–697.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (Oct. 2008). „Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Bode, Leticia (2017). „Closing the gap: gender parity in political engagement on social media". In: *Information, Communication & Society* 20.4, pp. 587–603.

Bourdieu, Pierre (1984). *Distinction: A Social Critique of the Judgement of Taste*. Original work published 1979. Cambridge, MA: Harvard University Press.

Campbell, A., Converse P., Miller W., and Stokes D. (1960). *The American voter*.

Carlsen, Hjalmar Bang and Jonas Toubøl (2024). „Beyond Cheap and Biased: Informal Volunteering on Social Media during the COVID-19 Crisis". In: *OSF*.

Cho, Wendy K Tam (1999). „Naturalization, socialization, participation: Immigrants and (non-) voting". In: *The Journal of Politics* 61.4, pp. 1140–1155.

Conover, Michael D, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer (2012). „Partisan asymmetries in online political activity". In: *EPJ Data Science* 1.6, pp. 1–19.

Coscia, Michele and Frank Neffke (2017). *Network Backboning with Noisy Data*. arXiv: 1701.07336 [physics.soc-ph].

Dalton, Russell (2002). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Third. Chatham House Publishers.

Dancygier, Rafaela M., Karl-Oskar Lindgren, Sven Oskarsson, and Kåre Vernby (2015). „Why Are Immigrants Underrepresented in Politics? Evidence from Sweden". In: *American Political Science Review* 109.4, pp. 703–724.

De Rooij, Eline A (2012). „Patterns of immigrant political participation: Explaining differences in types of political participation between immigrants and the majority population in Western Europe". In: *European sociological review* 28.4, pp. 455–481.

De Vreese, Claes H and Hajo Boomgaarden (2006). „News, political knowledge and participation: The differential effects of news media exposure on political knowledge and participation". In: *Acta politica* 41, pp. 317–341.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs.CL].

Dinesen, Peter Thisted, Rune Slothuus, and Rune Stubager (July 2020). „Danish Public Opinion: Stability, Change and Polarization". In: *The Oxford Handbook of Danish Politics*. Ed. by Peter Munk Christiansen, Jørgen Elklit, and Peter Nedergaard. Oxford: Oxford University Press, pp. 259–277.

DST (2023). *Danmark bruger sociale medier mest i EU*. https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=46771. Accessed: 2024-07-08.

DST (2024a). *Immigrants and Descendants*. https://www.dst.dk/da/Statistik/emner/borgere/befolkning/indvandrere-og-efterkommere. Accessed: 2024-07-08.

DST (2024b). *Population Statistics*. https://www.dst.dk/en/Statistik/emner/borgere/befolkning. Accessed: 2024-07-08.

Earl, Jennifer (2016). „Protest Online: Theorizing the Consequences of Online Engagement". In: *The Consequences of Social Movements*. Ed. by Lorenzo Bozi, Marco Giugni, and Katrine Uba. Cambridge: Cambridge University Press, pp. 363–400.

Earl, Jennifer and Katrina Kimport (2011). *Digitally enabled social change: Activism in the internet age*. Mit Press.

Elliott, Thomas and Jennifer Earl (2018). „Online protest participation and the digital divide: Modeling the effect of the digital divide on online petition-signing". In: *New Media & Society* 20.2, pp. 698–719.

Enikolopov, Ruben *et al.* (2020). „Social Media and Protest Participation: Evidence from Russia". In: *Econometrica* 88.4. Accessed 21 Apr. 2024, pp. 1479–1514.

Ennser-Jedenastik, Laurenz, Christina Gahn, Anita Bodlos, and Martin Haselmayer (2022). „Does social media enhance party responsiveness? How user engagement shapes parties issue attention on Facebook". In: *Party Politics* 28.3, pp. 468–481. eprint: https://doi.org/10.1177/1354068820985334.

Familieretshuset (2024). *Godkendte Fornavne*.

Fonseca, António, Catarina Pontes, Sérgio Moro, Fernando Batista, Ricardo Ribeiro, Rita Guerra, Paula Carvalho, Catarina Marques, and Cláudia Silva (2024). „Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns". In: *Heliyon* 10.11, e32246.

Gil de Zúñiga, Homero, Aaron S Veenstra, Emily K Vraga, and Dhavan V Shah (2014). „Digital democracy: Reimagining pathways to political participation". In: *Journal of Information Technology & Politics* 7.1, pp. 36–51.

Gomez-Gonzalez, Carlos, Cornel Nesseler, and Helmut M Dietl (2021). „Mapping discrimination in Europe through a field experiment in amateur sport". In: *Humanities and Social Sciences Communications* 8.1, pp. 1–8.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, *et al.* (2011). „The Dynamics of Protest Recruitment through an Online Network". In: *Scientific Reports* 1, p. 197.

Greenacre, M. J. (2017). *Correspondence analysis in practice*. Academic Press.

Halupka, Max (2014). „Clicktivism: A Systematic Heuristic". In: *Policy & Internet* 6, pp. 115–132.

Halupka, Max (2018). „The legitimisation of clicktivism". In: *Australian Journal of Political Science* 53.1, pp. 130–141.

Hansen, Kasper Møller and Rune Stubager (2017). *2015-valget: Oprør fra udkanten. Folketingsvalget 2015*. København: Jurist- og Økonomforbundets Forlag.

Hargittai, Eszter (2020). „Potential biases in big data: Omitted voices on social media". In: *Social science computer review* 38.1, pp. 10–24.

Jöst, Prisca, Matthias Krönke, Sarah J. Lockwood, and Ellen Lust (2024). „Drivers of Political Participation: The Role of Partisanship, Identity, and Incentives in Mobilizing Zambian Citizens". In: *Comparative Political Studies* 57.9, pp. 1441–1474.

Karpowitz, Christopher F and Tali Mendelberg (2014). *The silent sex: Gender, deliberation, and institutions*. Princeton University Press.

Kim, Mihee (2016). „Facebook's Spiral of Silence and Participation: The Role of Political Expression on Facebook and Partisan Strength in Political Participation". In: *Cyberpsychology, Behavior, and Social Networking* 19.12. PMID: 27849357, pp. 696–702.

Klanja, Marko *et al.* (2018). „Measuring Public Opinion with Social Media Data". In: *The Oxford Handbook of Polling and Survey Methods*. Ed. by Lonna Rae Atkeson and R. Michael Alvarez. Online edition published in 2015. Oxford: Oxford University Press.

Kostelka, Filip, André Blais, and Elisabeth Gidengil (2019). „Has the gender gap in voter turnout really disappeared?" In: *West European Politics* 42.3, pp. 437–463.

Kostelka, Filip and Jan Rovny (2019). „Its Not the Left: Ideology and Protest Participation in Old and New Democracies". In: *Comparative Political Studies* 52.11, pp. 1677–1712.

Larson, Jennifer M. *et al.* (2019). „Social Networks and Protest Participation: Evidence from 130 Million Twitter Users". In: *American Journal of Political Science* 63.3. Accessed 21 Apr. 2024, pp. 690–705.

Lilleker, Darren, Karolina Koc-Michalska, and Bruce Bimber (2023). „Women learn while men talk?: revisiting gender differences in political engagement in online environments". In: *Women in the Digital World*. Routledge, pp. 46–62.

Lonkila, Markku and Pertti Jokivuori (2023). „Sharing and liking as youth nano-level participation. Finnish students civic and political engagement in social media". In: *Journal of Youth Studies* 26.6, pp. 803–820. eprint: https://doi.org/10.1080/13676261.2022.2049731.

Lutz, Christoph, Christian Pieter Hoffmann, and Miriam Meckel (2014). „Beyond just politics: A systematic literature review of online participation". In: *First Monday* 19.7.

McAdam, Doug (1982). *Political Process and the Development of Black Insurgency 1930-1970*. University of Chicago Press.

McAdam, Doug (1992). „Gender as a Mediator of the Activist Experience: The Case of Freedom Summer". In: *American Journal of Sociology* 97.5, pp. 1211–1240.

McGregor, Shannon C (2019). „Social media as public opinion: How journalists use social media to represent public opinion". In: *Journalism* 20.8, pp. 1070–1086.

Meer, Tom W. G. van der, Jan W. van Deth, and Peer L. H. Scheepers (2009). „The Politicized Participant: Ideology and Political Action in 20 Democracies". In: *Comparative Political Studies* 42.11, pp. 1379–1403.

Morozov, Evgeny (2009). „The Brave New World of Slacktivism". In: *Foreign Policy*.

Morselli, Davide, Maite Beramendi, Zakaria Bendali, and Olivier Fillieule (2023). „A Longitudinal Approach to Online Collective Identity Work: The Case of the Gilets Jaunes in the Var Department". In: *Mobilization: An International Quarterly* 28.3, pp. 301–322.

Paxton, Pamela, Sheri Kunovich, and Melanie M Hughes (2007). „Gender in politics". In: *Annu. Rev. Sociol.* 33.1, pp. 263–284.

Prior, Markus (2005). „News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout". In: *American Journal of Political Science* 49.3, pp. 577–592.

Rau, Eli G. (2021). „Partisanship as Cause, Not Consequence, of Participation". In: *Comparative Political Studies* 55.6, pp. 1037–1069.

Al-Rawi, Ahmed (2022). „Hashtagged Trolling and Emojified Hate against Muslims on Social Media". In: *Religions* 13.6.

Rehurek, Radim and Petr Sojka (May 2010). „Software Framework for Topic Modelling with Large Corpora". English. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. http://is.muni.cz/publication/884893/en. Valletta, Malta: ELRA, pp. 45–50.

Rogowski, Jon C. (2014). „Electoral Choice, Ideological Conflict, and Political Participation". In: *American Journal of Political Science* 58.2, pp. 479–494.

Schöll, Nikolas, Aina Gallego, and Gaël Le Mens (2023). „How Politicians Learn from Citizens Feedback: The Case of Gender on Twitter". In: *American Journal of Political Science* n/a.n/a, pp. 557–574. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12772.

Shahin, S., J. Nakahara, and M. Sánchez (2024). „Black Lives Matter goes global: Connective action meets cultural hybridity in Brazil, India, and Japan". In: *New Media & Society* 26.1, pp. 216–235.

Strandberg, Kim (2013). „A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections". In: *New Media & Society* 15.8, pp. 1329–1347.

Suk, J., A. Abhishek, Y. Zhang, S. Y. Ahn, T. Correa, C. Garlough, and D. V. Shah (2021). „#MeToo, Networked Acknowledgment, and Connective Action: How Empowerment Through Empathy Launched a Social Movement". In: *Social Science Computer Review* 39.2, pp. 276–294.

Theocharis, Yannis (2015). „The Conceptualization of Digitally Networked Participation“. In: *Social Media + Society* 1.2.

Verba, Sidney, Nancy Burns, and Kay Lehman Schlozman (1997). „Knowing and caring about politics: Gender and political engagement“. In: *The journal of politics* 59.4, pp. 1051–1072.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady (1995). *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

Vesnic-Alujevic, Lucia (2012). „Political participation and web 2.0 in Europe: A case study of Facebook“. In: *Public Relations Review* 38.3, pp. 466–470.

Vochocova, Lenka, Vaclav Stetka, and Jaromir Mazak (2016). „Good girls dont comment on politics? Gendered character of online political participation in the Czech Republic“. In: *Information, Communication & Society* 19.10, pp. 1321–1339.

Whitford, Andrew B., Jeff Yates, and Holona L. Ochs (2006). „Ideological Extremism and Public Participation“. In: *Social Science Quarterly* 87.1, pp. 36–54.

Wiltfang, Gregory L and Doug McAdam (1991). „The costs and risks of social activism: A study of sanctuary movement activism“. In: *Social Forces* 69.4, pp. 987–1010.

Wojcieszak, Magdalena, Andreu Casas, Xiao Yu, Jonathan Nagler, and Joshua A. Tucker (Sept. 2022). „Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity“. In: *Science Advances* 8.39, eabn9418.

3 The Like Effect - Political Responsiveness on Social Media

By August Lohse, Tobias Priesholm Gårdhus, Snorre Ralund, and Hjalmar Bang Carlsen

Abstract

This study explores the influence of social media feedback, such as Facebook "likes", on political responsiveness. Social media platforms provide politicians with detailed, targeted information about public preferences, facilitating direct engagement. Recent research indicates that this feedback can play a key role in political responsiveness. We examine whether these effects are generalizable across different feedback types, topics, political periods, and types of politicians. Our analysis builds on a dataset of all political posts and user interactions on Facebook from Danish MPs active from 2011 to 2018. Our results demonstrate a high level of responsiveness to social media feedback across all feedback types, topics, periods, and types of politicians. Although responsiveness varies among different periods and politician types, an overall pattern of responsiveness remains. This research provides valuable insights into the complexities of political responsiveness in the digital age, emphasizing the critical role of social media feedback in contemporary politics.

3.1 Introduction

At the core of political science lies a fundamental question: how responsive are politicians to the will of the people? This question is crucial for understanding democratic governance and has inspired a wealth of research dedicated to exploring the alignment between public preferences and political actions (e.g., Stimson *et al.*, 1995; Page and Shapiro, 1983; Jones and Baumgartner, 2004). Findings from these studies consistently indicate a notable congruence between the policy priorities of the electorate and the decisions made by their representatives.

For politicians to be responsive, they must have access to accurate and timely information about public opinion. Traditional channels such as opinion polls (Page and Shapiro, 1983) and insights from political elites (Stimson *et al.*, 1995) have served this purpose. However, the rise of social media presents a new frontier. Recent research by Barberá *et al.* (2019) highlights that social media platforms, through the content generated by both the public and the media, offer a novel and direct channel for gauging public opinion.

Social media also offers a much more direct form of information on public opinion. Social media allows politicians to interact with the public directly and receive immediate feedback through "likes", "shares", "comments", etc. providing them with a direct gauge of public sentiment towards their specific messages. Emerging evidence suggests that politicians and political parties are responsive to this type of feedback, with studies showing a link between high engagement on social media posts and the likelihood of those topics being addressed in future posts (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023). For instance, Schöll *et al.* (2023) found that politicians are responsive to positive feedback related to gender issues, while Ennser-Jedenastik *et al.* (2022) observed that party organizations are responsive to social media feedback during election campaigns. These findings hint at a broader connection between social media interactions and political responsiveness, yet they both focus on responsiveness in rather specific conditions, such as the highly politicized topic of gender and during an election. This raises the question of how responsive politicians are to social media feedback *in general*.

Our research aims to fill this gap by examining how responsiveness to social media feedback varies across different types of feedback, topics, time periods, and the characteristics of the politicians themselves. We explore how factors such as ideological position, extremity, and electoral incentives influence politicians' engagement with social media feedback across various policy areas and stages of the electoral cycle. With social media increasingly shaping political discourse, comprehending the nuances of politicians' responsiveness to social media feedback is vital. Since social media feedback often originates from a polarized segment of users (Wojcieszak *et al.*, 2022), it is crucial to determine the extent to which politicians consider this feedback to understand the broader impact of social media on democratic governance.

In the following sections, we situate our work within the existing literature on political responsiveness and explore the potential impact of social media as a mechanism for this responsiveness. We formulate several hypotheses based on prior research, guiding our subsequent analysis. We then outline our methodology for assessing politicians' responsiveness to social media feedback and describe our comprehensive dataset of social media posts and how we label them using deep learning models. Finally, we present our analysis and discuss the implications and generalizability of our findings, offering new insights into the evolving relationship between social media and political responsiveness.

3.2 Political Responsiveness

Research into political responsiveness has long been a central focus in political science. In essence, the aim is to understand how elected representatives address the needs and concerns of the public and how the public in turn can hold politicians accountable for the way they represent them. Within the study of political responsiveness, there have traditionally been two groups of studies; those dealing with responsiveness in terms of policy *preference* (e.g. Page and Shapiro, 1983; Soroka and Wlezien, 2009; Stimson *et al.*, 1995), and those dealing with policy *priorities* (e.g. Jones and Baumgartner, 2004; Edwards and Wood, 1999; Sulkin, 2005; Neundorf and Adams, 2018; Barberá *et al.*, 2019; Schöll *et al.*, 2023).

Studies of policy preference responsiveness examine the degree to which politicians and constituents have the same opinions on policy, such as being pro-guns or supporting universal health care. Responsiveness in terms of policy priorities instead asks the question of what policies or issues the politicans and their constituents value highly (Jones and Baumgartner, 2004). In both, scholars aim to measure the degree of congruence between elected officials and the public, using any number of measures of policy position and priority such as roll-calls, voting behavior, ideological positions, or attention devoted to a policy area (Caughey and Warshaw, 2018).

Responsiveness in the political system can arise via two channels, either through politicians being elected based on the congruence with voters or dynamically by politicians adapting to the voters while in office (Caughey and Warshaw, 2018). Similarly, the alignment of politicians and the public can arise either as a top-down process where the issues and solutions brought forward by politicians influence the constituents' opinions of what is important or as a bottom-up process where the relationship is reversed so that the constituents' issue attention shapes the political landscape (Wagner and Meyer, 2014; De Vries and Marks, 2012; Klüver and Sagarzazu, 2016; Green-Pedersen, 2019). Here we focus on the way politicians adapt their policy priorities to those of the public while in office. As such we focus on what we could call the dynamic bottom-up policy priority responsiveness on social media, examining the temporal congruence of how politicians and the public prioritize various policy areas.

3.2.1 Social Media Feedback as a Mechanism for Political Responsiveness

For politicians to be responsive to the public, they need information about the public's policy priorities. Traditionally, scholars have proposed that politicians can infer a rough consensus view of public opinion through a mix of information from political actors such as community leaders, the media, and other politicians (Stimson *et al.*, 1995), or that politicians can judge the "temperature" of the public opinion in terms of more or less policy on an area (Wlezien, 1995). Other research has pointed out the role of opinion pools (Page and Shapiro, 1983), domain experts (Lee, 2022), NGOs (De Bruycker and Rasmussen, 2021) or even direct communication between the public and politicians such as letter-writing (Schilozman *et al.*, 2012) as key sources of information enabling political responsiveness.

Recent work suggests that social media can also serve as an important channel of responsiveness in two key ways. The first is that politicians can gather information on the priorities of the public by examining what they write about on social media (Barberá *et al.*, 2019). Specifically, Barberá *et al.*, 2019 shows that politicians follow what the public and the media write about on social media, posting on the same topics afterwards (Barberá *et al.*, 2019). However social media also enables direct interactions between politicians such as direct messaging (Chen *et al.*, 2018) and other affordances such as "liking" or "retweeting" politicians' posts to show support for the message. Social media feedback is arguably uniquely suited as an information channel for politicians to adapt to the public, as it is abundant, available in real-time, and natively quantified into comparable metrics (Schöll *et al.*, 2023). As such politicians can gauge public opinion in extremely granular detail, faster, and more easily by relying on social media feedback, compared to more traditional channels such as opinion polls (Schöll *et al.*, 2023).

Two recent studies have examined the role of such feedback in political responsiveness and shown that politicians tend to post more about policy areas where they have previously received more feedback (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023). Specifically, Schöll *et al.*, 2023 examines Spanish politicians' political responsiveness on the issue of gender to social media feedback over a 3-year period, finding that there is an increased likelihood of posting about gender issues based on prior feedback (Schöll *et al.*, 2023). Ennser-Jedenastik *et al.*, 2022 on the other hand examine 35 different policy issues during the 2017 Austrian election period and show that parties tend to be responsive to social media feedback (Ennser-Jedenastik *et al.*, 2022). While these studies advance our knowledge about the role of social media feedback in political responsiveness considerably, there are also some limitations to the generalizability of the results. Both studies examine rather special cases, as the issue of gender, for example, went from a niche issue to a major political topic in the period under study in Schöll *et al.*, 2023 and Ennser-Jedenastik *et al.*, 2022 focus solely on an election period (Schöll *et al.*, 2023; Ennser-Jedenastik *et al.*, 2022). As such we still lack knowledge of how responsive politicians are to social media feedback *in general*. This includes how they respond to various kinds of social media feedback on more mundane political issues both in and out of election periods as well as how this depends on the characteristics of the politicians. Here we therefore develop a number of hypotheses stating how we expect responsiveness to differ across these dimensions. We begin formulating our main and most general hypothesis concerning politicians' responsiveness to social media feedback across topics:

H1: Politicians are more likely to post on a topic if they received more social media feedback on that topic in the past.

The first heterogeneity we examine is the type of social media feedback, as different forms of feedback might influence politicians in various ways. To explore this, we analyze politicians' responsiveness to different kinds of feedback. Specifically, we consider likes, which represent a low-effort form of feedback from social media users, as well as shares and comments, which require more effort and carry potential social repercussions or costs associated with publicly expressing political sentiments. Politicians might be more responsive to high-effort feedback, such as comments and shares, compared to low-effort feedback like likes, as high-effort feedback could be a stronger signal of support. However, it's important to note that comments and, to a certain extent, shares are not always as uniformly positive in the same manner likes are. Comments can also be negative or hostile, and shares can be used to mock or criticize the author. Given these nuances and considering that the latest study in the area finds no significant difference in the impact of different feedback types on Twitter (Schöll *et al.*, 2023), we do not anticipate any difference in politicians' responsiveness based on the type of feedback received. Thus, our hypothesis is formulated as follows:

H2: Politicians are equally responsive to different kinds of social media feedback.

To generalize the responsiveness of politicians across different significant periods in politics, we examine responsiveness during election periods and while politicians are in government. Recent studies have focused either solely on election periods (e.g. Ennser-Jedenastik *et al.*, 2022) or do not examine it explicitly (e.g. Barberá *et al.*, 2019; Schöll *et al.*, 2023). While elections might intuitively seem like a time when politicians would be more responsive to potential voter feedback, they often plan their preferred issue agendas well in advance of election periods (Ennser-Jedenastik *et al.*, 2022). Given this context, it is reasonable to expect that politicians adhere more closely to a predetermined plan during elections and are less responsive to voter feedback. Although there is evidence that they are still somewhat responsive during elections (Ennser-Jedenastik *et al.*, 2022), we anticipate that their responsiveness is greater between election periods. Therefore, we examine:

H3: During elections, politicians are less responsive to social media feedback.

Past research on responsiveness has largely focused on two ways in which being in government can affect responsiveness; one focusing on policy responsibility and another on resources. Research highlighting the role of policy responsibility in responsiveness has argued and shown that government parties are more constrained in their issue priorities due to their focus on policy making and their policy responsibility, which in turn makes them unable to compromise their policy priorities to respond to shifts in voter attention (Klüver and Spoon, 2016; Belchior *et al.*, 2024). Similarly, politicians in government are expected to address and take ownership of a wide range of issues, which forces them to sometimes address unpopular issues, especially when confronted with the press (Norton, 1993, pp. 35–37). Importantly, this research operates on a party level using manifesto changes as measures of responsiveness. We investigate the individual-level politicians' day-to-day communication on Facebook, where a different dynamic might be in play. Addressing an issue in this context is arguably less of an ideological signal than when parties do so in political manifestos. However, this theory would still predict a decrease in responsiveness when politicians go into government.

The other theoretical approach highlights the role of resources available to the politicians as a consequence of their role in government. Theories on agenda setting hold that parties in government have the financial resources to address issues and the administrative resources to devise policies on a large number of issues (Baumgartner and Jones, 2009, pp. 48–58). In a social media context, Ennser-Jedenastik *et al.*, 2022 have found that smaller parties are less responsive than larger parties, and argue that it is in part because of the resources of big parties to monitor social media feedback and devise communications on a variety of topics (Ennser-Jedenastik *et al.*, 2022). As such, prior research thus predicts both lower and higher responsiveness as a consequence of being in government. We therefore create two competing hypotheses:

H4a: During government, politicians are less responsive to social media feedback.
H4b: During government, politicians are more responsive to social media feedback.

In terms of the responsiveness of different politicians, we focus on the heterogeneous effects based on the gender of the politicians, their electoral incentives, their ideological leaning, and ideological extremity. Prior research finds that women are more responsive to social media feedback, at least when writing about gender issues (Schöll *et al.*, 2023). However, the authors do not claim that the gender differences generalize across topics, and we find no reason to have a prior expectation other than the gender effect is domain-specific to gender issues. As such we would expect men and women to be equally responsive to social media feedback across most topics.

H5: Women and men are equally responsive to social media feedback.

Regarding electoral incentives, one might expect a relationship between a politician's electoral success and their responsiveness to the public. Specifically, politicians seeking re-election, especially those who received a low number of votes in the previous election, might be more motivated to engage with and respond to the public Stimson *et al.*, 1995. Some studies also suggest that electorally marginal and insecure politicians are more opportunistic in their political communication (Grimmer, 2013). To secure their place in parliament, these marginal politicians tend to prioritize public opinion over special interests, party strategies, and other influences. As such we might expect that politicians who receive few votes are more strongly incentivised to be responsive to the public.

Conversely, it is well established that smaller parties, in terms of vote counts, are generally less responsive to the public (Ennser-Jedenastik *et al.*, 2022; Tarik Abou-Chadi and Mortensen, 2020). The mechanism between responsiveness and party size, as discussed in Ennser-Jedenastik *et al.*, 2022 and Tarik Abou-Chadi and Mortensen, 2020, is attributed to the resources available for monitoring and creating talking points on numerous topics. While there to our knowledge, are no studies examining this at the individual politician level, the mechanism of resource limitation may be even pronounced at the individual level, where political aides are even more scarce. Consequently, marginal candidates may tend to specialize in a few topics and be more hesitant to follow public opinion and discuss new issues. Therefore, we propose competing hypotheses stating that electoral incentives may either positively or negatively affect a politician's responsiveness:

H6a: Politicians who receive fewer votes are more responsive to social media feedback.
H6b: Politicians who receive more votes are more responsive to social media feedback.

We explore the impact of ideological position and extremism on politicians' responsiveness. Both popular belief and recent results hold that right-wing politicians tend to be more populist and thus responsive to the public (Pedrazzani and Segatti, 2022). Specifically Pedrazzani and Segatti, 2022 shows that right-wing politicians exhibit stronger ideological alignment with their voters, meaning they more closely mirror their voters in terms of ideology. While ideological alignment and responsiveness are distinct concepts the findings of Pedrazzani and Segatti, 2022 lead us to expect that right-wing politicians are more likely to respond to social media feedback.

Additionally, recent research also show that ideological extremism is associated with weaker alignment between voters and politicians and reduced responsiveness to public opinion Pedrazzani and Segatti, 2022; Bischof and Wagner, 2020. Consequently, we propose two hypotheses stating that right-leaning politicians are more responsive to social media feedback, whereas ideologically extreme politicians are less:

H7: Ideologically right-leaning politicians are more responsive to social media feedback.
H8: Ideologically extreme politicians are less responsive to social media feedback.

With our theoretical expectations established for the varying effects of social media feedback on responsiveness considering the heterogeneities of topic, feedback type, timing, and political ideology, we now move on to discuss our method for measuring responsiveness to social media feedback.

3.3 Measuring Responsiveness

We conceptualize responsiveness as a learning process wherein politicians take actions by posting on certain topics and then learn from the feedback received to inform future actions. This conceptualization perceives responsiveness as the allocation of more attention to a topic when that topic is prioritized by the target audience. Implicitly, this approach assumes that politicians' expressed issue attention serves as a reliable proxy for their actual issue priorities, or at least that politicians use their communication to convey issue attention, which audiences perceive accordingly. This approach aligns with previous research arguing that posting on a topic is a valid measure of attention paid to a topic and thus the priority one places on it (Barberá *et al.*, 2019; Schöll *et al.*, 2023).

To model the responsiveness process, we draw inspiration from (Schöll *et al.*, 2023) and create variables representing the *feedback advantage* of each topic for each post which we use to model responsiveness. The feedback advantage is a variable that for each politician individually measures how well a specific topic has fared in the recent past in terms of feedback, in relation to other topics. Intuitively, the feedback advantage of a topic should be positively correlated with the probability of posting on a topic in the future. However, the feedback from the public is not the only factor impacting the politicians' choice of topic to address. We also want to measure and control for two other key factors that may influence topic choice: the politicians' *topical interest* and the general *topic poularity* on the political agenda at the time of posting. Here we develop our methodology for measuring all three.

3.3.1 The Feedback Advantage

In order to measure the responsiveness of politicians, we calculate the feedback advantage of each topic *prior* to each post and use this to predict the topic of each post. This means that for each post we calculate I values representing the feedback advantage of the I different topics. To calculate the feedback advantage of a given focal topic i for a post at a specific timestep t, we take the difference between the recent feedback given to topic i and the average recent feedback to other topics J:

$$\Delta F_{i,t} = \overline{F}_{i,t} - \overline{F}_{-i,t} \tag{3.1}$$

Intuitively, $\Delta F_{i,t}$ captures the extent to which topic i has an advantage in terms of recent feedback compared to the average state of all other topics. To calculate $\Delta F_{i,t}$, we need to choose the period we observe recent feedback in and how we aggregate the recent feedback. Here, we allow politicians to evaluate the feedback advantage of each topic by "looking" N posts into the past and aggregating the feedback by taking the average. In this way, $\overline{F}_{i,t}$ is the average feedback given to topic i in the last N posts, excluding the feedback given to the post in question posted at time t, and can thus be defined as:

$$\overline{F}_{i,t} = \frac{1}{N} \sum_{n=1}^{N} F_{i,t-n} \tag{3.2}$$

Similarly, we can calculate $\overline{F}_{-i,t}$ by taking the average feedback in the last N posts for each of the other topics j in J and then averaging this as:

$$\overline{F}_{-i,t} = \frac{1}{J} \sum_{j}^{J} \overline{F}_{j,t} \tag{3.3}$$

An important feature of the feedback advantage is that it relates not only to the feedback given on topic i but also indirectly to the feedback given on other topics. If other topics receive more feedback, $\Delta F_{i,t}$ will fall as a consequence. Similarly, a well-received post on topic i will reduce the feedback advantage of other topics. This property, which we could call a zero-sum assumption of issue attention, ensures that for every post, we update $\Delta F_{i,t}$ for all topics regardless of whether the topics are present in the post. This assumption aims to replicate scenarios where a politician, for example, discusses environmental politics and the feedback is unfavorable. Our approach would then predict a higher likelihood of shifting to different topics, even without receiving *direct* feedback on those particular topics.

Empirically, we calculate the $\Delta F_{i,t}$ using the raw log-transformed feedback counts for each post using a look-back window of 10 posts. We do this separately for each politician and each of the three types of aggregate measures of feedback that we examine: likes, comments, and shares. While we use the average of the feedback on the last 10 posts when calculating $\overline{F}_{i,t}$, we could use other aggregations such as the cumulative sum or the max feedback. However, we believe that a rolling average provides the simplest, most intuitive measure of prior feedback for a topic, as it captures the overall level of recent feedback.

Similarly, the period of recent feedback could be defined using discrete time rather than a number of posts, which we actually use later to measure the relative popularity of each topic.

To estimate responsiveness, we estimate a linear probability model for each topic, using the feedback advantage calculated using the feedback given *prior* to each post, to predict the topic of the post[1]. If politicians are responsive, there should be a positive relationship between the feedback advantage of a topic and the likelihood of politicians posting on that topic at the next time step.

3.3.2 Topical Interest and Popularity

The politicians' interest in a topic which we call $T_{self,i,t}$ can be driven by intrinsic interest, ideology, or even institutional constraints such as a party spokesperson position. Either way, it can be measured as their relative propensity to post on a topic in the recent past and can be expressed as:

$$T_{self,i,t} = \frac{P_{self,i,t}}{\sum_j^I P_{self,j,t}} \tag{3.4}$$

where $P_{self,i,t}$ denotes the number of posts on topic i in the last 10 posts. As $T_{self,i,t}$ only captures the recent past, we also include MP fixed effects in the models to control for differences in more long-term, deep-held individual preferences for specific topics.

To measure the relative popularity of the topic, for each post, we count the number of posts on each topic by all *other* politicians in the last seven days, and use the share of posts of topic i as the measure:

$$T_{other,i,t} = \frac{P_{other,i,t}}{\sum_j^I P_{other,j,t}} \tag{3.5}$$

where $P_{other,i,t}$ denotes the number of posts on topic i in the last 7 days made by all other politicians. $T_{other,i,t}$ thus intuitively captures the popularity of the topic among other politicians in the recent past and serves as a proxy for the overall popularity or salience of topics on the political agenda at the time the post is being made. As such, it theoretically also captures prior media attention and other factors that affect all politicians.

We choose to measure the popularity of topics using a number of days instead of a number of posts to better capture the popularity of the topic at the specific moment in time each post is being made. However, we also include monthly time-fixed effects to control for any more long-term changes in topic popularity

[1]We estimate a linear probability model with each topic as the outcome separately as there to our knowledge are no inferential models that can account for the multilabel nature of the data and thus model everything at once.

and levels of social media feedback. Finally, all models are calculated with clustered standard errors at the individual MP level.

3.4 Data

The data for this paper contains all Facebook posts, made by Danish Members of Parliament (MPs) active on Facebook in the period from the beginning of 2011 to early 2018[2]. In total, the data contains 128.204 posts from 123 unique MPs. The included MPs were elected to Parliament either during the 2011 or 2015 general election, and we include only the posts these politicians made in the periods when they were in parliament, even though prior or later data is sometimes available. The data comes from the politicians' public Facebook pages which are to be confused with standard user profiles. Facebook pages are public pages used by organizations, brands, or public figures to communicate to their social media audience. The data was obtained using the Facebook API.

We measure politicians' responsiveness by aggregating feedback counts for each post, specifically tracking the number of likes, shares, and comments. Following established research methods, we add 1 and log-transform our feedback variables to address their highly skewed distribution, where a few posts receive extreme amounts of feedback while most receive very little (Schöll *et al.*, 2023). This transformation ensures better model fits and reduces the risk of extreme outliers driving effects.

Our Facebook data is enriched with information on election periods and government tenure. During the period covered by the data, Denmark held two general elections, one in 2011 and another in 2015. The prime minister of Denmark can call an election at any time, leading to election periods that typically last about three weeks. Posts made during these times are coded as occurring during election periods, while all other posts are coded as outside election periods. Similarly, we create binary variables to distinguish posts made while a politician's party was in government. In Denmark, the government usually consists of a coalition of one or more parties. Some politicians serve as ministers, while others remain as members of Parliament. Being part of the government does not necessarily mean holding an official governmental role, but it does imply being part of the ruling party and typically discussing and promoting government talking points in Parliament.

In our models investigating the heterogeneous effects of different politician types, we include a gender dummy variable and three continuous variables: standardized vote count, ideological placement, and ideological extremity. The vote count represents the standardized number of votes each politician received in the most recent parliamentary election.

[2]The four North Atlantic members of the Danish Parliament from Greenland and the Faroe Islands are not included in this analysis. Similarly, we remove politicians who were inactive in the period, meaning they did not have at least one post in half the years of the data.

Ideological placement is determined by the party affiliation of each politician. We use data from a comprehensive survey of Danish politics, where voters rank each party on a 0-10 ideological scale from left to right (Hansen and Stubager, 2017). We average these party positions to assign an ideology score to each politician based on their party's average ideological placement at the time of their post. As individual-level data for each politician is unavailable, using party ideology as a proxy provides a strong approximation of individual MP ideology.

For ideological extremity, we measure the absolute distance of each politician from the center of the ideological scale (5), resulting in an extremity score. This method ensures that both left-wing and right-wing politicians are given comparable scores if they are equidistant from the center. For example, a politician with an ideology score of 7 (relatively right-wing) and another with a score of 3 (relatively left-wing) both receive an extremity score of 2.

3.4.1 Identifying Valid and Consistent Topics

The examination of political responsiveness necessitates the definition of a coherent set of topics that politicians prioritize their attention toward. To empirically explore responsiveness, it is imperative to identify and label the data with the topics that politicians engage with. Historically, the literature has pursued two primary approaches for labeling political text: the purely deductive and the fully inductive methods.

In the deductive approach, exemplified by Baumgartner's work (2019) on categorizing topics and subtopics in US politics, researchers aim to create a comprehensive taxonomy that ensures comparability across different studies. Trained researchers utilize a codebook to meticulously label each sentence in political text with appropriate topical and subtopical codes (Green-Pedersen and Mortensen, 2019). While this method ensures replicability, its resource-intensive nature restricts its suitability for big data contexts such as ours. Moreover, the purely deductive topic classification may not fully cover the rich information present in the data, potentially leading to mismatches between predefined topics and the actual topics manifested in the data, especially when shifting the domain from political manifestoes to social media such as in our case.

On the other hand, the fully inductive approach to topic classification involves employing topic modeling algorithms, such as Latent Dirichlet Allocation (LDA), Hierarchical Stochastic Block Model (HSBM), BERTopic, etc., to classify text into topics (Blei *et al.*, 2003; Gerlach *et al.*, 2018; Grootendorst, 2022). Researchers then validate and assign names to the topics by examining keywords associated with each, using criteria such as face validity to validate them (Carlsen and Ralund, 2022). While this approach efficiently identifies meaningful topics, it may also produce numerous small and highly specific topics, that typically only exist in the data for a short period of time. This might be ideal when studying who tends to lead the political agenda (Barberá *et al.*, 2019), but when we want to examine the responsiveness to feedback, we need to examine topics that are consistently relevant in politics over time, such that they are actually valid choices to post on at most times.

In this paper, we therefore adopt a hybrid strategy for labeling topics that combines deductive and inductive methods, augmented by qualitative examination. We initiate our topic categorization with the list of topics provided by Green-Pedersen and Mortensen (2019)'s top-level topic classification. However, since Green-Pedersen and Mortensen (2019) topic classification is designed for political manifestos, we also utilize an unsupervised HSBM topic model (Gerlach *et al.*, 2018) to estimate empirical topics that may not be covered by Green-Pedersen and Mortensen (2019) or to identify subtopics that warrant consolidation. Through this process, we consolidate and combine the initial 21 topics from Green-Pedersen and Mortensen (2019) into 17 topics and introduce two topics related to religion and the EU, resulting in a total of 19 intermediate topics.

Subsequently, we undertake empirical examinations and validation of these 19 intermediate topics to ensure their relevance in our dataset. This involves creating keyword lists for each topic based on the HSBM and prior domain knowledge. We refine these lists using word similarities calculated through the word2vec algorithm (Mikolov *et al.*, 2013). We then continuously sample posts based on the keyword lists and review the content of them, in order to confirm the presence of intended topics in the data and ensure a comprehensive understanding of topic manifestations in our specific corpus. This iterative approach facilitates the grouping of related topics that may not be immediately apparent.[3] For instance, we combine issues related to religion and immigration after observing their intermingling in posts containing relevant keywords. The final outcome of this approach is a set of 15 topics, with an additional "non-topical" category, resulting in a total of 16 topics, as illustrated in Table 3.1. The 16th non-topical topic is assigned to posts that do not have any explicit references to politics. Similarly, they do not contain any mentions of activities that could be interpreted as political, such as a visit to a teachers union or a talk with a foreign delegation. Instead, these posts usually contain content related to the politicians' everyday life, family, hobbies etc.

Topic label	Topic name	Share of posts	Share of labels
0	Agriculture, Environment & Food	4.46%	4.11%
1	Crime & Justice	4.44%	4.08%
2	Culture	6.80%	6.26%
3	Defence	1.29%	1.19%
4	Economy & Business	8.04%	7.40%
5	Education	5.42%	4.99%
6	Climate & Energy	3.44%	3.17%
7	EU	3.88%	3.57%
8	Foreign Policy	5.72%	5.27%
9	Health Care	3.59%	3.31%
10	Immigration and Religion	8.16%	7.52%
11	Infrastructure	3.47%	3.19%
12	Labour	8.71%	8.02%
13	Social Policy & Welfare	7.31%	6.73%
14	Science & Technology	1.75%	1.61%
15	Non Topical	32.13%	29.58%

Table 3.1: **Topics in the data.** Total share of posts does not add to 100% because posts can be labeled with multiple or no topics. There is a relatively large share of posts which are nontopical. These typically contain text related to the politicians' everyday life, family, hobbies etc.

[3] For a detailed documentation of our topic creation process and its relationship to Green-Pedersen and Mortensen (2019) topic classifications, please refer to Appendix 6.2.1

The meticulous process of identifying and validating these topics ensures their consistency and relevance to the discursive landscape on social media over time. Crucially, this enables us to examine responsiveness across topics and time, as we have acess to data on all topics at all times as seen in figure 3.1.

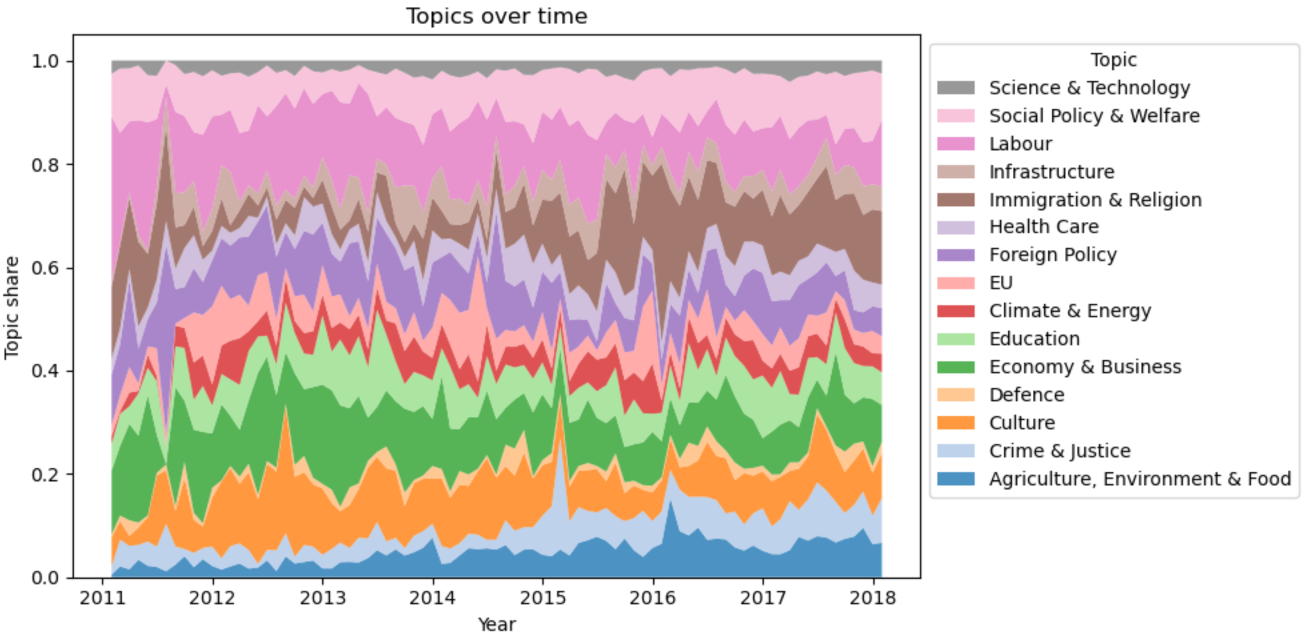


Figure 3.1: Topic distribution over time. The figure shows the share of each topic for each month. The non-topical label has been left out for visual clarity

3.4.2 Labelling Topics

Having identified a comprehensive set of topics, the next crucial step in our analysis involves assigning these topics as labels to the posts. While a straightforward approach could simply entail using the aforementioned curated list of keywords for this task, we instead pursue a supervised classification strategy using deep learning transformer models, which enables us to capture a lot more contextual information and semantic complexity.

To train the supervised classifier, we manually label a subset of 7,500 posts, ensuring adequate representation for each topic. We construct this labeled dataset, by sampling 250 posts for each topic (totaling 3,750 posts) by selecting posts containing keywords from our list of topical keywords. Within each topic, we further sample from the subset of posts that have the highest proportion of topical keywords specific to that particular topic. Additionally, we randomly sample an additional 3,750 posts from the remaining data, ensuring that the labeled dataset encompasses posts not exclusively centered on specific topics. This approach provides a balanced representation of posts across various topics, while also considering the diversity of discussions on social media.

The labeled dataset is split into a training set of 6,000 posts and a hold-out test set of 1,500 posts. To ensure intercoder reliability, three separate labelers label 2,000 posts each from the training set and all

1,500 posts in the test set. The labeling process for the training set is iterative, involving discussions and minor adjustments to the codebook and topic definitions to achieve consistency and accuracy. For the test set, each labeler independently assigns labels to calculate an unbiased intercoder reliability measure.

The agreement among labelers is highly favorable, with 97.1% of the posts (1,456 posts) having consistent labels from at least two labelers. Considering all three labelers 75.5% (1,133 posts) show complete agreement. Moving forward the validated 1,456 posts are retained for the test set, as they ensure robust evaluation of model performance.

To classify the topics of the remaining unseen posts, we fine-tune several pre-trained Danish large language models with transformer architectures (Certainly, 2022; CFHC, 2023; Kummervold *et al.*, 2021; Snæbjarnarson *et al.*, 2023; Devlin *et al.*, 2018). We fine-tune the models by conducting masked language modeling for 1 epoch on the complete dataset and subsequently adding a classification layer for 10 additional epochs on the training and evaluation sets. The best performing model is the fine-tuned dfm-encoder-large-v1 (CFHC, 2023), which is also the largest Danish language model to date (CFHC, 2023).

The best model demonstrates a sample average test F1 score of 0.78, with slightly higher precision than recall. Overall, the model exhibits satisfactory performance in predicting the topics present in the text, comparable to other similar studies (Schöll *et al.*, 2023). For a detailed overview of the model's performance on each topic, the training strategies employed, and the performance off all tested models we refer to appendix 6.2.2.

3.5 Results

To examine the general dynamics of political responsiveness, we now delve into the general patterns of how politicians engage with social media feedback across different topics and feedback types. Figure 3.2 illustrates the association between the feedback advantage ($\Delta F_{i,t}$) and the probability of a post featuring topic i. Notably, for all topics, the feedback advantage is significantly ($p < 0.01$) positively correlated with the likelihood of the post containing that specific topic. The effects are calculated while controlling for the politicians' own topical interest ($T_{self,i,t}$), the topic popularity ($T_{other,i,t}$) as well as fixed effects for politician and time (month). This suggests that the feedback advantage of a topic plays a pivotal role in shaping the politicians' choices for future posts, even when accounting for agenda, individual idiosyncrasies, and time-specific effects.

This relationship holds true for all three types of feedback measured likes, shares, and comments. Shares generally exhibit the smallest effect size of around 0.1, implying that a 1-unit increase in feedback advantage for shares corresponds to approximately a 10% increase in the probability of the subsequent post focusing on topic i. Both likes and comments demonstrate slightly stronger effect sizes for most topics, hovering closer to 0.13. In general, the pattern seems to be that shares have the lowest effect, then likes and finally comments, but the differences are not stark. As such, it does not in general seem like the influence of

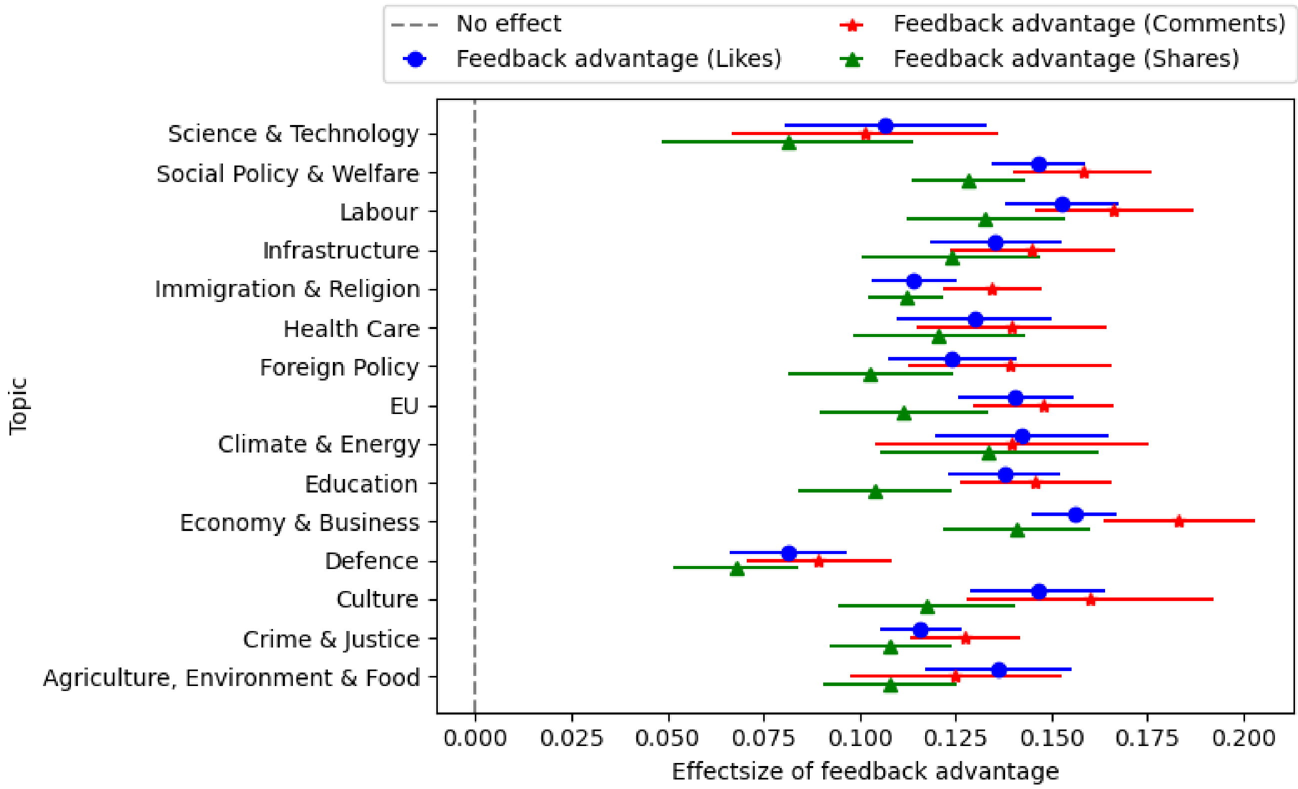


Figure 3.2: Effect size of feedback advantage on topic probability. This figure depicts the effect size of feedback advantage for predicting topic i using linear probability models. The models are estimated with topical interest and topic popularity, fixed effects on the individual and monthly basis and clustered standard errors.

high-effort feedback types, such as comments and shares, provide politicians with more of a signal of support than likes.

The specific quantity of likes, comments, or shares required for a 1-unit increase in feedback advantage varies over time, contingent on the prior feedback to topic i and all other topics. Given the log transformation of the feedback variables, a 1-unit increase in the average difference between feedback to topic i and the average feedback to other topics corresponds to an approximately 3-fold increase in the actual feedback to topic i, holding the feedback to other topics constant. This means that the amount of likes to topic i needs to increase by a factor of about 3, with likes to other topics remaining at the same level, for the probability of topic i to increase the around 13%.

To contextualize these coefficients, we examine the pooled standard deviation of the feedback advantage to gauge how much politicians typically alter their behavior after each post. Consider the feedback advantage of the topic *Crime & Justice* as an illustrative example, with an effect size of roughly 0.12 and a standard deviation of approximately 0.41 as seen in appendix 6.2.3. The feedback advantage of the topics *Crime & Justice* is thus on average expected to change by 0.41 for every post, resulting in a change of $0.12 \times 0.41 = 0.05$ - equating to about a 5% change in the probability of the next post addressing *Crime* *and Justice*. This means that on average, politicians change their probability of posting on *Crime & Justice* roughly 5% after every post.

Another way to examine the effects of feedback on topic probability is to compare the effects of the feedback advantage, to the effect of the general popularity of the topics and the politicians' own tendency to post on the topic. If figure 3.3 we see that this comparison for likes, across all topics. What we can see here is, that feedback is not the most important thing in terms of predicting what a politician posts on next. The politicians' topical interest is a lot more important than the feedback they receive. If we again use the topic of *Crime & Justice* as an example, we can see that for every extra post concerning *Crime & Justice* in the last 10 posts, the probability that the post at time t is also concerned with *Crime & Justice* increases by roughly $0.25 * 0.1 = 2.5\%$, as one post would equal a 0.1 increase in the topical interest variable. The pattern is the same for shares and comments as seen in appendix 6.2.4. Interestingly we see little to no effects of topic popularity at the time of the post. The topic popularity is insignificant for most topics, and even has a small negative relation to topics probability for most.

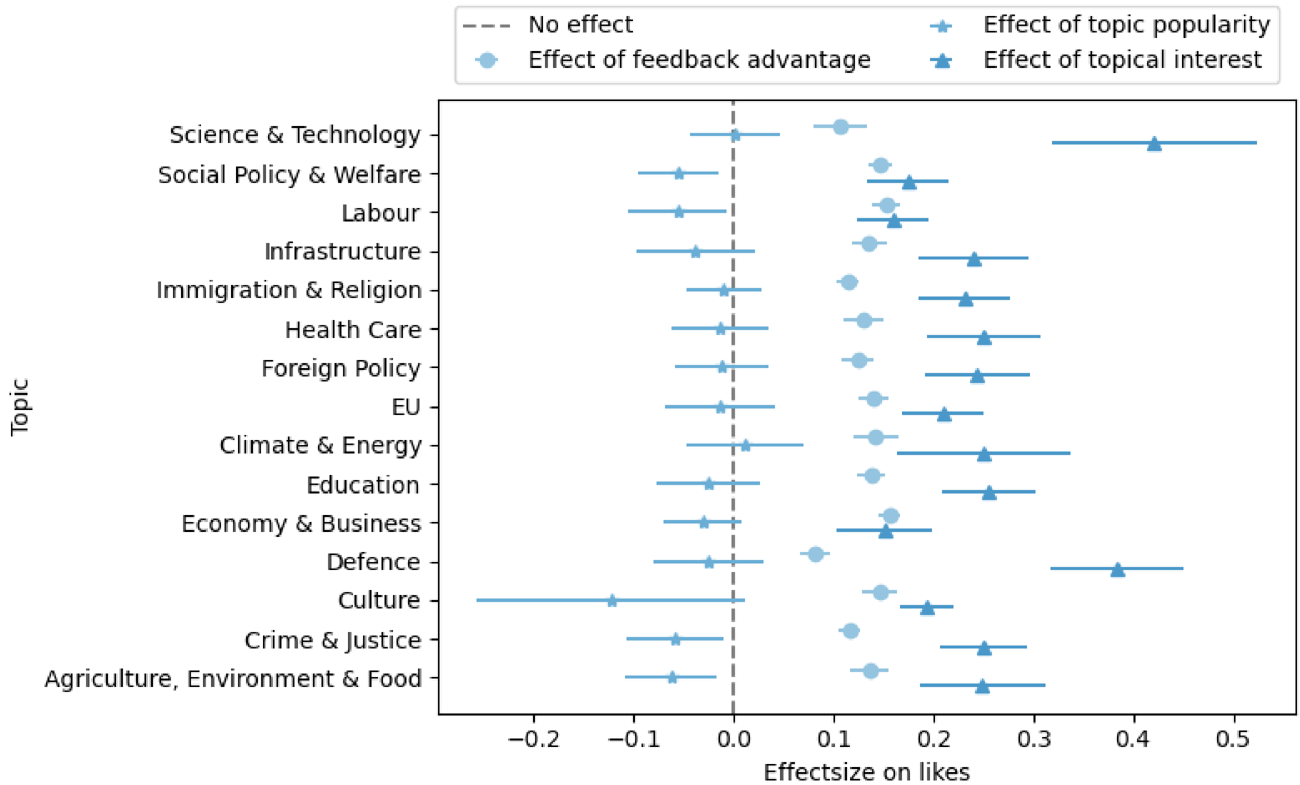


Figure 3.3: **Effect size of feedback advantage, topical interest and topic popularity for Likes on topic probability.** This figure depicts the effect size of feedback advantage for predicting topic i using linear probability models. The models are estimated with topical interest and topic popularity, fixed effects on the individual and monthly basis and clustered standard errors.

In summary, our findings reveal a consistent pattern of responsiveness to social media feedback across various topics and feedback types. There is a robust relationship between prior social media feedback and subsequent topic choice. The effect of feedback seems smaller than the effect of politicians' own tendency to post on a topic but still has a statistically and practically significant effect on politicians' future choice of topics. As such we argue that we find strong evidence in favor of H1 concerning politicians' responsiveness to social media feedback across topics. Shares have a slightly weaker relationship with future topic choices than likes and comments, but all feedback types prove relevant for determining future topics. We thus argue that overall we find strong support for H2 suggesting that politicians are equally responsive to different types of social media feedback. While there are small differences, they vary across topics, all types of feedback are positively related to future topic choices and the difference between them is negligible.

To ascertain the universality of these patterns, we now explore the heterogeneous effects of social media feedback on topic choice in different significant political periods as well as how this responsiveness varies across politician-specific traits. In the following analysis, we focus solely on the effects of likes while noting that similar results hold for comments and shares as seen in the appendix. We also focus on the three most common topics in our data; *Labour*, *Immigration & Religion*, and *Economy & Business*. This is done for visual and analytical clarity, but when important we comment on the pattern across all topics. Full plots and regression tables with all the topics are available in the appropriate appendixes. All models are estimated with feedback advantage, topical interest, topic popularity, the variable for heterogeneous effects, as well as an interaction variable between the feedback advantage and the heterogeneous effect.

3.5.1 Responsiveness in Different Periods

We begin our examination of responsiveness in different relevant political periods, by examining political responsiveness during elections. Overall, our findings indicate that politicians remain responsive both during and between elections. In Figure 3.4 (left), we observe that the feedback advantage of likes predicts an increased probability of posts on all the examined topics, both during and between elections. The main effect is thus significant for all topics during both periods. The same is true for all other topics as seen in appendix 6.2.7.

Our results also reveal, that for the largest three topics, politicians are less responsive to feedback during elections compared to outside election periods. Similar to the main effect, we also find this pattern across all topics as seen in appendix 6.2.7. The negative and significant interaction term between the election period and the feedback advantage suggests that politicians need a larger feedback advantage on a topic to persuade them to post on it. This is in line with our initial expectations and may suggest that during campaigns, politicians follow more carefully planned strategies, reducing their reliance on immediate social media feedback to choose what topics to post on. We thus find strong support for H3 suggesting that politicians are less responsive during elections. Notably, politicians are still responsive during election periods, just less than between elections.

Turning to the periods when politicians are in government, Figure 3.4 (right) demonstrates that politicians, regardless of their government status, are responsive to feedback for all topics. The feedback advantage is significantly associated with topic choice for both government and opposition politicians. This holds true across all topics as seen in appendix 6.2.8.

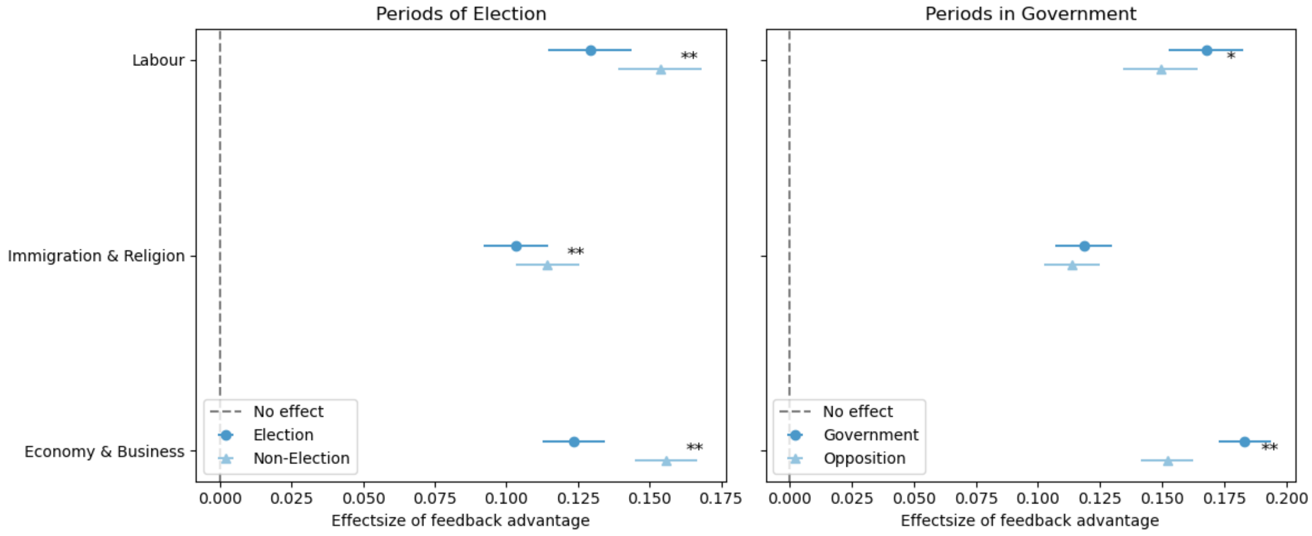


Figure 3.4: Heterogeneous effects of election and government periods. Left: Effect size of feedback advantage (Likes) for predicting topic i during election periods. Right: Effect size of feedback advantage (Likes) for predicting topic i during periods when politicians are part of the government or not. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

In line with theoretical expectations of hypothesis H4b, we see that at least for *Labour*, and *Economy & Business* there is a significant positive relationship between being in government and responsiveness. It thus appears that politicians in government are to some extent more responsive than those in opposition. This relationship is however not significant for *Immigration & Religion*, though the coefficients point in the same direction. In appendix 6.2.8 we can see that when looking across all topics, there is a consistent pattern of the direction of the interaction term, but it is only statistically significant for a little over half of the topics. We thus find no evidence for the prediciton of hypothesis H4a, that politicians should be less responsive during government. The findings do however lend some support for the notion that politicians in government would exhibit more responsiveness, though it is not as conclusive as with election periods.

In summary, our results suggest that politicians are generally responsive to social media feedback across different periods in the electoral cycle and whether or not they are in government. Politicians appear less responsive during elections and to some extent more responsive when in government, both in line with prior theoretical expectations.

3.5.2 Responsiveness Among Different Politicians

We now delve into an exploration of the responsiveness of various types of politicians, investigating the potential impact of gender, electoral incentives, ideology, and ideological extremity on their responsiveness to social media feedback.

Commencing with an analysis of gender and electoral incentives, Figure 3.5 (left) reveals that both men and women demonstrate responsiveness to social media feedback. Across the three major topics, no significant difference in responsiveness between men and women is discernible. We even see that the coefficients of the interaction term point in both directions, with men being more responsive on *Labour* and *Immigration & Religion*, and women on *Economy and Business*. If we examine all topics in appendix 6.2.9 we can see that there are only two topics with a significant difference *Defense* and *Education*, and even here we observe a bidirectional pattern, with women being more responsive to *Education* and men displaying higher responsiveness to *Defense*. Thus, there is no consistent and overarching pattern of gender-based differences in responsiveness. As expected we find strong evidence for H5, and the notion that any gender differences in responsiveness are domain-specific.

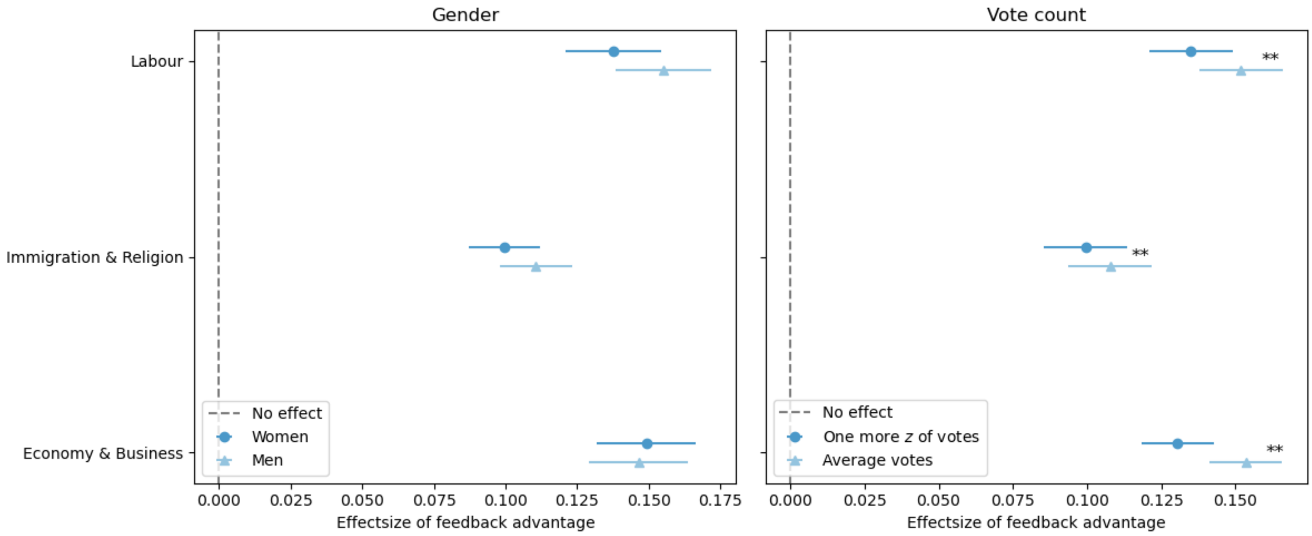


Figure 3.5: Heterogeneous effects of gender and vote count. Left: Effect size of feedback advantage (Likes) for predicting topic i based on gender. Right: Effect size of feedback advantage (Likes) for predicting topic i based on vote counts at last election. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Examining responsiveness based on electoral incentives in figure 3.5 (right), we observe a pattern indicating that politicians, irrespective of their electoral success, demonstrate general responsiveness to social media feedback. Notably, both politicians with high and low vote counts appear affected by the feedback they receive.

However, in line with our hypothesis H6a, but contrary to hypothesis H6b there is a discernible trend where politicians with fewer votes exhibit higher responsiveness. This effect is consistent across all topics except one as seen in appendix 6.2.7. As such we find no evidence for H6b suggesting that politicians with more votes are more responsive. One suggestive explanation of this finding could be, that politicians facing lower levels of electoral support perceive social media as a vital avenue to connect with constituents, understand public sentiment, and potentially enhance their appeal. In a sense, they have more of an incentive to be responsive to social media feedback, as they are close to not getting reelected.

Now turning to consider the ideology of the politicians in figure 3.6 (left), we see that politicians, regardless of ideological position appear to be responsive to social media feedback. Similarly, they also appear to be equally responsive to social media feedback on the left and right wing of politics. This finding remains consistent across all topics except *Climate & Energy*, where the left-leaning politicians are more responsive than the right-leaning politicians as seen in appendix 6.2.11. This demonstrates that ideology, as measured on a left-right scale, does not significantly alter politicians' responsiveness. This is an interesting finding, that runs counter to our prior expectations that right-leaning politicians are more responsive. As such, we do not find any evidence for H7 that the right-leaning politicians should be any more responsive than their left-leaning colleagues.

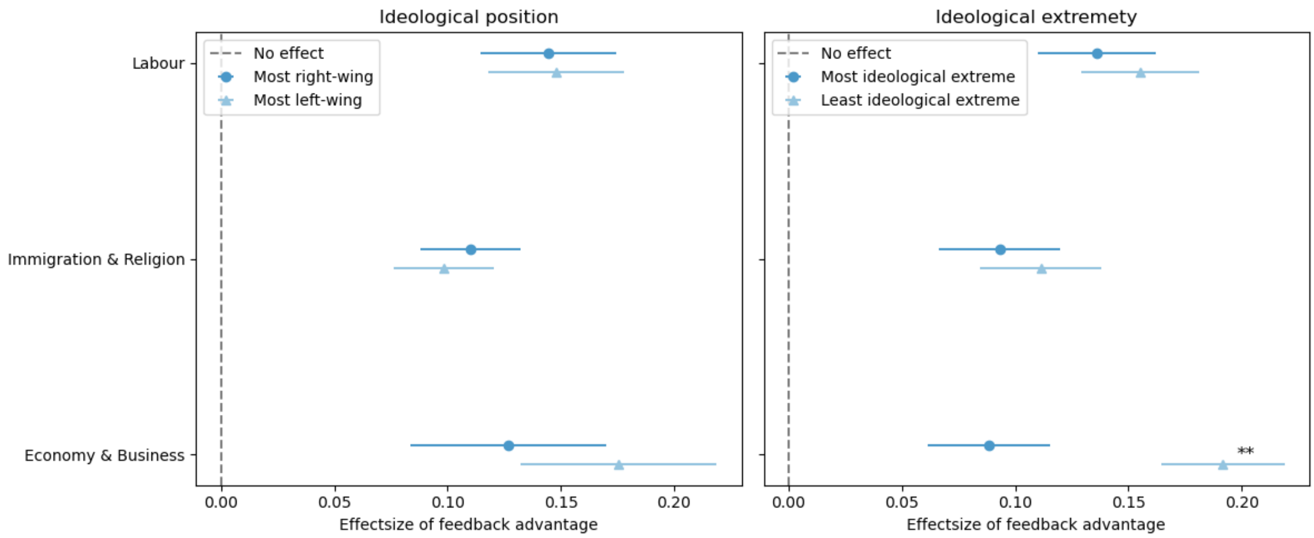


Figure 3.6: Heterogeneous effects of ideology and ideological extremity. Left: Effect size of feedback advantage (Likes) for predicting topic i based on ideology. Right: Effect size of feedback advantage (Likes) for predicting topic i based on ideological extremity. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Shifting the focus from absolute ideological positions to relative positions from the middle of the ideological scale i.e. their ideological extremism, figure 3.6 (right) reveals an interesting pattern. First of all, we again see that politicians, regardless of ideological extremism, exhibit responsiveness to social media feedback. We also see that for *Economy & business*, there is a large significant difference, in the direction that the more ideologically extreme politicians are also less responsive. The direction of the effect holds for almost all topics, as seen in appendix 6.2.12, but for most topics, the effects are statistically insignificant. This suggests, that to the extent that ideological responsiveness has an influence on responsiveness to social media feedback, centrist politicians are the most responsive. However, this effect appears to be at best domain-specific.

These findings are somewhat in line with our prior theoretical expectations and provide some support for H8, however, the result is only suggestive as the difference is not significant for a number of topics and even has a reverse association on some topics. One speculative explanation could be, that ideologically extreme politicians also focus on very specific and/or niche topics that are ideologically passionate about, and as such listen less to feedback from the public, potentially diminishing their responsiveness to popular topics on social media. This line of reasoning would be in line with Ennser-Jedenastik *et al.*, 2022 who also that show that niche parties are less responsive.

3.6 Discussion and Conclusions

The central finding of the study is that there is a clear pattern of political responsiveness to social media feedback amongst politicians, across all types of feedback topics, periods, and types of politicians. This finding reinforces prior research suggesting that social media may serve as a valuable channel for politicians to gauge public policy preferences and priorities (Ennser-Jedenastik *et al.*, 2022; Schöll *et al.*, 2023), and cement social media feedback as a viable mechanism driving political responsiveness in the social media age.

In line with the initial expectations, the study does not find any major differences in political responsiveness to different types of social media feedback. As such it appears that neither the effort level of providing the feedback nor the ambiguity of the feedback dominates the politicians' interpretation of the feedback type. While all politicians are responsive to social media feedback we see that in line with prior expectations, politicians are more responsive between elections and while in government. This may suggest that politicians follow more tight plans during elections. In terms of politicians in government, they may have higher agenda-setting power and/or more resources to address all topics and can thus more easily switch between topics based on social media feedback, whereas politicians outside of government, may to a large extent have to address whatever is on the agenda. In terms of gender, we don't find any differences, which is what we expected. However, when examining vote count we find that politicians who received fewer votes last election exhibit a higher degree of responsiveness to social media feedback. Finally, we also examine ideology where we somewhat surprisingly find no differences based on ideology. However, when examining ideological extremity, we find some suggestive evidence that the more centrist politicians are more responsive to social media feedback, but more research is needed to determine this for certain.

Having established that politicians are generally responsive to social media feedback, we need to consider the implications of this for modern democratic governance. On the positive side, social media allows for more immediate and direct communication between politicians and their constituents, enabling elected officials to better comprehend public preferences and priorities. This direct feedback mechanism may facilitate greater accountability and responsiveness to the needs of the electorate, bolstering democratic governance.

However, challenges emerge concerning the representativeness of social media feedback. Social media feedback does not fully reflect the broader public's views and is instead a biased sample from within their own social media echo chamber (Wojcieszak *et al.*, 2022). Politicians must therefore be cautious not to over-rely on social media feedback as the sole source of information, as it may lead to responsiveness towards skewed policy priorities and an incomplete understanding of their constituency's diverse needs. Instead, a balanced approach that incorporates diverse information channels, including traditional polling data and offline interactions, can complement social media feedback to provide a more comprehensive picture of public preferences and priorities.

While this study's findings provide valuable insights, it is crucial to acknowledge that political responsiveness and the influence of social media feedback can be context-specific. Different political systems, cultural norms, and societal dynamics may influence how politicians engage with social media and respond to feedback. While we attempt to generalize prior research we still only examine one country and the main social media platform there. Future research should therefore explore how these contextual factors shape the dynamics of political responsiveness on social media to gain a more nuanced understanding of the phenomenon, specifically by examening more countries, electorial systems, and social media platforms.

Furthermore, while we show a consistent pattern of responsiveness to social media feedback, on social media, this may not translate into political representation in other fora. Politicians may very well optimize their social media feedback by choosing their more popular topics but then actually work on policy in completely different areas. In other words, social media issue attention, may not match the issue attention of the same politician in another domain. Future research should seek to examine the link between political responsiveness on social media, and other domains.

In conclusion, this study makes a significant contribution to the study of political behavior in the digital age, uncovering the pivotal role of social media feedback as a mechanism driving political responsiveness. By demonstrating politicians' responsiveness across diverse topics, feedback types, periods, and types of politicians the research advances our understanding of the mechanisms that underlie political responsiveness in representative democracies in the social media age. The findings underscore the potential of social media as a tool for direct political engagement, while also highlighting the importance of a balanced and comprehensive approach to political representation. As we continue to navigate the complex landscape of digital political communication, further research into the contextual and cross-national variations of political responsiveness on social media remains essential for a general understanding of democratic governance in the digital era.

Bibliography

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker (2019). „Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data". In: *American Political Science Review* 113.4, pp. 883–901.

Baumgartner, Frank R., Christian Breunig, and Emiliano Grossman (Mar. 2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press. eprint: https://academic.oup.com/book/11237/book-pdf/50004199/9780192572226_web.pdf.

Baumgartner, Frank R. and Bryan D. Jones (2009). *Agendas and Instability in American Politics*. Second. University of Chicago Press, pp. 48–58.

Belchior, Ana Maria, Hugo Ferrinho Lopes, Luís Cabrita, and Emmanouil Tsatsanis (2024). „Party policy responsiveness at the agenda-setting and decision-making stages: The mediating effect of the types of government and promise". In: *International Political Science Review* 45.3, pp. 316–335.

Bischof, Daniel and Markus Wagner (2020). „What Makes Parties Adapt to Voter Preferences? The Role of Party Organization, Goals and Ideology". In: *British Journal of Political Science* 50.1, pp. 391–401.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (Mar. 2003). „Latent Dirichlet Allocation". In: *J. Mach. Learn. Res.* 3.null, pp. 993–1022.

Carlsen, Hjalmar Bang and Snorre Ralund (2022). „Computational grounded theory revisited: From computer-led to computer-assisted text analysis". In: *Big Data & Society* 9.1, p. 20539517221080146. eprint: https://doi.org/10.1177/20539517221080146.

Caughey, Devin and Christopher Warshaw (2018). „Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 19362014". In: *American Political Science Review* 112.2, pp. 249–266.

Certainly (June 2022). *Certainly has trained the most advanced Danish Bert Model to date*.

CFHC (2023). *Centre-for-humanities-computing/danish-foundation-models: A project for training foundational Danish language model*.

Chen, Kaiping, Nathan Lee, and William Marble (Nov. 2018). „How Policymakers Evaluate Online versus Offline Constituent Messages". In: Available at SSRN: https://ssrn.com/abstract=3251651 or http://dx.doi.org/10.2139/ssrn.3251651.

De Bruycker, Iskander and Anne Rasmussen (2021). „Blessing or Curse for Congruence? How Interest Mobilization Affects Congruence between Citizens and Elected Representatives". In: *JCMS: Journal of Common Market Studies* 59.4, pp. 909–928. eprint: https://onlinelibrary-wiley-com.ep.fjernadgang.kb.dk/doi/pdf/10.1111/jcms.13146.

De Vries, Catherine E. and Gary Marks (June 2012). „The struggle over dimensionality: A note on theory and empirics". en. In: *European Union Politics* 13.2, pp. 185–193.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805, pp. 1–16. arXiv: 1810.04805.

Edwards, George C. and B. Dan Wood (1999). „Who Influences Whom? The President, Congress, and the Media". In: *American Political Science Review* 93.2, pp. 327–344.

Ennser-Jedenastik, Laurenz, Christina Gahn, Anita Bodlos, and Martin Haselmayer (2022). „Does social media enhance party responsiveness? How user engagement shapes parties issue attention on Facebook". In: *Party Politics* 28.3, pp. 468–481. eprint: https://doi.org/10.1177/1354068820985334.

Gerlach, Martin, Tiago P. Peixoto, and Eduardo G. Altmann (2018). „A network approach to topic models". In: *Science Advances* 4.7, eaaq1360. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aaq1360.

Green-Pedersen, Christoffer (July 2019). *The Reshaping of West European Party Politics: Agenda-Setting and Party Competition in Comparative Perspective*. en. Oxford University Press.

Green-Pedersen, Christoffer and Peter B. Mortensen (Mar. 2019). „82The Danish Agendas Project". In: *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press. eprint: https://academic.oup.com/book/0/chapter/159764743/chapter-pdf/50004241/oso-9780198835332-chapter-8.pdf.

Grimmer, Justin (2013). „Appropriators not position takers: The distorting effects of electoral incentives on congressional representation". In: *American Journal of Political Science* 57.3, pp. 624–642.

Grootendorst, Maarten (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv: 2203.05794 [cs.CL].

Hansen, Kasper Møller and Rune Stubager (2017). *2015-valget: Oprør fra udkanten. Folketingsvalget 2015*. København: Jurist- og Økonomforbundets Forlag.

Jones, Bryan D. and Frank R. Baumgartner (2004). „Representation and Agenda Setting“. In: *Policy Studies Journal* 32.1, pp. 1–24. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0190-292X.2004.00050.x.

Klüver, Heike and Iñaki Sagarzazu (Mar. 2016). „Setting the Agenda or Responding to Voters? Political Parties, Voters and Issue Attention“. en. In: *West European Politics* 39.2, pp. 380–398.

Klüver, Heike and Jae-Jae Spoon (2016). „Who Responds? Voters, Parties and Issue Attention“. In: *British Journal of Political Science* 46.3, pp. 633–654.

Kummervold, Per E, Javier De la Rosa, Freddy Wetjen, and Svein Arne Brygfjeld (2021). „Operationalizing a National Digital Library: The Case for a Norwegian Transformer Model“. In: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. Reykjavik, Iceland (Online): Linköping University Electronic Press, Sweden, pp. 20–29.

Lee, Nathan (2022). „Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers“. In: *American Political Science Review* 116.2, pp. 677–688.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781 [cs.CL].

Neundorf, Anja and James Adams (2018). „The Micro-Foundations of Party Competition and Issue Ownership: The Reciprocal Effects of Citizens Issue Salience and Party Attachments“. In: *British Journal of Political Science* 48.2, pp. 385–406.

Norton, Philip (1993). *Does Parliament Matter?* Harvester Wheatsheaf, pp. 35–37.

Page, Benjamin I. and Robert Y. Shapiro (1983). „Effects of Public Opinion on Policy“. In: *American Political Science Review* 77.1, pp. 175–190.

Pedrazzani, Andrea and Paolo Segatti (2022). „Responsiveness when parties are weak: A candidate-based analysis of voter-party congruence in Europe“. In: *Party Politics* 28.1, pp. 149–162. eprint: https://doi.org/10.1177/1354068820968091.

Schilozman, Kay Lehman, Sidney Verba, and Henry E. Brady. (2012). *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton University Press.

Schöll, Nikolas, Aina Gallego, and Gaël Le Mens (2023). „How Politicians Learn from Citizens Feedback: The Case of Gender on Twitter“. In: *American Journal of Political Science* n/a.n/a, pp. 557–574. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12772.

Snæbjarnarson, Vésteinn, Annika Simonsen, Goran Glava, and Ivan Vuli (May 2023). „Transfer to a Low-Resource Language via Close Relatives: The Case Study on Faroese“. In: *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*. Tórshavn, Faroe Islands: Linköping University Electronic Press, Sweden.

Soroka, Stuart N. and Christopher Wlezien (2009). *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge University Press.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson (1995). „Dynamic Representation“. In: *American Political Science Review* 89.3, pp. 543–565.

Sulkin, Tracy (2005). *Issue Politics in Congress*. Cambridge University Press.

Tarik Abou-Chadi, Christoffer Green-Pedersen and Peter B. Mortensen (2020). „Parties policy adjustments in response to changes in issue saliency“. In: *West European Politics* 43.4, pp. 749–771. eprint: https://doi.org/10.1080/01402382.2019.1609296.

Wagner, Markus and Thomas M. Meyer (Sept. 2014). „Which Issues do Parties Emphasise? Salience Strategies and Party Organisation in Multiparty Systems“. en. In: *West European Politics* 37.5, pp. 1019–1045.

Wlezien, Christopher (1995). „The Public as Thermostat: Dynamics of Preferences for Spending“. eng. In: *American journal of political science* 39.4, pp. 981–1000.

Wojcieszak, Magdalena, Andreu Casas, Xudong Yu, Jonathan Nagler, and Joshua A. Tucker (2022). „Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity“. In: *Science Advances* 8.39, eabn9418. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abn9418.

4 Analyzing Differences between Discursive Communities using Dialectograms

By Thyge Ryom Enggaard, August Lohse, Morten Axel Pedersen, and Sune Lehmann

Abstract

Word embeddings provide an unsupervised way to understand differences in word usage between discursive communities. A number of papers have focused on identifying words that are used differently by two or more communities. But word embeddings are complex, high-dimensional spaces and a focus on identifying differences only captures a fraction of their richness. Here, we take a step towards leveraging the richness of the full embedding space, by using word embeddings to map out *how* words are used differently. Specifically, we describe the construction of *dialectograms*, an unsupervised way to visually explore the characteristic ways in which each community uses a focal word. Based on these dialectograms, we provide a new measure of the degree to which words are used differently that overcomes the tendency for existing measures to pick out low-frequency or polysemous words. We apply our methods to explore the discourses of two US political subreddits and show how our methods identify stark affective polarisation of politicians and political entities, differences in the assessment of proper political action as well as disagreement about whether certain issues require political intervention at all.

4.1 Introduction

The field of Natural Language Processing (NLP) has developed embedding models to nunumerically represent and quantify the content of texts. Here, words are represented as high-dimensional vectors that capture how words are associated in a training corpus. Embeddings serve various purposes, such as classifying or partitioning text, translating between languages and engaging in dialogues via chatbots like ChatGPT.

In social science, these models allow for a quantitative study of how social actors communicate. Word embeddings, in particular, enable researchers to quantify the similarities (or differences) between word use among language users by identifying which words are used most similarly in a particular corpus. This approach has been applied to examine the semantic neighborhoods of obesity-related words in a corpus of NY Times articles (Arseniev-Koehler and Foster, 2020). Word embeddings can also help identify how relatively similar words are to distinctions such as *woman* vs *man*. This approach has been used for instance to study how different occupations are associated with gender, ethnicity and class distinctions (Bolukbasi *et al.*, 2016; Caliskan *et al.*, 2017; Garg *et al.*, 2018; Jones *et al.*, 2020).

Beyond examining how words relate within a single corpus, embeddings also allow for mapping differences and similarities *between* corpora. Thus embeddings of corpora from two languages (English and Spanish) have been used to automatically translate between the languages (Mikolov, Le, *et al.*, 2013). In other work, embeddings of corpora from different points in time have been used to detect how the meaning of words used by a language community shifts over time (Hamilton *et al.*, 2016; Schlechtweg *et al.*, 2020).

Recently, a new kind of comparison has been pursued between corpora from communities communicating in the same language, period, and about the same topic (KhudaBukhsh *et al.*, 2021; Milbauer *et al.*, 2021; Azarbonyad *et al.*, 2017). Based on methods from machine translation and semantic change detection, it can be used to identify ideological differences in a language community, or what some have dubbed 'ideolects': "a language shared by an ideological group, which necessarily contains the private worldview of that group and may not be naively decipherable to those outside" (Milbauer *et al.*, 2021). So far, such work has focused on simply *detecting* which words are used differently by measuring the distance between word embeddings or by identifying words that mistranslate between embeddings.

But what do ideolects *consist* of? While semantic contexts might explain such mistranslations (such as *kkk/blm* being catagorized a hate group in the commentary on CNN/Fox respectively (KhudaBukhsh *et al.*, 2021)), no systematic way of accounting for the difference in embedding has so far been proposed. Here, we pursue the question of how to systematically identify and explain such ideolectical differences between language communities. We contribute to the literature by developing a method for both *detecting* and *explaining* the difference in how language communities use a given word through visualizations. We refer to these visualizations as *dialectograms*. Dialectograms show how the projection of words along the difference in the embedding of a focal word, captures the characteristic ways in which each corpus uses the focal word. In addition, this approach also allows for the construction of a new aggregate measure of difference in word use, which overcomes the tendency for cosine distance and mistranslations to pick out low-frequent and polysemous words (words with multiple senses, like *settlement*). By calculating the degree to which the characteristic use of each corpus is also *unique* to each corpus, our measure of *sense separation* is designed to single out words, that each corpus uses in distinct ways. As such we contribute to the current literature by developing a method that not only identifies relevant ideolectical differences, but also seeks to explain these, thus allowing for for in-depth analysis of group differences based on word embeddings than previously possible.

In the following, we begin by providing an example of a dialectogram, after which we describe its construction in some detail. We then introduce sense separation, our new measure of difference in word use. Finally, we use this new measure to carry out a detailed, exploratory analysis comparing two US political subreddits, r/Democrats (511.906 comments) and r/Republican (576.872 comments), from January 2011 to September 2022 (For a detailed description of our data collection, preprocessing and corpora construction, see appendix 6.3.1). We find stark affective polarization of partisan references, mainly characterized by derogatory cross-partisan discourse. We also identify non-partisan references, where the dialectograms suggest that the differences rather arise from divergent assessments of policy relevance. Finally, we show that on average across all words, the characteristic word use in r/Democrats tends to focus on electoral and legislative issues, while the characteristic work use in r/Republican tends to focus on (contentious) topics of policy.

4.2 Results

To begin, figure 4.1 shows the dialectogram for the word *republican*[1]. The words associated with the characteristic way *republican* is used in r/Republican are located in the upper right quadrant; they centre around *candidates* and how they are perceived (e.g. *misunderstood*, *constructive*, *hopeful*). In addition, there is a range of words related to Reddit itself (e.g. *post*, *sub*, *comment*), that stem from internal talk about the republican subreddit[2]. Opposed to this is the characteristic use of *republican* in r/Democrats, located in the lower left quadrant, which appears to be mainly derogatory (e.g. *filth*, *treasonous*, *brainwash*).

The words that both communities associate more with their own use of *republican* are located in the lower right quadrant. These tend to be more descriptive in the form of near-synonyms (*republicans*, *gop*), names of republican politicians (e.g. *goldwater*, *donald*), movements (*maga*, *tea*), or *conservative* and *evangelical* associations. Words that each community associate more with the other community's characteristic use are located in the upper left quadrant. Again, we see that these words tend to be descriptive, this time in the form of references to democrats (e.g. *democrats*, *leftist*, *pelosi*). Inspecting the translation of *republican* between the corpora (provided on the axis labels in Figure 4.1), this is due to the strong discursive equivalence of *republican* and *democrat* in these subreddits; the way r/Republican uses *republican* is most similar to the way r/Democrats uses *democrat*, while the way r/Democrats uses the word *republican* is most similar to the way r/Republican use *democrat*.[3]

We will return to a more extensive analysis of the partisan divide in section 4.2.3 below. But first, we describe how the dialectogram in Figure 4.1 is constructed (section 4.2.1), as well as how dialectograms enable us to suggest a new measure of difference in use (section 4.2.2).

4.2.1 Constructing Dialectograms

Constructing the dialectogram in Figure 4.1 involves three overall steps. First, we embed each corpus, which requires choosing an embedding model. Second, we align the embeddings, such that the embedding

[1]One criticism of unsupervised text analysis, like we deploy here, is the reliance on a relatively low number of words to interpret the fitted models. For this reason, we provide full-page visualisations, with a large number of annotated words. While the annotations might occasionally clutter, we believe they overall provide a more detailed yet still readable way to interpret the dialectograms.

[2]We conclude this based on sampling and reading comments in the r/Republican subreddit, that include these words

[3]To translate a focal word from one corpus to the other, we find the word in the later embedding that is closest to the embedding of the focal word in the former. For more information, see appendix 6.3.2.

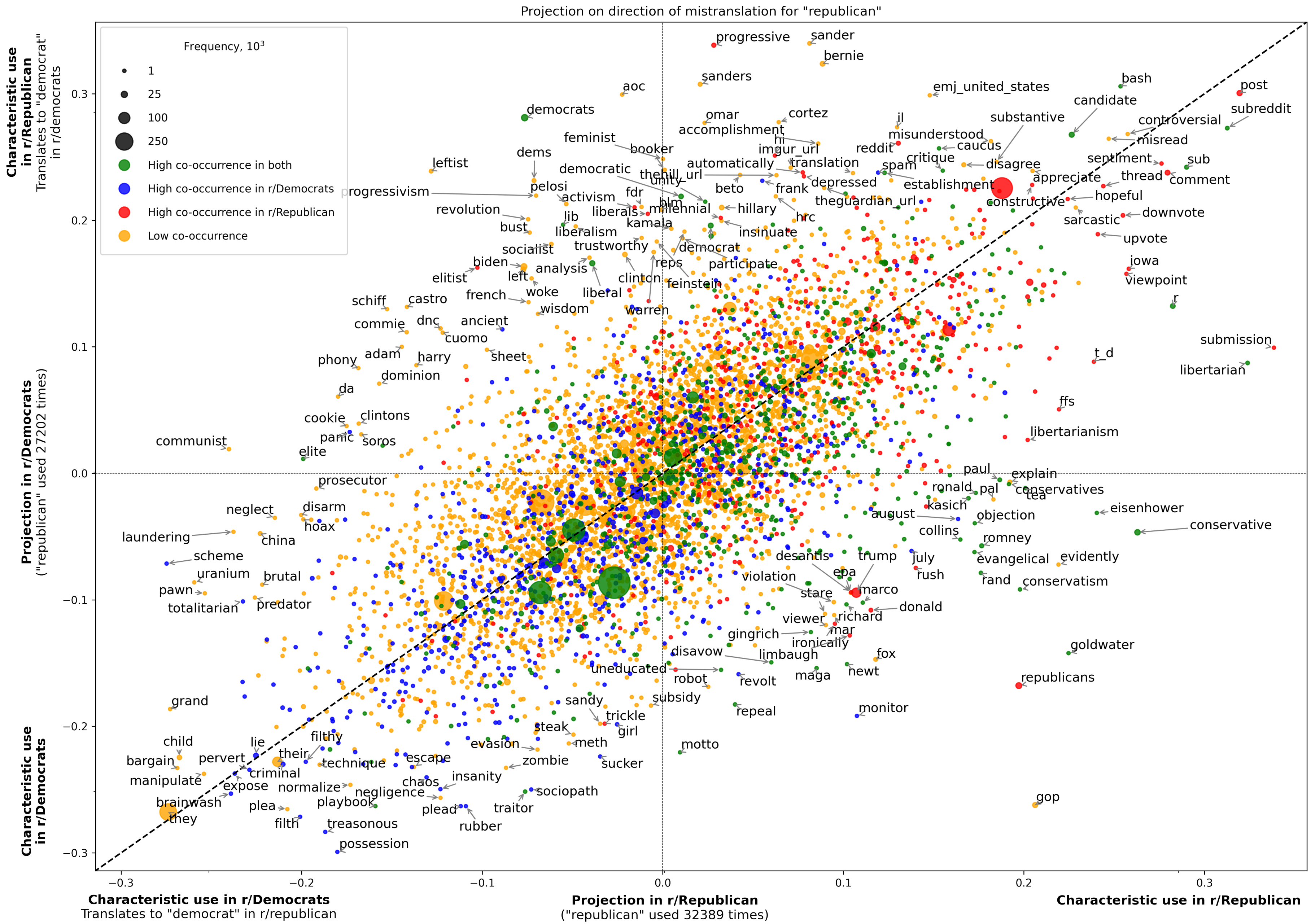
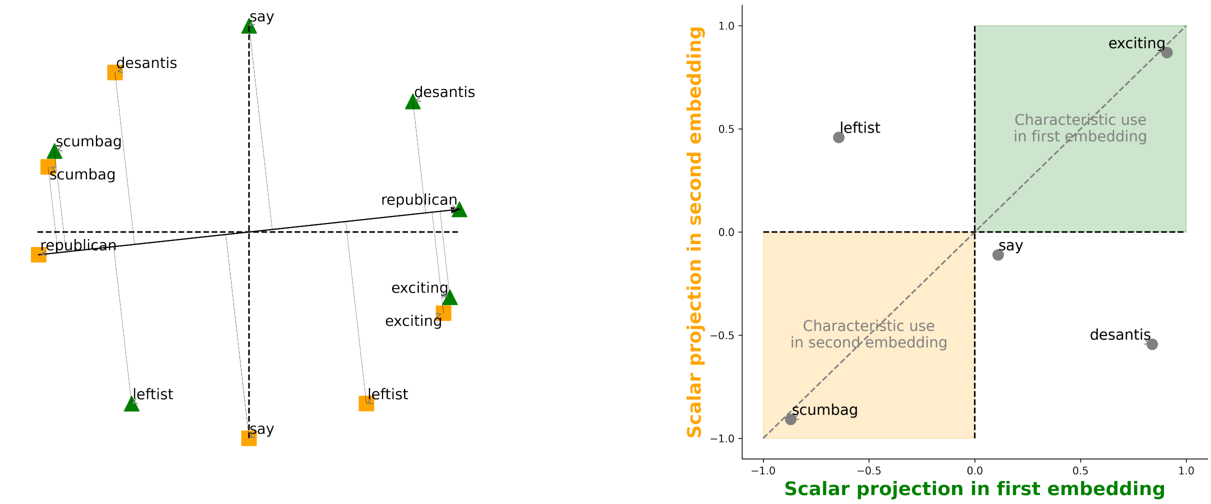


Figure 4.1: Projection of words on the offset for the embeddings of *republican*. The dispersion of words along the x-axis might appear greater than along the y-axis, as the figure has been stretched to the page. Words are coloured according to their co-occurrence with *republican*; see Equation 4.2 for the definition of high co-occurrence.



(a) Projecting onto the offset: Given two aligned embeddings (one marked with green triangles, the other with orange squares) and a focal word (here *republican*), we first identify the vector corresponding to the difference between the embeddings of the focal word in the two corpora, marked as the solid black line. We then project the remaining vocabulary onto this offset, to create a graded measure of how each corpus relatively associates each word with the two embeddings of the focal word.

(b) Comparison of scalar projections: Plotting the two scalar projections of each word against each other, we can distinguish between the characteristic ways each corpus associates the focal word (*exciting* vs. *scumbag*, areas marked with green and orange respectively). In addition, words in the two remaining quadrants (*desantis*, *leftist*) are relevant to the distinction between the characteristic uses, but the corpora do not agree to which characteristic use they belong. Finally, words close to the centre are not relevant to the difference in use.

Figure 4.2: Illustration of our geometric approach to accounting for the disputed conceptions of a word

of words from different corpora can be compared. This requires choosing an alignment method. Third, to create the dialectogram for a focal word, we project each embedding onto the direction corresponding to the difference in the embeddings of the focal word. For each word, we obtain two projections, corresponding to the projection from each of the two corpora. We then construct the dialectogram by plotting these scalar projections against each other as seen in Figure 4.2. Below, we consider each of the steps in turn.

Embedding the corpora

In an exploratory analysis like ours, there is no straightforward way to decide which embedding model is best. We therefore devise a validation task mirroring the task of finding difference in meaning. Specifically we evaluate how well the models identify the degree to which a word has been swapped with another random word within its frequency decile and part-of-speech category, to evaluate if the model captures the new meaning the swapped word is being bestowed by its swapped context. By varying the degree to which we swap words within each pair, we are able to compare how well the different embedding models recover the degree of change based on a range of different measures. In appendix 6.3.3, we describe the validation task further and provide the results of comparing two static embedding models (SGNS-Word2Vec and GloVe) and one contextual embedding model (DistilBERT) (Mikolov, Chen, *et al.*, 2013; Pennington *et al.*, 2014; Liu *et al.*, 2019; Sanh *et al.*, 2020). Overall, we find little difference in model performance except that the static models are able to capture both the degree of swap and the swapped words themselves.

As GloVe embeddings are more stable than SGNS embeddings, we present our results based on GloVe embeddings here, but our method is not restricted to this choice of model (Wendlandt *et al.*, 2018; Antoniak and Mimno, 2018).

Aligning the embeddings

Training two GloVe embeddings on the same corpora will not yield the exact same numerical representation though the embedings might still reflect the same structural representation of the corpus. To make the embeddings trained on r/Democrats and r/Republican comparable, we therefore align the embedding spaces by identifying a linear transformation between them. As part of our validation task, we compare two approaches to identifying such linear transformations, Procrustes Analysis and Canonical Correlation Analysis, and find they work equally well (Schönemann, 1966; Artetxe *et al.*, 2016; Balbi and Misuraca, 2006; William, 2011). As Procrustes Analysis has the desirable property of leaving all pairwise distances within each embedding space unaffected, we use it here. For a description of the two alignment methods, see appendix 6.3.2; for the comparison in the validation task, see appendix 6.3.3.

Comparing the projections

Having obtained two aligned embedding spaces, we can now create the dialectogram of a focal word i. First, we identify the vector offset between the embeddings of word i, $O_i = w_{1,i} - w_{2,i} \in \mathbb{R}^D$, where $w_{k,i} \in \mathbb{R}^D$ is the embedding of word i in corpus $k \in \{1, 2\}$. Intuitively, the vector offset O_i captures the way word i is used differently in the two corpora. While there are likely many similarities in how a word is used, this approach allows us to hone in on the difference, no matter how small[4].

To measure how the remaining words are related to the difference in use of the focal word, we project the embeddings onto this vector offset (for an illustration, see Figure 4.2a). Specifically, we obtain the scalar projections from the embedding of corpus k, $E_k \in \mathbb{R}^{N \times D}$ as

$$\alpha_i^k = \frac{E_k O_i}{\| O_i \|_2} \in \mathbb{R}^N \tag{4.1}$$

By plotting the scalar projections from each embedding against each other, we can distinguish and interpret the different ways in which the two corpora use the focal word (see Figure 4.2b) for an illustration). Overall, we distinguish between three scenarios.

First, words projecting relatively close to zero in both embeddings (*say* in the example in Figure 4.2b) are unrelated to the difference in how the focal word is used. Second, words projecting positively in both

[4] In practice, we find that only plotting the projections of words, that co-occur at least once with the focal word in either of the corpora, makes the dialectograms easier to interpret. While the projection of non-co-occurring words might provide a way to assess the most implicit associations within each corpus, we found they muddled our interpretation. Further, we leave out the three highest frequent words (*be*, *do*, *have*) in order to illustrate word frequency by marker size, as well as the focal word itself.

embeddings (*exciting*, upper right quadrant marked with green) are, in both aligned embeddings, closer to the embedding of the focal word in the first (green) embedding, and hence account for the characteristic way in which the first corpus uses the word. Similarly, words projecting negatively in both embeddings (*scumbag*, lower left quadrant marked with orange) account for the characteristic way in which the second corpus uses the word. In general, these make up the *relative* difference in how the corpora use the focal word (*republican*) - both corpora might still use the word in both ways, just to varying degrees.

Third, words projecting positively in the first embedding but negatively in the second (*desantis*, lower right quadrant) are more associated with each corpus' own characteristic use. Likewise, words projecting negatively in the first embedding but positively in the second (*leftist*, upper left quadrant) are, in each corpus, more associated with the other corpus' characteristic use. Both corpora associate the words in these quadrants with one of the characteristic uses but do not agree to which characteristic use they belong.

4.2.2 Identifying Words Used Most Differently

A dialectogram can be constructed for any word that appears in both of the aligned embeddings. As our vocabulary consists of 5.738 unique words, it is infeasible to inspect the dialectograms of all words. If a researcher has a particular theoretical or empirical emphasis, this might provide a way to select which words to create dialectograms of. Another approach to identify such differences, would be to measure the degree to which words are embedded differently in the two copora[5].

While cosine distance and mistranslation have traditionally been used to measure difference in embeddings space, their tendency to pick out low-frequent or polysemous words make them less than ideal. Here we describe these two challenges in turn, and then introduce our measures of sense separation which builds on the intuition of dialectograms to pick out words that carry the most difference in meaning across corpora.

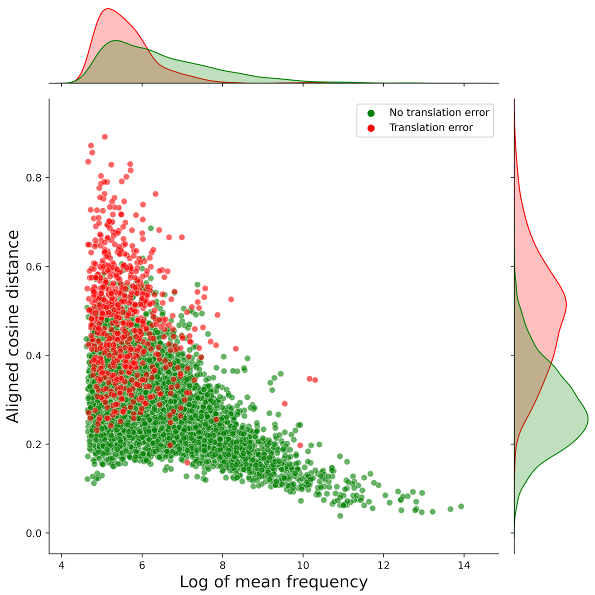
Difference in use and frequency

Figure 4.3a shows the relation between aligned cosine distance and the log of mean frequency across the two corpora. Cosine distance correlates negatively with frequency (Spearman's correlation: -0.51), implying that words exhibiting high cosine distance tend to be used less. As a larger distance makes translation errors more likely, mistranslations are also more prevalent among low-frequent words. This issue can partially be overcome by focusing on the mistranslating words that have the highest frequency.

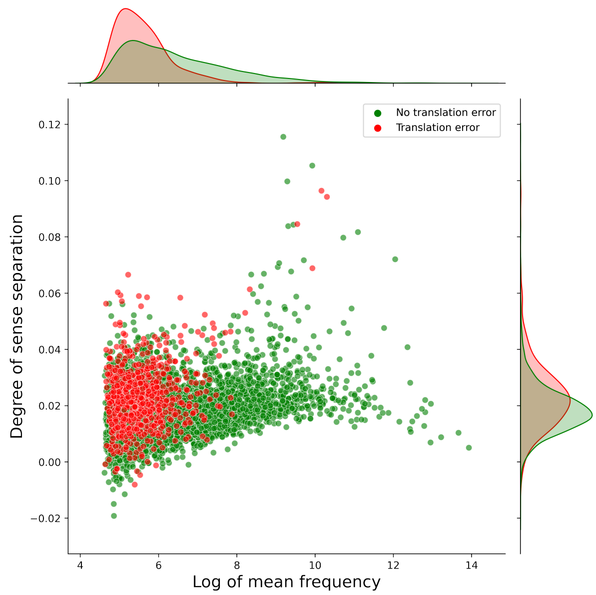
In the case of semantic change detection, this negative correlation has been interpreted as a substantial feature of language (Hamilton *et al.*, 2016). In our comparative setting, it seems unlikely, however, that the degree to which a word is used differently between two corporas should be determined by the frequency of the word. Instead, the negative correlation at hand may be a property of the embedding models, an

[5]Several measures of difference between embeddings exist and as part of our validation in appendix 6.3.3, we describe and compare some of these

explanation supported by controlled experiments (Dubossarsky *et al.*, 2017). As such, using cosine distance or high-frequent mistranslations as a method for identifying focal words that are used differently, would introduce undue bias towards low-frequent words. In appendix 6.3.2 section we describe this challenge further, along with our (failed) attempts to remove the correlation.



(a) Cosine distance: Scatter plot of the log mean frequency of each word against its aligned cosine distance. Cosine distance correlates negatively with frequency, why it tends to identify low-frequent words as used most differently.



(b) Sense separation: Scatter plot of the log mean frequency of each word against its degree of sense separation. Sense separation identifies a range of higher-frequent words as used most differently; these also include the highest-frequent mistranslations.

Figure 4.3: Relation between measures of difference in use and frequency. Words are coloured according to whether their two embeddings are each others' nearest neighbours across embeddings (no translation error) or not.

One adjustment does however support the interpretation of the dialectograms: projecting each embedding to the orthogonal complement of the direction that varies most with the logarithm of word frequency. Intuitively, projecting the embeddings to the complement of this direction should remove the information about frequency implicitly embedded in the vector spaces.

Unfortunately, this strategy does not remove the negative correlation, which suggests that the relation between frequency and embedding position is not encoded in a single direction. However, for words that are used much more by one corpus than the other, this adjustment does balance the projections in the dialectogram around the origin. For this reason, we apply the adjustment to the embeddings, before we align them (for a description of this adjustment, see appendix 6.3.2).

Difference in use and polysemy

In addition to correlating with frequency, cosine distance and high-frequent mistranslations also tend to pick out polysemous words when applied to our corpora. The word *settlement* e.g. has a high cosine distance; by inspecting the dialectogram, the two characteristic uses correspond to the sense of an inhabited place vs. an agreement. The dialectogram also indicates that both corpora use the word in both senses, as evidenced by the presence of words in both of the characteristic uses, that co-occur highly with *settlement* in both of the corpora. In general, even if the polysemous word is used in the same set of senses by both corpora, differences in the degrees to which each sense is used result in differences in the embedding of the word (Arora *et al.*, 2018).

In some cases, such differences might be particularly interesting for social scientists. For example, we can see that with the word *flood* r/Republican tends to talk relatively more about a metaphorical flood of illegal immigration, while r/Democrats are relatively more occupied with the management of a fluid flood. But there are also cases, where the difference in the degree to which the senses of a polysemous word are used, might be less interesting. The word *sheet* e.g. exhibits high cosine distance - inspecting the dialectogram, we interpret this as the difference between a sheet of toilet paper and a fact sheet, such as published by institutions like the White House. Therefore, using cosine distance or high-frequent mistranslations to identify focal words, would tend to pick out words, where the two corpora discuss completely different issues, rather than different perspectives on the same issues.

Sense Separation

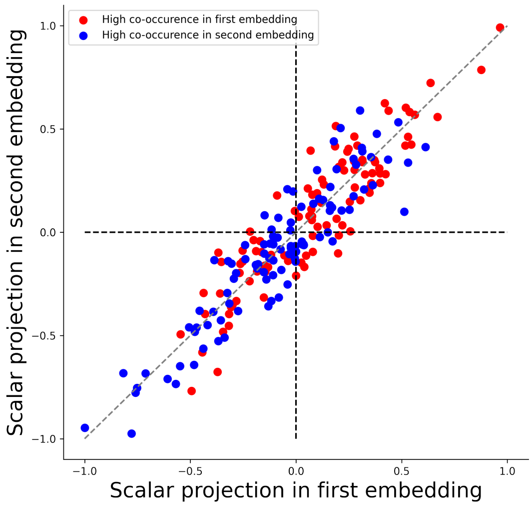
Based on the dialectograms, we devise sense separation which is a new measure of difference in word usage between copora which measures the degree to which the characteristic use of each corpus is also unique to that corpus. The intuition behind sense separation is illustrated in Figure 4.4. Here, we show two hypothetical dialectograms, that correspond to a low and high degree of sense separation. In these hypothetical dialectograms, the projected words are coloured according to whether they co-occur highly with the focal word in one or the other corpus. Our sense separation measure captures the degree to which these highly co-occurring words are separated between the two characteristic uses. In particular, Figure 4.4a illustrates a case where there is almost no separation between the uniquely high co-occurring words (red and blue dots); in this case, the sense separation measure will be low. In contrast, Figure 4.4b illustrates a case, where the separation is much more pronounced; in this case, the sense separation measure is high.

To calculate the sense separation for any focal word we identify the words in each corpus, that co-occur more than expected if co-occurrence was independent. If $C^k_{i,j}$ is the co-occurrence count between focal word i and context word j in a corpus k, N^k_c is the sum of all co-occurrence counts in that corpus and N_w is the number of words, the criteria is:[6]

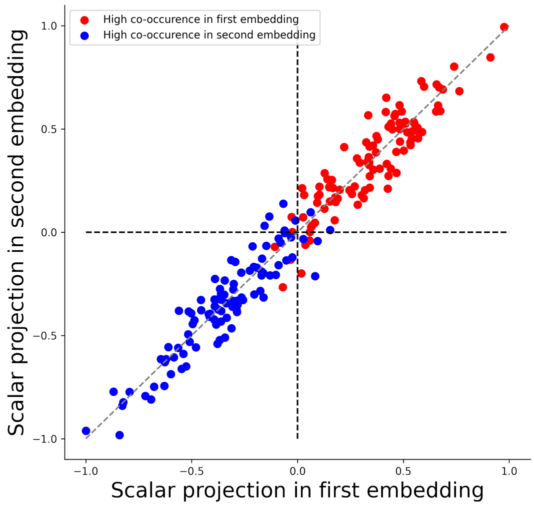
$$EC^k_{i,j} = \frac{C^k_{i,j} \cdot N^k_c}{\sum_{h=1}^{N_w} C^k_{i,h} \cdot \sum_{h=1}^{N_w} C^k_{h,j}} > 1 \tag{4.2}$$

Given the sets of words co-occuring highly with the focal word in each corpus, we identify the subset of these words, that only co-occur highly in one of the corpora, but not both, $HC^1_i = \{j \in \{1, .., N_w\} \mid$

[6] This is equivalent to cases where the pointwise mutual information (PMI) between the words is positive.



(a) Low sense separation: The words that co-occur highly with the focal word in one or the other corpora (red and blue dots) are not separated between the upper right and lower left quadrant.



(b) High sense separation: The words that co-occur highly with the focal word in one or the other corpora (red and blue dots) are much more separated between the upper right and lower left quadrant.

Figure 4.4: Illustration of the intuition behind sense separation: Our measure of difference captures the degree to which the words, that co-occur highly with the focal word in only one of the corpora, are separated between the two characteristic uses.

$EC^1_{i,j} > 1 \wedge EC^2_{i,j} \leq 1\}$ for the first corpus and likewise for the second. Intuitively, these words mark the associations with the focal word that are distinct for each corpus. For each set of words, HC^1_i and HC^2_i, we calculate the mean of their scalar projections ($\alpha^1_{i,j}$ and $\alpha^2_{i,j}$) on the offset for the focal word - in the case of HC^1_i,

$$\overline{HC^1_i} = \frac{1}{|HC^1_i|} \sum_{j \in HC^1_i} \frac{\alpha^1_{i,j} + \alpha^2_{i,j}}{2} \in [-1, 1] \tag{4.3}$$

and likewise for HC^2_i. Taking the average of the scalar projections is equivalent to projecting the words onto the dashed line running along the main diagonal in the dialectogram; taking the mean of these averages hence captures the overall position of the uniquely high co-occurring words along the main diagonal. Finally, to get our measure of sense separation we subtract the two means,

$$S_i = \overline{HC^1_i} - \overline{HC^2_i} \tag{4.4}$$

As Figure 4.3b shows, the sense separation measure is less correlated with frequency than cosine distance (Spearman's correlation: 0.16). Further, the highest-frequent mistranslations also show a high degree of sense separation. As such we find that sense separation intuitively captures not just distance in embedding space, but the difference directly relevant to the difference in meaning, and does so while being less biased by frequency and still capturing the highest-frequent mistranslations.

4.2.3 Comparing r/Republican and r/Democrats

Table 4.1 shows the 25 words used most differently in the two corpora, measured by (i) cosine distance, (ii) highest frequent mistranslations and (iii) sense separation (see appendix 6.3, Tables A.46, A.47 and A.48 for top 100 words). Only six words appear in more than one list (*republican*, *republicans*, *democrats*, *liberal*, *newt* and *cruz*), and none in all three, highlighting that these measures capture distinct differences between the embeddings.

Highest cosine distance	Highest-frequent mistranslations	Highest sense separation
voluntary	republican	r
juvenile	democrats	biden
com	republicans	comment
🤪	liberal	democrats
gingrich	cruz	republican
sheet	leftist	liberal
👏	unfortunately	gop
settlement	subreddit	left
keystone	sander	trump
newt	warren	she
spur	😂	he
❤️	kasich	conservative
karl	yep	hillary
airline	bs	clinton
captain	harris	republicans
mission	desantis	democrat
port	orange	race
rescue	rubio	sanders
rove	approve	pbs
cough	son	progressive
🤡	🤣	cruz
🇺🇸	mitch	obese
url	ted	fox
lastly	rand	broadband
forbes_url	ah	newt

Table 4.1: **Words used most differently, as measured in three ways**

We see that some words are clearly partisan, such as names of politicians (e.g. *biden*, *cruz*, *gingrich*) and party references (e.g. *republicans*, *democrats*, *gop*), well-known ideological positions (*liberal*, *conservative*), concrete political issues (such as the *keystone* pipeline), and reference to media (*forbes_url*, *pbs*, etc.).

However, the lists also comprise non-partisan words, including emojis (such as 👏) and various internet/platform references (*subreddit*, *r*, *com*, *url*). This is important, since it suggests that ideolectcial differences are not restricted to the specific uses and meanings of explicitly ideological and partisan words and concepts but extends to the much broader of domain of political language and social media discourse as such. It also suggests that dialectograms, and other so-called 'computational hermeneutic' methods (Mohr *et al.*, 2015), might be of much more general use in the social sciences.

To explore these points and possibilities, we now use dialectogram to examine partisan and nonpartisan discourse and to visualize differences between distinct word uses in the two corpora[7].

Partisan Discourse

As Figure 4.1 shows, the difference in how the two corpora use *republican* consist of within party candidate neutral evaluation versus across-party derogatory talk. Inspecting the dialectograms of other partisan references and politicians' names, this pattern consistently re-appears, indicating that our method may the used for evaluating partisan affiliations more generally. Consider Figure 4.5, which projects US politicians along the mean direction of offsets[8,9]. To see what discursive variation the mean offset might corresponds to, we also plot the words that overall project most negatively or positively on the mean offset in each embedding.

We see the appearance of the same opposition between derogatory words (lower left quadrant) versus evaluating and appraising candidates (upper right quadrant), suggesting strong degree of affective political polarization between the corpora (Iyengar *et al.*, 2012). In line with prior research, cross-party animosity seems particularly strong as compared to within-party appraisal (Rathje *et al.*, 2021). While the positive associations remain within rather neutral descriptions of political candidates (e.g. *substantive*, *pragmatic*, *enthusiasm*), the negative associations do not seem bound to an electoral or legislative domain but are outright derogatory (*filth*, *bastard*, *rapist*). We also see that there are some differences between the words the two corpora associate with each end of the spectrum - notions of *commie* and *communist* e.g. project negatively in r/Republican but neutral in r/Democrats, which in turn associates the upper right quadrant more with *activism* and *progressivism*.

We also see that *conservatism* and *conservative* project positively in r/Republican but more neutral in r/Democrats, and that r/Republican uses *communist* in ways that are very similar to a derogatory term, whereas this word projects more neutral in r/Democrats. The fact that the two corpora have so different understandings of common and ostensibly neutral concepts suggest that do the two groups don't just disagree about who is the filth, bastard and rapist, but about the basic meaning of fundamental ideological, political and societal concepts at large.

[7] While we only show and discuss the dialectograms for a few selected words what follows, the dialectograms for all words listed in Table 4.1 can be found at the project GitHub repository: https://github.com/A-Lohse/Dialectograms

[8] Any shared distinction is not unique in the direction of the offsets - in the embedding of r/Republican, we e.g. expect derogatory words to project negatively for republican politicians, but positively for democratic politicians. Hence we calculate the mean offset across all politicians' offsets (Republicans and Democrats), calculate the cosine similarity between each politician's offset and this mean offset, flip the direction of the offset for all politicians, whose cosine similarity to the mean is negative, and finally re-calculate the mean offset, before we project all politicians onto it.

[9] We identify politicians referenced in the corpora by inspecting words tagged as proper nouns

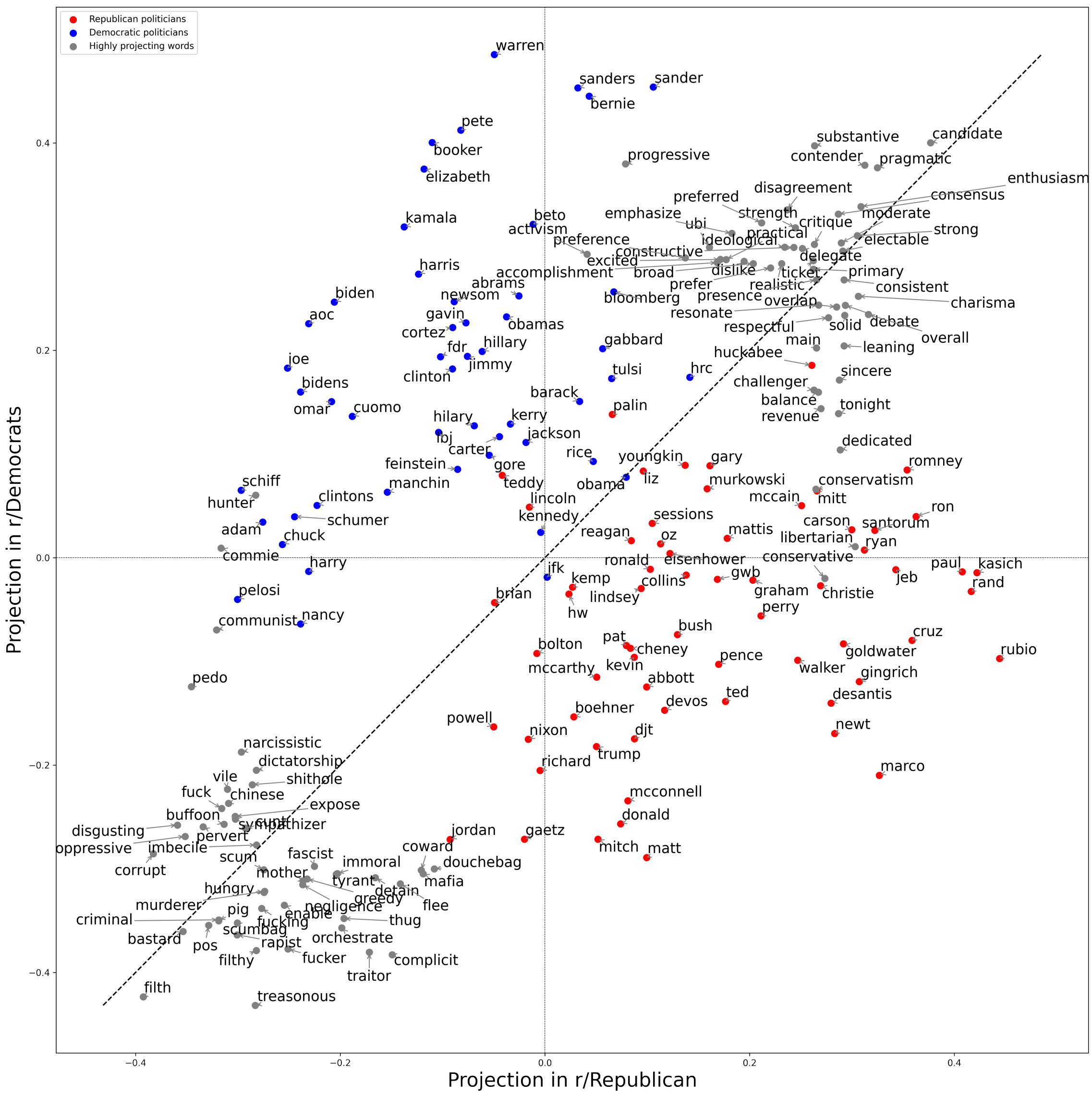


Figure 4.5: **Projection of words on the mean offset of US politicians.**

Non-partisan discourse

While the explicitly partisan words exhibit affective polarization, such polarization is less evident, if appearing at all, when we inspect the non-partisan words in Table 4.1. For example, Figure 4.6 shows the projection on the offset for *juvenile*. Rather than any clear-cut polarization, what seems to be at play here are two different conceptions of what *juvenile* means. On one hand, in r/Republican the concept is associated with *violence*, *insults* and *disrespect*, and with policies demanding *arrest* and *incarceration*. There is also salience in terms of demographic categories, particularly race and ethnicity (*latino*, *black*, *white*, *hispanic*). On the other hand, the use of *juvenile* in r/Democrats is more connoted with *reform*, *justice*, *system*, with political emphasis on the rights of the *defender* and financial implications (*bail*, *economic*). While both corpora associate their own characteristic use with notions like *abusive* and *embarrassing*, they associate the other corpus' use more with *ego*.

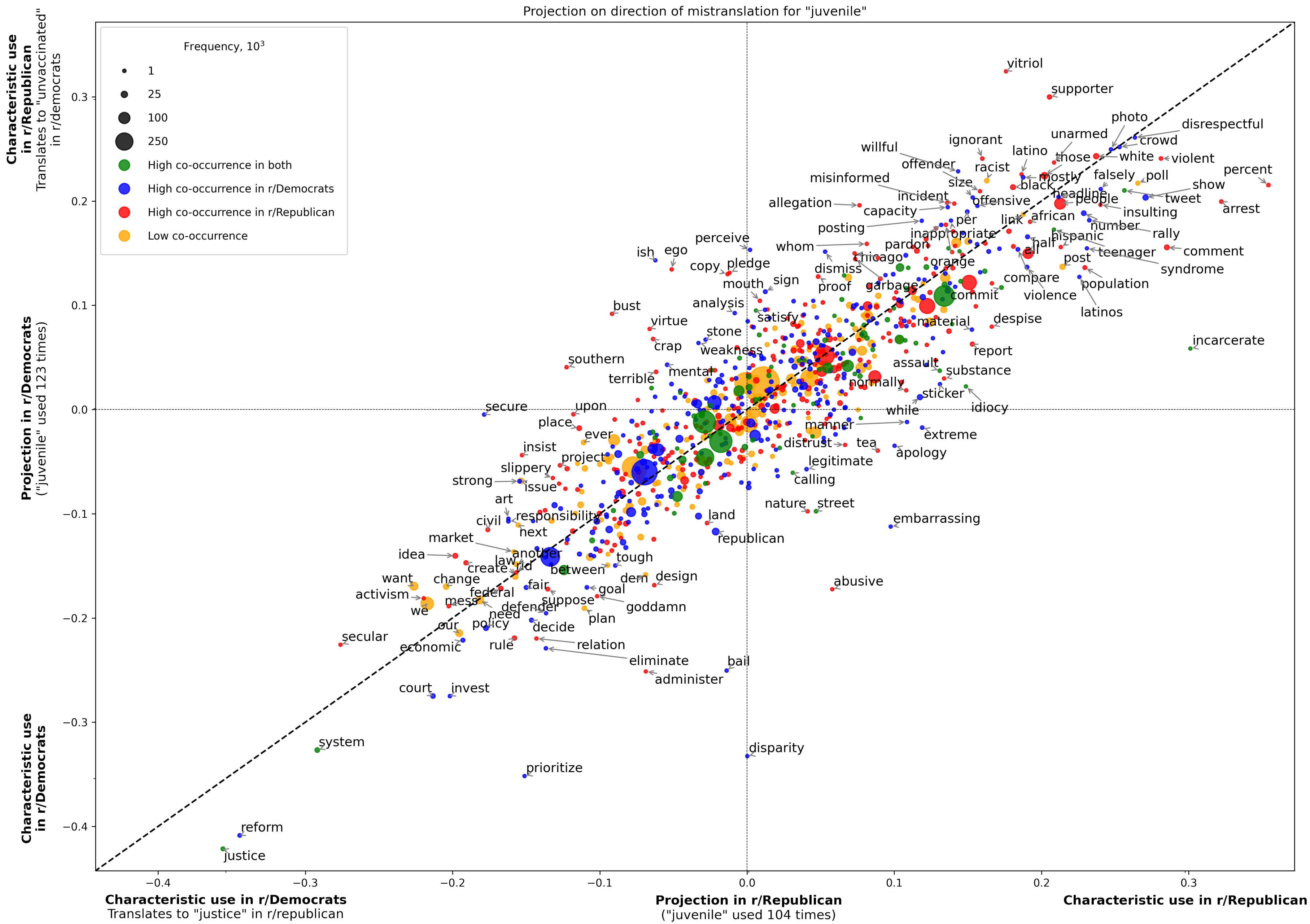


Figure 4.6: Projection of words on the offset for the embeddings of *juvenile*. The dispersion of words along the x-axis might appear greater than along the y-axis, as the figure has been stretched to the page. Words are coloured according to their co-occurrence with *juvenile*; see Eq. 4.2 for the definition of high co-occurrence.

Figure 4.6, then, is a visualization of differences between the two corpora that are not affective (as in Figure 4.6), but policy oriented. We find other such non-affective differences between r/Republican and r/Democrats around the *keystone* pipeline, where the former is dominated by industrial references (*gas*, *oil*, *production*, *drive*, etc.), and the latter with political-economic ones (*unemployment*, *gdp*, *benefit*) and illegality (*jail* and *sentence*). This would seem to indicate that dialectograms are able to capture not just relatively obvious partisan differences but also more minute discursive differences between the two corpora at hand. Which is turn raises a bigger methodological question and opportunity: might the new method presented in this article be used to analyze and compare other corpora as well?

Overall differences in the characteristic use of words

To explore the over differences between the two corpora, let us return to Table 4.1. While inspecting the dialectograms of these words, we noticed that, across several plots, r/Democrats tends to focus on electoral entities and processes, whereas r/Republican is relatively more characterized by (contentious) topics of politics[10].

To examine this more closely we aggregate the overall differences between the characteristic uses in the two corpora, by counting how often the mean projection of each word across all offsets is greater than 0.2 and less than -0.2. This is equivalent to words that most often project heavily towards the characteristic use in r/Republican (upper right quadrant) or in r/Democrats (lower left quadrant). We then subtract the two counts, focusing on words that tend to project heavily towards the characteristic use in either corpus, but not both (such as the partisan associations identified above).

Table 4.2 shows the 30 words with the highest and the lowest values, corresponding to those words that are used most characteristically in r/Republican and r/Democrats, respectively. As can be seen, the result substantiates our earlier findings: while words with an electoral (e.g. *election*, *voter*, *lose*, *seat*), legislative (*policy*, *majority*, *governor*), and judicial meaning (*court*, *judge*, *scotus*) are used most frequently in r/Democrats, the most characteristic word usage in r/Republican conversely revolves around words related to gender/women (*marriage*, *fetus*), immigration (*immigrant*, *immigration*), healthcare (*disease*, *vaccinate*, *covid*), policing (*police*, *officer*), macroeconomics (*spending*, *budget*, *dollar*) as well as notions of *freedom*, *government* and *corrupt*. Consider, as an illustration of this pattern, the word *suppression* (Figure 4.7).

We see how words associated with legislative and voter suppression tend to be salient in r/Democrats, while r/Republican is more occupied with the suppression of speech. Specifically, whereas *legislator*, *gerrymandering* and *senate* project heavily towards r/Democrats's characteristic use of suppression, the characteristic use of suppression in r/Republican is conversely more similar to *speech*, *censor* and *freedom*. This indicates a stark discursive difference in the use of the term *suppression* when it comes to both who

[10]This difference in emphasis is in line with the subreddits' self-description; while r/Republican is presented as "a partisan subreddit [..] a place for Republicans to discuss issues with other Republicans", r/Democrats is conversely introduced as a subreddit that "offers daily news updates, policy analysis, links, and opportunities to participate in the political process. We are here to get Democrats elected up and down the ballot." (r/Republican, 2022; r/Democrats, 2022)

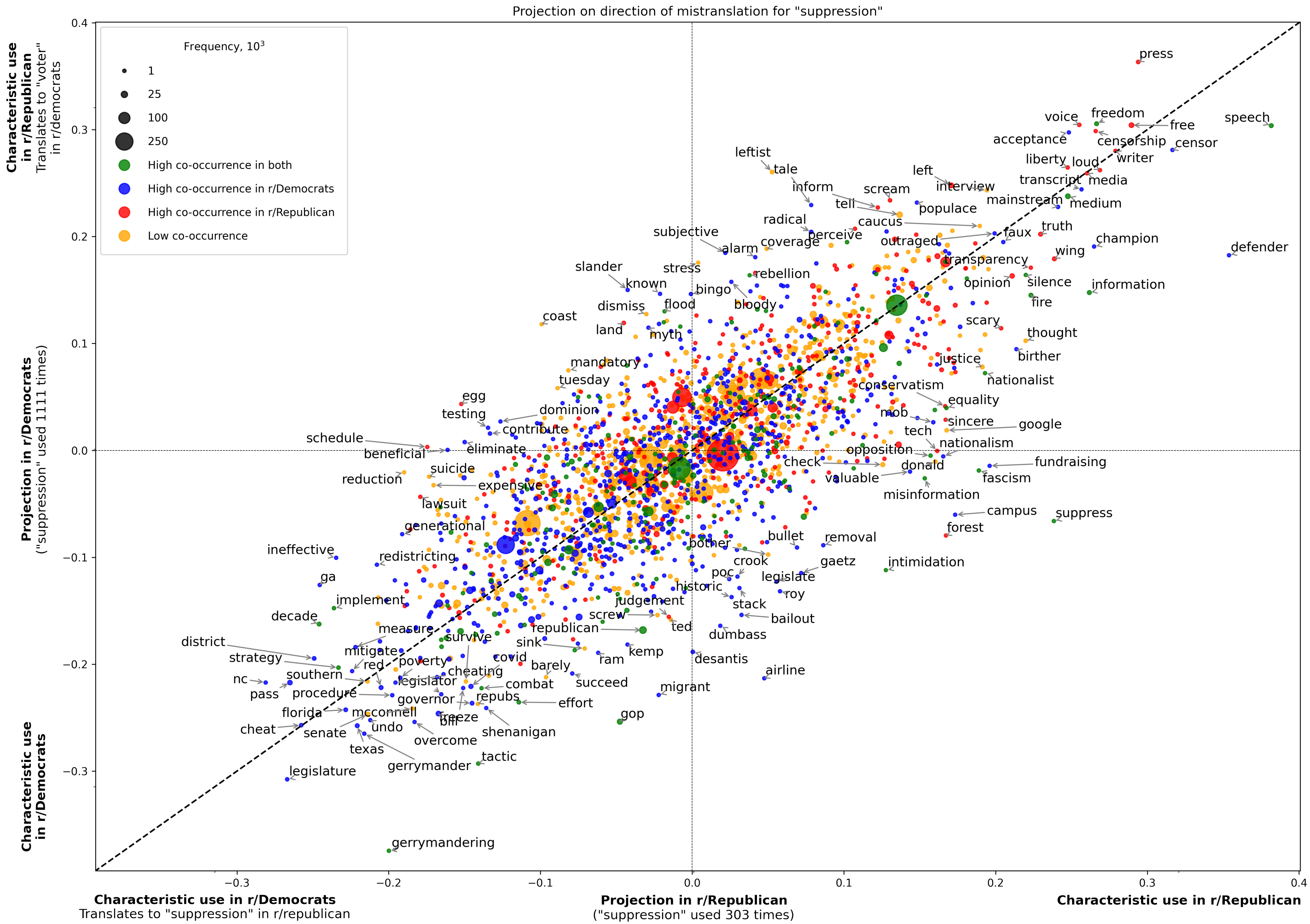


Figure 4.7: **Projection of words on the offset for the embeddings of *suppression*.** The dispersion of words along the x-axis might appear greater than along the y-axis, as the figure has been stretched to the page. Words are coloured according to their co-occurrence with *suppress*; see Eq. 4.2 for the definition of high co-occurrence.

Most common words in the characteristic use of r/Republican	Most common words in the characteristic use of r/Democrats
company	supporter
illegal	comment
spending	court
marriage	bernie
subreddit	loan
reddit	judge
woman	r
kasich	day
police	democratic
biden	change
covid	hillary
study	end
immigration	sanders
corrupt	seat
government	his
immigrant	sander
cruz	stop
dollar	election
law	propaganda
budget	he
ask	majority
paul	volunteer
romney	lose
fetus	expand
vaccinate	voter
officer	scotus
city	gerrymander
serious	governor
disease	policy
freedom	white

Table 4.2: **Most descriptive words.** Words that most unanimously project heavily towards the characteristic use in r/Republican (mean projection greater than 0.2) or in r/Democrats (mean projection less than -0.2).

and what is the object of suppression. Indeed, it would not be too far-fetched to make the claim that, when the members of the two Reddit fora we have here investigated use this word, they are talking about entirely different concepts all together.

Needless to say, the potential implications of this finding are as remarkable as they as chilling for everyone who is concerned about the increasingly polarization of US political discourse and indeed politics and society more generally.

4.3 Discussion and Conclusions

The word embedding method allows for the study of how words are used within and between corpora. By encoding the totality of relations between all pairs of words, this method can be used to map distinctions of discourse into the structure of the geometric spaces (Mikolov, Sutskever, *et al.*, 2013; Kozlowski *et al.*, 2019; Grand *et al.*, 2022). Mapping the full range of word associations, however, makes it hard to discover fine-grained differences in word usage of social scientific interest - e.g. how Republican and Democratic supporters write differently about the current President of the United States, Joe Biden.

Here, we have presented a solution to this challenge. By focusing on the relative use of words within corpora, we show how comparing projections along the direction of difference in the embedding space captures the most characteristic differences between language communities, no matter how minuscule this difference might be in quantitative terms. Our approach also offers a novel way to measure differences in word use, which overcomes the tendency for existing cosine distance and mistranslation approaches to pick out low-frequent and polysemous words. By measuring the degree to which the characteristic uses in the dialectogram are also unique to each community, our sense separation measure identifies words that are used most distinctly by the communities, thereby allowing for an unsupervised exploration of the discursive differences between the two communities. First, measures of the degree to which words are embedded differently can be used to *detect* words that the two communities use differently. Second, by constructing dialectogram of these words, we can *map* what these discursive differences consist of. While we have applied this approach to two contemporaneous political corpora, the same approach could also be used to compare other types of corpora or to explore how a discourse evolves over time. Further, while we have compared two language communities, the approach could be extended to compare multiple communities, by aligning the multiple embeddings into a single, shared space.

Our method has its limitations. First, while we interpret dialectograms like Figure 4.5 to show out-group discourse, what we actually observe is the word saliency among out-group politicians. Technically, this could include cases of denying the application of the derogatory words at hand to out-group members - although if so, one would expect a word like 'not' to project in similar ways. Still, it is not completely evident to what extent computational maps of digital discourse support an interpretation of the underlying meaning of this discourse (for a discussion, see Lee and Martin, 2015). Also, as with all found data, one should be careful when generalizing (Salganik, 2018). Republican and Democratic Reddit users probably differ systematically from their supporters in the general population. To further validate our new method and measure, further research could assess the degree to which survey-based approaches result in similar associations as those identified in the unsupervised examination of observational data (in the case of stance detection, see Joseph *et al.*, 2021).

We also encountered three technical issues. First, many social media platforms use partitions such as subreddits, but these do not necessarily correspond to homogeneous strands of discourse. We mitigated against this by excluding users who were fairly active in both corpora, but one could also attempt to computationally partition the documents within a single corpus based on the distribution of words within it, e.g. using higher-order matrix factorization of word-word-author matrices[11]. Second, exploring the development of discursive alignment over time results is an obvious extension. We attempted to do so,

[11] Unlike topic modeling, which attempts to split documents into separate topics, this would rather partition the documents into two (or more) parts, that each speaks to the same set of topics, but in different ways

by converting our static GloVe embedding to contextual representations of every word (token) in every time-stamped comment (Khodak *et al.*, 2018; Rodriguez *et al.*, 2022). This approach, however, amplified the correlation between cosine distance and frequency suggesting the unlikely conclusion that the later periods with more activity saw less difference in use. Finally, as discussed in appendix 6.3, converting contextual representations to static embedding resulted in quite rigid centroids, as measured by the perfect self-translation between the corpora, even with masked language modeling. While we believe contextual representations provide a promising path for this type of analysis and could provide a way of studying variation over time, the implications of the pre-training corpus and architecture of language modeling are still unclear.

Bibliography

Antoniak, Maria and David Mimno (Dec. 2018). „Evaluating the Stability of Embedding-based Word Similarities“. en. In: *Transactions of the Association for Computational Linguistics* 6. ZSCC: 0000085, pp. 107–119.

Arora, Sanjeev, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski (Dec. 2018). „Linear Algebraic Structure of Word Senses, with Applications to Polysemy“. en. In: *Transactions of the Association for Computational Linguistics* 6, pp. 483–495.

Arseniev-Koehler, Alina and Jacob G. Foster (June 2020). „Machine learning as a model for cultural learning: Teaching an algorithm what it means to be fat“. en. In: *arXiv:2003.12133 [cs]*. ZSCC: 0000004 arXiv: 2003.12133.

Artetxe, Mikel, Gorka Labaka, and Eneko Agirre (2016). „Learning principled bilingual mappings of word embeddings while preserving monolingual invariance“. en. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, pp. 2289–2294.

Azarbonyad, Hosein, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps (Nov. 2017). *Words are Malleable: Computing Semantic Shifts in Political and Media Discourse*. en. arXiv:1711.05603 [cs].

Balbi, Simona and Michelangelo Misuraca (2006). „Rotated canonical correlation analysis for multilingual corpora“. en. In: *In JADT*, pp. 99–106.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai (June 2016). *Quantifying and Reducing Stereotypes in Word Embeddings*. en. arXiv:1606.06121 [cs, stat].

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan (Apr. 2017). „Semantics derived automatically from language corpora contain human-like biases“. en. In: *Science* 356.6334, pp. 183–186.

Dubossarsky, Haim, Daphna Weinshall, and Eitan Grossman (Sept. 2017). „Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models.“ In: *Conference on Empirical Methods in Natural Language Processing*. MAG ID: 2759556264 S2ID: ab370db3fe0ee0cb272f13ebbb32894de90bec8d, pp. 1136–1145.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (Apr. 2018). „Word embeddings quantify 100 years of gender and ethnic stereotypes“. en. In: *Proceedings of the National Academy of Sciences* 115.16, E3635–E3644.

Grand, Gabriel, Idan Asher Blank, Francisco Pereira, and Evelina Fedorenko (Apr. 2022). „Semantic projection recovers rich human knowledge of multiple object features from word embeddings“. en. In: *Nature Human Behaviour* 6.7, pp. 975–987.

Hamilton, William L, Jure Leskovec, and Dan Jurafsky (2016). „Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change“. en. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1489–1501.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes (Sept. 2012). „Affect, Not Ideology“. en. In: *Public Opinion Quarterly* 76.3, pp. 405–431.

Jones, Jason, Mohammad Amin, Jessica Kim, and Steven Skiena (2020). „Stereotypical Gender Associations in Language Have Decreased Over Time“. en. In: *Sociological Science* 7. ZSCC: 0000008, pp. 1–35.

Joseph, Kenneth, Sarah Shugars, Ryan Gallagher, Jon Green, Alexi Quintana Mathé, Zijian An, and David Lazer (Sept. 2021). *(Mis)alignment Between Stance Expressed in Social Media Data and Public Opinion Surveys*. arXiv:2109.01762 [cs].

Khodak, Mikhail, Nikunj Saunshi, Yingyu Liang, Tengyu Ma, Brandon Stewart, and Sanjeev Arora (May 2018). „A La Carte Embedding: Cheap but Effective Induction of Semantic Feature Vectors“. en. In: *arXiv:1805.05388 [cs]*. ZSCC: 0000054 arXiv: 1805.05388.

KhudaBukhsh, Ashiqur R, Rupak Sarkar, Mark S Kamlet, and Tom Mitchell (2021). „We Don’t Speak the Same Language: Interpreting Polarization through Machine Translation“. en. In: *Association for the Advancement of Artificial Intelligence* AAAI-21, p. 9.

Kozlowski, Austin C., Matt Taddy, and James A. Evans (Oct. 2019). „The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings“. en. In: *American Sociological Review* 84.5, pp. 905–949.

Lee, Monica and John Levi Martin (Feb. 2015). „Coding, counting and cultural cartography“. en. In: *American Journal of Cultural Sociology* 3.1, pp. 1–33.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (July 2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. en. arXiv:1907.11692 [cs].

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (Sept. 2013). „Efficient Estimation of Word Representations in Vector Space“. en. In: *arXiv:1301.3781 [cs]*. ZSCC: 0020188 arXiv: 1301.3781.

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever (Sept. 2013). „Exploiting Similarities among Languages for Machine Translation“. en. In: *arXiv:1309.4168 [cs]*. arXiv: 1309.4168.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (Oct. 2013). „Distributed Representations of Words and Phrases and their Compositionality“. en. In: *arXiv:1310.4546 [cs, stat]*. ZSCC: 0025348 arXiv: 1310.4546.

Milbauer, Jeremiah, Adarsh Mathew, and James Evans (2021). „Aligning Multidimensional Worldviews and Discovering Ideological Differences“. en. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. ZSCC: 0000000, pp. 4832–4845.

Mohr, John W., Robin Wagner-Pacifici, and Ronald L. Breiger (2015). „Toward a computational hermeneutics“. In: *Big Data & Society* 2.2.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (2014). „Glove: Global Vectors for Word Representation“. en. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ZSCC: 0022883. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543.

r/Democrats (2022). *Pushshift*.

r/Republican (2022). *Pushshift*.

Rathje, Steve, Jay J. Van Bavel, and Sander van der Linden (June 2021). „Out-group animosity drives engagement on social media“. en. In: *Proceedings of the National Academy of Sciences* 118.26, e2024292118.

Rodriguez, Pedro L, Arthur Spirling, and Brandon M Stewart (2022). *Embedding Regressions: Models for Context-Specic Description and Inference*. en. ZSCC: 0000000.

Salganik, Matthew J. (2018). *Bit by bit: social research in the digital age*. OCLC: on1012406622. Princeton: Princeton University Press.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf (Feb. 2020). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. en. arXiv:1910.01108 [cs].

Schlechtweg, Dominik, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi (Aug. 2020). *SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection*. en. arXiv:2007.11464 [cs].

Schönemann, Peter H. (Mar. 1966). „A generalized solution of the orthogonal procrustes problem“. en. In: *Psychometrika* 31.1, pp. 1–10.

Wendlandt, Laura, Jonathan K. Kummerfeld, and Rada Mihalcea (2018). „Factors Influencing the Surprising Instability of Word Embeddings“. en. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. ZSCC: 0000067 arXiv: 1804.09692, pp. 2092–2102.

William, Press (May 2011). *Canonical Correlation Clarified by Singular Value Decomposition*.

5

Fixing Fieldnotes: Developing and Testing a Digital Tool for the Collection, Processing, and Analysis of Ethnographic Data

By Sofie L. Astrupgaard, August Lohse, Emilie M. Gregersen, Jonathan H. Salka, Kristoffer Albris, and Morten A. Pedersen[1]

Abstract

Ethnographic fieldnotes can contain richer and more thorough descriptions of social phenomena compared to other data sources. Their open-ended and flexible character makes them especially useful in explorative research. However, fieldnotes are typically highly unstructured and personalized by individual researchers, which make them harder to use as a method for data collection in collaborative and mixed methods research. More precisely, the unstructured nature of ethnographic fieldnotes presents three distinct challenges: 1) Organizability it can be difficult to search and sort fieldnotes and thus to get an overview of them, 2) Integrability it is difficult to meaningfully integrate fieldnotes with other more quantitative data types such as more such as surveys or geospatial data, and 3) Computational Processability it is hard to process and analyze fieldnotes with computational methods such as topic models and network analysis. To solve these three challenges, we present a new digital tool, for the systematic collection, processing, and analysis of ethnographic fieldnotes. The tool is developed and tested as part of an interdisciplinary mixed methods pilot study on attention dynamics at a political festival in Denmark. Through case examples from this study, we show how adopting this new digital tool allowed our team to overcome the three aforementioned challenges of fieldnotes, while retaining the flexible and explorative character of ethnographic research, which is a key strength of ethnographic fieldwork.

5.1 Introduction

Fieldnotes are qualitative social scientific data whose depth and richness can be incomparable to other data forms. Their open-ended and flexible nature makes them apt for explorative research and grounded theory (FitzGerald and Mills, 2022; Phillippi and Lauderdale, 2018), as well as particularly useful for studying settings marked by radical unpredictability, contingency, and uncertainty, such natural disasters and political

[1]This paper was co-first authored by Sofie Læbo Astrupgaard and August Lohse.

conflicts (e.g. Gill *et al.*, 2015). Ethnographic field data can play a vital role in "big data" research. On the one hand, fieldnotes can thus serve as "ground truth" for big data and black-boxed computational research (Blok *et al.*, 2017; Krieg *et al.*, 2017). On the other hand, fieldnotes also contain highly spatially and temporally granular information, and making them amenable for computational analysis could help social scientists to discover new patterns and produce ground-breaking results (Bjerre-Nielsen and Glavind, 2022; Blok and Pedersen, 2014; Munk, 2019; Nelson, 2020).

Traditionally, fieldnotes have been handwritten in situ into notebooks and over recent decades increasingly typed into text-processing programs. Ethnographers typically develop their own idiosyncratic and not always transparent ways of writing, storing, and accessing their fieldnotes (e.g., Abramson *et al.* 2018, Sanjek 1990, see also Albris *et al.* 2021). This makes it hard to combine and compare different researchers' fieldnotes, and to integrate them with other kinds of data, including not least quantitative ones. For the same reason, it is still uncommon for ethnographers to work in teams, especially interdisciplinary collaborations[2]. A further drawback arising from conventional ways of producing and handling fieldnotes is the fact that it is extremely time-consuming. For one thing, the process of typing (and sometimes re-typing) fieldnotes into text-processing software such as Word or commercial qualitative analysis software such as NVivo is very cumbersome. What is more, the subsequent coding of this data into themes in accordance with research questions will often involve many hours of often tedious work. Indeed, researchers are often forced to limit their analysis to chunks of their raw data based on their own recollection of best material and cases.

In this paper, we explore the advantages to be gained from logging and structuring ethnographic fieldnotes into a formalized digital corpus already from the onset of collecting them in the field, which makes it easier to sort, search, and analyze such ethnographic data afterwards. More specifically, we highlight three key challenges but therefore also opportunities pertaining to how fieldnotes are commonly collected, processed, and analyzed: 1) *Organizability* it can be difficult to search and sort fieldnotes and thus to get an overview of them, 2) *Integrability* it is difficult to meaningfully integrate fieldnotes with other more quantitative data types such as surveys or geospatial data, and 3) *Computational Processability* it is hard to process and analyze fieldnotes with computational methods such as topic models and network analysis. In what follows, we propose a solution to each of these three challenges in the form of a digital tool that allows for the systematic collection, processing, and analysis of ethnographic fieldnotes.

The EthnoPlatform, as we have called our pilot version, is a digital data architecture enabling fieldworkers to collect ethnographic notes in a structured format (see also Astrupgaard *et al.*, Forthcoming). The idea to increasingly structure ethnographic data collection is far from new, and we draw inspiration from a range of scholars who have also seen this potential (e.g., Nippert-Eng 2015, pp. 36-43, see also Bernard 2006, pp. 398-405, Lyon 1999). Via examples from a mixed methods pilot study of political attention in Denmark, we show how adopting our tool for structuring and digitizing fieldnotes allowed our team to overcome the three challenges of traditional practices around taking fieldnotes, while retaining several key strengths of

[2] Several articles and books have come out over recent years document the potentials and challenges of conducting ethnographic research in teams, where each member collect fieldnotes in the traditional, low-tech manner (see for example: Boyer and Marcus, 2021; Bunkenborg *et al.*, 2022; Hastrup *et al.*, 2016; Korsby and Stavrianakis, 2018)

qualitative social scientific methods. More specifically, we argue that by structuring fieldnotes in a tabular format, the EthnoPlatform 1) ensures organizability of fieldnotes by making them sortable and searchable, 2) enhances integrability with other data types such as geospatial data, and 3) improves computational processability by making fieldnotes amenable for computational processing and analyses via social data science methods such as network analysis and methods for automated text analysis.

What follows below is divided into two parts. In the first, we begin by providing a critical overview of existing digital tools for ethnographic data collection, processing, and analysis. We then introduce the EthnoPlatform by describing its main functionalities as a digital tool and our motivation for creating an infrastructure that compels and nudges ethnographers to settle upon a shared data structure as well as identifying key *tags* for their fieldnotes prior to embarking upon the fieldwork. Then, in the second part, we first introduce the interdisciplinary research project on political attention dynamics that served as the context for our pilot study. Via selected examples from this pilot, we then show how using this new digital tool made it possible to process and analyze the collected ethnographic fieldnotes in new ways that overcome the three aforementioned limitations of traditional fieldnote production and handling. We conclude by reflecting on how the EthnoPlatform makes ethnographic research more efficient and systematic and how it can contribute to front-line research combining mixed methods and data science (Grigoropoulou and Small, 2022; Marda and Narayan, 2021)

5.1.1 Background

Hopes and aspirations that digital tools would revolutionize qualitative research, including ethnography, have existed for quite some time (e.g. Hymes 1965, see also Kemper *et al.* 1992). The first applications and platforms very much reflected the state of software development in the early periods of the internet. The earliest examples such as ETHNOGRAPH (Seidel and Clark, 1984) or Anthropac (Borgatti, 1989) by now look somewhat antiquated, although they were on the forefront of the development at their launch. Since then, the usability, design, and power of software applications for qualitative research have vastly improved. A key distinction to make when surveying the landscape of qualitative research software, including commercial and noncommercial solutions, is whether the software is to be used in the data collection phase, in the data analysis phase, across these phases, as well as in any interlocking "sub-phases." Most qualitative researchers are familiar with research software in the data analysis phase, which is known as Computer Assisted Qualitative Data Analysis Software, often referred to in abbreviated form as "CAQDAS." Most well-known examples are NVivo (Lumivero, 2023), Dedoose (SocioCultural Research Consultants, 2021), ATLAS.ti (ATLAS.ti, Scientific Software Development GmbH, 2023), TAMZ Analyzer (Weinstein, 2006), or MAXQDA (VERBI Software, 2021), to name but a few. While all these platforms are different, what unifies them is the ability to process and analyze qualitative datasets through coding, indexing, and basic visualizations. Some of these platforms, like Dedoose, are cloud-based, to some extent oriented to teamwork, and have over time become quite dynamic and flexible to use.

However, existing digital platforms for qualitative data analysis, ranging from commercial software such as Nvivo to Dedoose have several limitations (Fielding 2012, pp. 126-133, Paredes *et al.* 2017, p. 1562). Apart from the fact that much of this software is proprietary and thus difficult to afford for individuals or institutions with limited resources, their inflexibility and closed environmental design makes them cumbersome to use for ethnographic research. In particular, the inbuilt and often encapsulated structure for the labeling of data, which is a common feature of most existing platforms, puts constraints on the iterative nature of grounded theory analysis, which requires that researchers can continuously re-type and re-label their data as their analysis progresses (Nelson, 2021). Moreover, applications such as Dedoose and Nvivo do not incorporate in situ data collection in a way that is aligned with the affordances that smartphones and other handheld devices can offer ethnographers in the field, as they operate on computers only. Nor is Dedoose or Nvivo designed to be interoperable with backend programming languages such as Python or R, although these applications have indeed expanded their usage with, for instance, gathering data from social media platforms.

Examples of software aimed at the data collection phase of the ethnographic research process with a specific focus on fieldnotes are sparse. Many ethnographers have promoted the use of generic note taking software such as Evernote (Wang, 2012). However, while such applications can be used on handheld devices, is usable for teams, and can even do simple tagging of themes within notes, they are not designed for research, and therefore lack further functionalities and design capabilities. Crowd sourced data collection or "mobile ethnography" represents another group of applications. A case in point is the Indeemo mobile ethnography app (Indeemo, 2018), which is designed explicitly for handheld devices, aimed at letting research participants contribute to the data collection through written interactions, photos, and videos. It is, however, not a software platform optimized for the writing of fieldnotes and analysis of these across a team of researchers. This is also the case with similar platforms and apps such as EthnoAlly (Favero and Theunissen, 2018), the EthOS platform (EthOS, 2023), or QualMobile (Sago, 2023), as many other competing alternatives on the market.

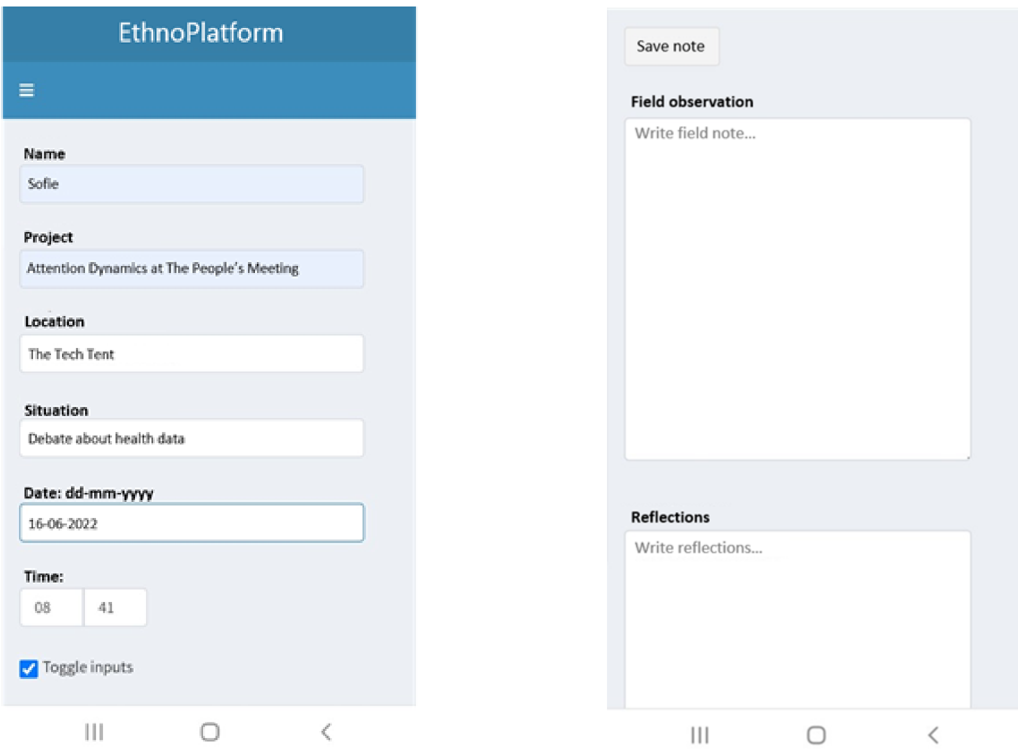
To the best of our knowledge then, there is currently no tool, software, nor platform (commercial or non-profit) that can facilitate the writing and storage of fieldnotes on the move (i.e., on handheld devices), while being easy to use and adjustable to the wishes of the individual or group of ethnographers who are using it and allow computational processing of fieldnotes. This motivated us to develop our own digital tool, and in the following section, we describe the tool it is built on and how it works.

5.2 A Digital Data Architecture for Ethnographic Fieldnotes

Our overarching ambition has been to develop a digital tool that helps to overcome the three limitations pertaining to existing ways of producing and handling fieldnotes, namely, organizability, integrability, and computational processability. Unlike existing digital tools for the handling of qualitative data, it had to be accessible for ethnographers in the field through a handheld device, usable for collaborative data collection and storage, and allowing for export of fieldnotes in computationally processable file-formats. Crucially, all these objectives should be met without restricting the flexibility and open-endedness of the ethnographic method and explorative and abductive social scientific analyses too much. As we are going to describe in detail in the following pages, the solution we developed involves a *tagging approach* as the basic principle undergirding the data infrastructure of our digital tool. This, we shall argue, provides a viable path out of each of the three above challenges.

5.2.1 A Tagging Approach

The tagging approach implies that researchers develop a set of tags before entering the field. These Tags can be closed categories pertaining to each fieldnote, or they can be open categories that allow for more elaborate accounts, including a virtually mandatory tag for field observations. An example of a fieldnote template from our own case study, which we will describe in more depth later, can be seen in Figures 5.1a and 5.1b, which show the interface of the EthnoPlatform. Here, we used 8 tags in total. When initiating a new fieldnote, the ethnographer is met with a template of the pre-defined tags. In our case, these consisted of 6 close-ended tags: "Name" (of ethnographer), "Project", "Location", "Situation", "Date", and "Time" (see Figure 5.1a) ensuring that each note contains the same contextual information. Beneath the close-ended tags, what we refer to as meta data, there are two additional open-ended tags with larger text fields; "Field observation" and "Reflections" (see Figure 5.1b). "Field observation" is for descriptive fieldnotes pertaining to a particular setting or situation playing out, and the second field is for adding analytical or methodological reflections.



(a) Interface of the EthnoPlatform. Close-ended tags.

(b) Interface of the EthnoPlatform. Open-ended tags.

Figure 5.1: Interfaces of the EthnoPlatform.

In combination, the tags make out a fieldnote template that can be filled out by any member of a team. Every time an ethnographer writes a fieldnote, she makes sure to retrieve information that correspond to the pre-defined tags, thereby, the template ensures that all fieldnotes contain the same type of information, making them more aligned through the tags.[3] We want to emphasize that the tags exemplified in the template in Figures 5.1a and 5.1b were chosen because they made sense for our particular study at a Politics festival. In other studies, another set of tags might be more useful such as "Keywords", "Names (or pseudonyms) of interlocutors present", or "Summary of fieldnote".[4] In the coming section, about how our tool enhances organizability, we elaborate how the template tags can be used to sort and retrieve specific fieldnotes. By using such templates, the ethnographers are forced to follow a structure, and the fieldnotes then become less individualistic and idiosyncratic, as the single ethnographer thus becomes more detached from her notes. Moreover, if there are multiple researchers working on a project, they can better read and understand each other's fieldnotes.

5.2.2 Developing the EthnoPlatform

Certainly, it is the possibility to collect fieldnotes in a common structure via tags that makes up the key feature of our tool, The EthnoPlatform. We have developed two test-versions[5] that are based on the tagging approach, but in the first test-version, we used the already existing off-the shelf survey tool, SurveyXact. Here, the fieldnote template corresponded to questions in a survey that was distributed to ethnographers through a link that they could access on their devices and fill out whilst observing in the field. Based on lessons learned from the first version, the EthnoPlatform was refined, and we developed the next test version as a web application.[6] The fieldnote template was similar to the first version, though this time the fieldnote archive was more accessible in the field, making it easier to add and edit notes on the go. The interface of our second version is shown in Figures 5.1a and 5.1b. Here the ethnographer fills out information in the fields below each tag, and then, she clicks "Save note" to store it in a cloud-based archive. The ethnographer can read and edit the note in the field or later. Thus, if there has not been enough time

[3] For an example of single-researcher use of rigorous fieldnote templates, see Lyon, 1999 who also utilizes digital infrastructures in organizing and sharing fieldnotes. For team-based use in an interdisciplinary setting, see Astrupgaard *et al.*, Forthcoming.

[4] As one reviewer rightfully pointed out, having short abstracts or summaries as part of the template, can be utilized for training machine learning models, for example, LLMs for identifying and linking specific fieldnotes. Furthermore, the use of a header as the first tag with keywords and an indicator of whether there is sensitive information in the fieldnote can ease anonymization and protection of interlocutors.

[5] We are currently in the final stages of developing an app-based version of the EthnoPlatform called EthNote (Gregersen *et al.*, Forthcoming"), with funding received from the Carlsberg Foundation. The app will be open-source and is already accessible at: `https://ethnote.org.`

[6] The EthnoPlatform was developed using the Shiny framework, which is an open-source library written for the programming language R allowing for the creation of simple interactive web-apps that are very customizable while also easy to deploy. The platform was hosted on a secure Digital Ocean droplet (server) that fulfilled GDPR requirements for safe storage. All ethnographic data was written to a MongoDB database on this server and could later be accessed. The source code for the beta version of the EthnoPlatform can be made available upon request. Our app-based version of the EthnoPlatform, EthNote, supports offline fieldnotes by storing the data locally on the given device until the user is connected to the internet. The user can then click on a sync button in which case the app automatically erases all locally stored data and uploads it to the cloud. While many of the functionalities of the app such as accessing team members' fieldnotes or uploading different media require an internet connection, the offline fieldnote section does allow the app to be used in areas where it can be difficult to maintain or even access a stable connection.

to type in all information at once, it is possible to log on to the EthnoPlatform and elaborate or extend fieldnotes later on if needed.

5.2.3 Processing Fieldnotes in a Tabular Format

Because the ethnographic fieldnotes are structured via tags, they can easily be exported from the EthnoPlatform in a tabular format such as a CSV-file. In this data format, the tags are columns and fieldnotes are rows. The tabular format makes the collection of fieldnotes appear a lot like a quantitative data set, where the tags correspond to variables and fieldnotes to observations. Then, the fieldnotes can to a large extent be processed as quantitative data. If an ethnographer uses many tags, then the fieldnote template will provide better possibilities for quantitative analysis through aggregation and clustering of fieldnotes by tags. However, the open-endedness of a research project is reduced if there are many tags as it will make ethnographers less prone to delve into unexpected questions and dimensions emerging along the way in the field. For a project with an open-ended focus, where the ethnographer prefers to retain the traditional focus on writing and editing fieldnotes in a more idiosyncratic and flexible manner, she can decide to merely use tags such as "Text", "Place", and "Date". This will make the data collection less structured, reducing the possibilities of overcoming the aforementioned limitations pertaining to traditional ways of doing fieldnotes. Nevertheless, even in such cases, ethnographic researchers could still benefit from digitizing and organizing their fieldnotes via time and location metadata by utilizing the EthnoPlatform's efficiency as a handheld device for instantly logging fieldnotes into an online editable archive.

Which and how many tags to include should ideally be settled upon before entering the field. However, it is important to emphasize that when using the EthnoPlatform, the number of tags for fieldnotes does not have to be set in stone before the data collection begins. The fieldnote template can be edited after the ethnographic data collection has begun, and in case the ethnographer encounters new paths emerging in the field that make the pre-defined tags seem insufficient, then she can add or remove tags. Archived fieldnotes can be gone through and edited to include the revised tags if it seems sensible. The framework we developed generates a protocol of the changes made in the tag structure during the period of data collection, which ensures that potential changing of tags along the way is transparent.[7] Though the main purpose of our tool is to digitize and structure fieldnotes, the possibility of changing tags ensures a high degree of flexibility during data collection. Nevertheless, any revision of tag structure should preferably take place in the beginning of the ethnographic data collection to avoid a high degree of manual post-tagging and irregularities in the structure of the fieldnotes.

[7] When exporting the data in csv-format, all tags (original and added) will be present as columns. Fieldnotes that were created before a given tag, such as "location," was added to the template will display empty cells. More technically, each missing data point will have an empty string represented by two quotation marks (""). In this way, the user can keep track of added and discontinued tags in keeping with the tabular format. EthNote logs changes to templates automatically, which can be exported as an overview showing changes in the different template version. Additionally, the app has a template version count. When creating a new fieldnote, it will automatically include the current template version count thus providing a more low-tech transparency of tag changes.

5.2.4 Fixing Fieldnotes

There are several ways in which our tool improves organizability of fieldnotes: The tabular format of its data infrastructure means that fieldnotes are more easily sorted or grouped based on conditions through tags (e.g., "date" or "ethnographer"). This means that the fieldnotes can better be organized by one's own preferences or preferred tags which provide the ethnographer with an overview of the collected fieldnotes. Correspondingly, the integrability of the fieldnotes is also enhanced. If one or more tags of the fieldnotes matches features of another data set, then it is possible to integrate the data sets. For instance, having a "Time"-tag allows the ethnographer to combine relevant quantitative data that also has a time feature. This could be different kinds of data dependent on the specific project, and one could imagine integrating information about anything from weather conditions to social media posts with the ethnographic fieldnotes.

But that is not all. As we are going to see below, the digitized and structured format of the EthnoPlatform's digital data architecture also makes computational processing of fieldnotes easier. This means that we can use methods and techniques from social data science such as automated text analyses to either test pre-existing hypotheses or to discover new patterns in the fieldnotes.

5.3 Exploring Fieldnotes From an Interdisciplinary Case Study

In this second part of the article, we introduce the ethnographic research context in which the pilot test and development of the EthnoPlatform took place and we discuss how this digital tool enabled us to explore our ethnographic fieldnotes by using computational methods. Accordingly, what follows is structured around the three aforementioned limitations of traditional fieldnotes, organizability, integrability, and processability. First, we show how the structured format of our fieldnotes allow us to easily get an overview of the data using summarizing statistics as well as effortlessly search and sort through them. Thus, we can easily get an overview of our ethnographic fieldnotes. Then, we show how our fieldnotes can be integrated with geospatial data which allow us to better explore spatial dimensions of the festival site through our fieldnotes. Lastly, we show how our fieldnotes can be analyzed through more advanced means of computational processing such as topic modelling and network analysis, as we present a topical mapping to visualize thematic relations in the fieldnote corpus.

5.3.1 Collecting Ethnographic Fieldnotes at a Danish Politics Festival

We used the EthnoPlatform to collect fieldnotes as part of a case study at the Danish politics festival, The People's Meeting (In Danish: "Folkemødet"). The People's Meeting aims to facilitate a democratic dialogue between citizens and decision-makers by hosting numerous topical debates and political speeches organized by public and private stakeholders. Since its launch in 2011, it has grown to become Denmark's largest politics festival, and in 2022, the festival hosted 2500 events in its four-day duration with around

30.000 daily visitors (Folkemødet, 2022). Many events occur simultaneously causing event organizers to use different means to try and attract the attention of visitors who can freely choose which events to partake in. The People's Meeting thus served as an interesting case for us to study different aspects of attention dynamics and behavior as well as for testing our new tool.

The pilot was conducted as two studies, 2021 and 2022, with different analytical objectives but a shared focus on data pertaining to political attention.[8] In 2021, we experimented with new ways of collecting ethnographic fieldnotes collaboratively to use these in combination with other data types. We also sought to produce fieldnotes pertaining to similar situations and events occurring in different places around the festival site. Given the somewhat chaotic nature of such a large-scale event, we were compelled to reconsider the traditional practice of fieldnote collection and instead introduce a more structured format leading to the development of the first version of the EthnoPlatform. In 2022, we returned to the festival for the second part of the study with a second version of the EthnoPlatform, this time as a web application. Using the EthnoPlatform alongside detailed observation guides, we found that this format indeed could produce organizable, integrable, and computationally processable fieldnotes. The EthnoPlatform thus allowed us to accumulate similarly structured, ethnographic fieldnotes from many locations at the festival square at the same time which were instantly digitized and stored safely from the field in both parts of the case study.[9] We have now presented our own experiences with the tool in the data collection phase. In the next section, we turn to how the structured fieldnotes can be analyzed, and here, the first step was to export the data as CSV-files to then process them computationally using the programming language, Python.

5.3.2 Organizability - Searching and Sorting Fieldnotes

The first stage in any data processing is to obtain an overview of one's data. One way to do this is via summarizing statistics as presented in Table 1. The tabular data format of the fieldnotes makes it easy to get an overview of basic information about the dataset such as the amount of fieldnotes collected, their average length, and the number of ethnographers involved. In our example, we use information from the tags "Name" and "Field observation." From Table 5.1, we can thus infer that with our first and second version of the platform, we collected 144 fieldnotes in total, with an average length of 181 words. Similarly, we see that there have been 16 different ethnographers writing the fieldnotes in total. Looking at the 2021 and the 2022 data separately, we see that most of our data was collected in 2021 where a higher number of fieldnotes were collected and these were also longer in average.

[8] The overarching research project on political attention was supported by the H2020 European Research Council (grant number 834540) as part of the project: "The Political Economy of Distraction in Digitized Denmark" (DISTRACT). The more methodological project focused on computational ethnographic methods, which included the pilot (the Ethnoplatform), as well as the development of the more professional version of the digital tool (EthNote) was funded by University of Copenhagen's Data-Plus Program and by an infrastructure grant from the Carlsberg Foundation.

[9] For a more elaborate discussion of a potentials and pitfalls regarding standardized collection of fieldnotes in a team-based and interdisciplinary setting, see Astrupgaard *et al.*, 2022 or Astrupgaard *et al.*, Forthcoming

It is also possible to produce summarizing statistics that may be of more analytical interest through word frequencies. We can, for example, pick out the most frequently used words. In Table 1, we see the most common words in the data seem to be related to the stage and audience, but also to behaviors such as walking and talking. We also see that in 2021 there was a lot more written about the audience and the stage, whereas the 2022 fieldnotes seem to be more concerned with groups of people and the general atmosphere.

	Full Corpora	**2021 Corpora**	**2022 Corpora**
Number of notes	144	90	54
Mean length (words)	181.17	215.21	124.43
Number of ethnographers	16	7	10
Most frequent words	*Stage, walk, woman, audience, say*	*Stage, woman, audience, walk, say*	*Group, walk, atmosphere, event, chat*
Number of notes including "*attention*"	48	39	9
Most frequent co-occurrence with "*attention*"	*Stage, direct, talk, audience, work*	*Stage, direct, work, talk, debate*	*Direct, group, stage, talk, follow*

Table 5.1: **Summary of Corpora**

Since we were interested in obtaining an overview of how many of the collected fieldnotes directly mentioned our primary research concept ("attention"), we count the notes containing the word "attention." We also extract the words that most frequently appear next to "attention" in a sentence (co-occurrence). Doing this, we find that approximately a third of the fieldnotes explicitly mention attention, and that most of these fieldnotes are in the data from 2021. The five most common words associated with "attention" are almost similar in 2021 and 2022, which indicates that our observations of attention dynamics in 2021 and 2022 share similar content. We assume that the differences in the content of notes in 2021 and 2022 is somewhat due to each year's observations being guided by different instructions given to ethnographers on what kind of attention-related behavior to look for. Indeed, these relatively basic descriptive statistics can thus both reveal overall patterns, but more importantly, they can help to guide our qualitative reading of the fieldnotes.

As a next step of our exploratory analysis, we choose to focus on the temporal dynamics in the data. We find all sentences in the text within the "Field observation"-tag of all fieldnotes that include the word "attention" and derivatives of the word. Then we sort these sentences using "Time" and the "Date"-tags, constructing a timeline indicating how observations about attention differ during the day. An extract is shown in Figure 5.2, where we have highlighted the word "attention" in each sentence, and we have added information from the "Situation"-tag of the fieldnote in question to get a sense of the context.

Reading the notes in this way allows us to analyze temporal development in attention among the visitors at the People's Meeting. This approach reveals that the attention at the political debates might be more intense in the morning and that there at all times appear to be several external distractors, attracting attention away from the debates (such as phones, noisy bins, and demonstrations). While there might be many causes of these findings, our re-reading of the temporally organized sentences with the word "attention" reveals a pattern that we could then examine more thoroughly through detailed qualitative readings of the 48 notes in full length containing the word and its derivatives.

Figure 5.2: Fieldnote timeline. This is a extract of the timeline of sentences in fieldnotes containing the word "attention" and the situation-tag of the fieldnote.

Having searchable and sortable fieldnotes is, thus, both a big practical and analytical advantage. Researchers can, for example, easily discover new patterns in their own or others' ethnographic data, by searching for specified keywords or set of keywords, and extract all fieldnotes or specific sentences, where those words or derivations of those words are used. This is obviously very useful for exploring, for example, how a specific concept is used differently in the total dataset, as it reduces the time the researchers must use on skimming through their notes to find what they are looking for and can direct their attention to parts of the fieldnote corpora where they would not otherwise have looked. Sorting the notes can also be an important tool for researchers, and it is a lot easier to do computationally. Researchers can, for example, sort their corpora by time and examine their relationship and conversations with interlocutors over time. One could also imagine sorting fieldnotes by where they were recorded, in order to understand why some areas, seem to be better described than others. This way one could examine how notes with specific keywords are dispersed spatially by extracting notes with the keyword and sorting them by where they were recorded. One could go even further and search up the fieldnotes concerning a specific interlocutor that contains a specific keyword and sort these by time in order to read the notes in a very analytically focused manner. Thus, searching and sorting through the fieldnotes allows us to find patterns across the fieldnotes that can serve as independent analytical insights and/or guide further analysis.

5.3.3 Integrability - Combining Fieldnotes with Other Data Types

Besides examining the temporal dimensions of attention at The People's Meeting, one can also use our digital tool to explore its spatial dimension. In this part of our exploratory data analysis, we integrate our digitized fieldnotes and geographical data from OpenStreetMap to create a fieldnote map, which can be used as point of departure for further explorative research. As part of this process, we use the "Location", "Time", "Field observation", and "Name" (same as researcher ID) tags to create an interactive spatiotemporal map which we show a snapshot of in Figure 5.3. Here, each point denotes a fieldnote from the 2022-study, showing the exact location of its production plotted onto a map of the festival site. The colors of the pins indicate the authoring ethnographer, meaning that we see where each field observation is conducted and by whom throughout the day. Accordingly, we can explore the interactive visualization, and by hovering our computer mouse over the pins, we see various information about each fieldnote such as the time of day and the researcher ID. We also added the most important words in that fieldnote, by measuring the words with the highest TF-IDF loading (Spärck Jones, 1972). TF-IDF loading is a very commonly used measure of word importance from the field of Natural Language Processing (NLP) that can help find the words that sets a text apart from the rest of the corpora. Here, the technique can help us summarize the fieldnote contents, so we can explore them spatially as well.

Figure 5.3 allows us to see where the ethnographers in the team observed attention dynamics. In the specific case at hand, we see that data from the main stage zone (top left) appear to revolve around debates. The two notes highlighted in this zone in Figure 5.3, for example, pertain to a debate, where the audience's attention was distracted by noise from neighboring events (in one tent, the host was a stand-comedian making people laugh, and in another tent, there was a debate with an ambassador where the main focus of the audience seemed to be on the free tapas being served). Conversely, we also see that the fieldnotes from the harbor zone (bottom right) tend to focus more on how groups of people move through the zone and their interactions. Thus, the two highlighted notes illustrate that while some of the observed people in the zone stand still in small clusters (apparently discussing particular debates), other groups traverse through the harbor while chatting, in one instance buying a flower from the local florist. While a more thorough investigation would be required to confirm and substantiate this, this preliminary analysis indicates that activities in this zone are less centered around political debates than around the main stage.

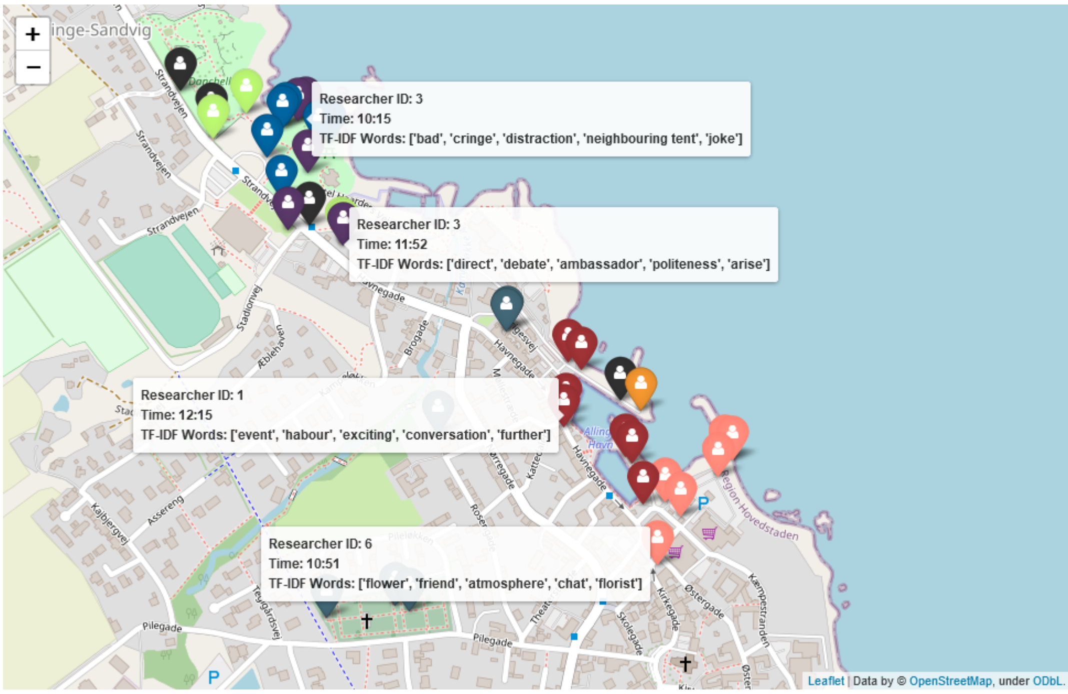


Figure 5.3: Map of data collection.

To sum up, these examples demonstrate that the combination and integration between digitized fieldnotes and geospatial data in the form of an interactive visualization offers allow us to identify new patterns in their data that we might not otherwise have seen. One could also imagine combining the ethnographic data with weather data or maybe even social media posts pertaining debates that are also described in fieldnotes. The tags indicating "time", "date", "location", and "situation" allow us to combine the ethnographic data with a range of other data types that can further enrich our ethnographic data and our analysis. Similar to the sorting and searching steps described in the previous section, the ability to combine fieldnote data with other data sources can augment and expands our exploratory analysis. As such, this kind of analysis can be used to dig deeper and more systematically in the ethnographic data, for example, by re-reading and comparing fieldnotes from specific locations and moments and thereby trace the flow of attention in and across specific areas.

5.3.4 Processability - Exploring Fieldnotes through Topical Mapping

The structured format of our fieldnotes also allows us to make use of more advanced computational techniques to explore the textual content more in depth. Several techniques and methods from the data science fields of NLP and Machine Learning (ML) are particularly well suited for analysis of ethnographic data as they allow for statistical exploitation of language structure in search for potentially meaningful semantic patterns (Evans and Aceves, 2016). One such technique is topic modelling, a statistical method for identifying topics in large text corpora (Blei, 2012). Topic models come in many varieties, and they usually involve treating the words in text as topics, that is, latent distributions estimated by optimizing the internal coherence within each topic and minimizing the overlap with the other topics in the text data

(Blei, 2012). This means that the model will generate a range of topics that are all, to put it simply, lists of words and associated weights that canprovided that the researcher is equipped with sufficient context and domain knowledge then be interpreted by the researcher as topics in a text corpus.

Qualitatively oriented social scientists have used these types of computational methods for finding thematic relations in ethnographic material (Fischer and Ember, 2018), to augment the coding of archival data (Nelson, 2020); for generating hypotheses from interview data (Karlgren *et al.*, 2020); and more generally for explorative purposes in keeping with the grounded or "abductive" nature of much anthropology and sociology (Brandt and Timmermans, 2021). Having fieldnotes in a tabular format exported as a CSV-file makes it easy to upload and process with programming languages such as Python or R to then apply ML or NLP techniques on the data. Here the tags containing descriptive accounts from the field are of key interest, however, if applied, other tags such as "Place" and "Time" can also be utilized to find textual patterns across time and space.

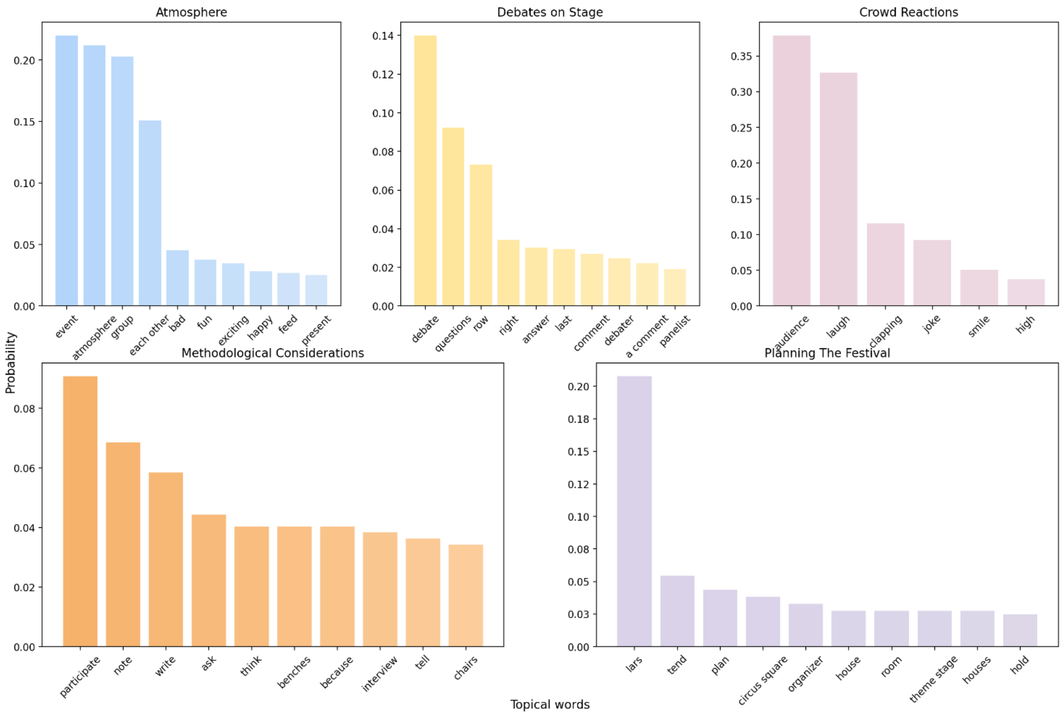


Figure 5.4: Fieldnote topics. Topics are identified using HSBM topic modelling.

In our own analysis, we decided to apply a hierarchical stochastic block model (HSBM) (Gerlach *et al.*, 2018) to model the fieldnote text data, in order to guide our exploratory search for overall patterns. In essence, HSBM topic models use co-occurrence information among words and documents, to find words in the corpus of fieldnotes that typically belong together (Gerlach *et al.*, 2018). We choose an HSBM model because it holds the desirable property, that it attempts to calculate the number of topics in the data as well and is thus not dependent on the researchers predefining a number of topics that the model should look for, unlike other popular topic models like LDA (Blei 2012, Gerlach *et al.* 2018, see also Carlsen and Ralund 2022).

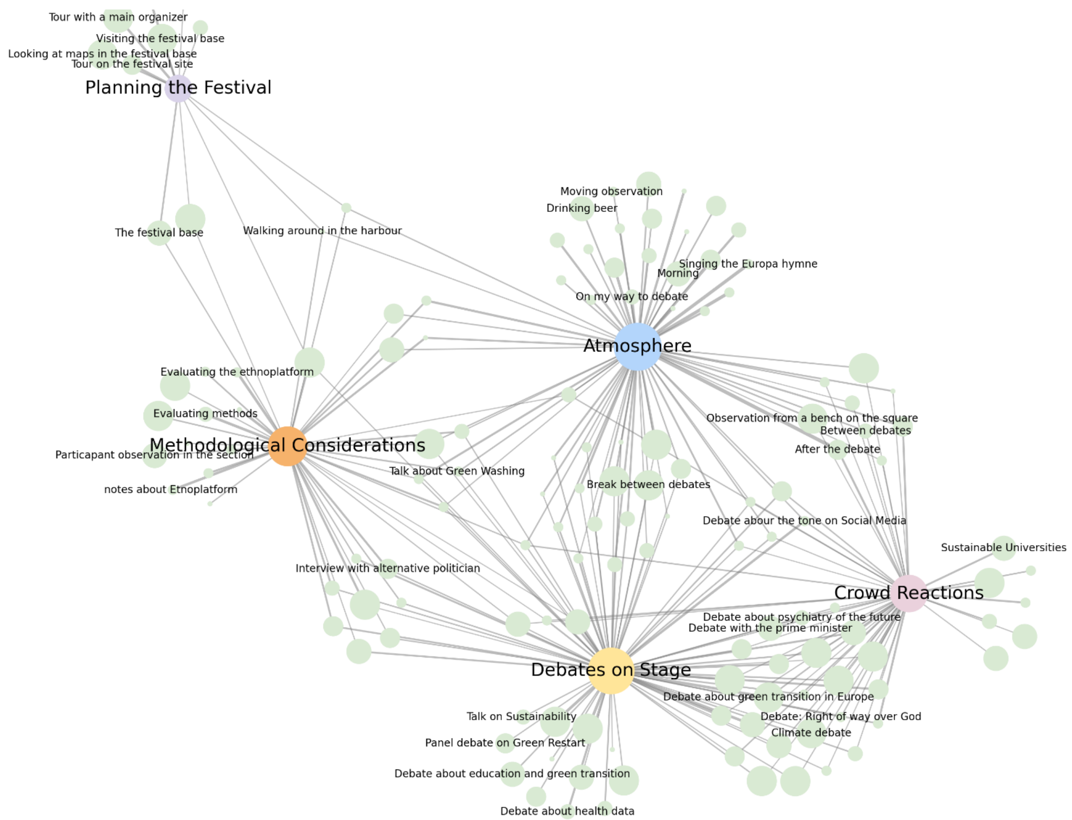


Figure 5.5: Topical mapping of fieldnotes.

Our initial model suggested that there were 31 topics in the data. However, after examining the words contained in each of the topics generated, as well as re-reading the fieldnotes related to the different topics, we find that five of the suggested topics represented the most relevant thematic topics in our data. In Figure 5.4, we present the five topics as generated by the HSBM model. The bars of the words in each chart indicate the importance weight of the specific word for the overall topic. For instance, the topic, we have named *Atmosphere*, is comprised of words like "event", "atmosphere", "exciting", and other words, which in both direct and indirect ways denote and revolve around the atmosphere at events and in groups. *Debates on Stage* is about what happens during and around debates. *Crowd Reactions* is centered on the reactions of the audience. *Methodological Considerations* differs by being more about methodological reflections that ethnographers had in the field. The last theme, *Planning the People's Meeting* is about the process of planning the festival; here the word "Lars" is the name of our key contact person in the organizing team. Each of our fieldnotes can then be labeled with the topic(s) that they contain, by using the weights generated by the model.We assign a topic if our model predicts that it is 50% or more likely that the fieldnote contains that topic.

Moving on in the exploration of topics in our data, we can use tools from network science to visualize and examine the relations between the topics and fieldnotes. In Figure 5.5, we have constructed what we call a *topical mapping* of the fieldnotes. This is a so-called bipartite network visualization, where we draw two types of nodes (points in the network), the green nodes represent fieldnotes, and the colored nodes represent topics. Each of the topic nodes have been annotated with their topic names, and the fieldnote nodes have been annotated with their "Situation"-tag. Likewise, the size of the fieldnote nodes corresponds to the length of the text from the "Field observation"-tag of each fieldnote. Having visualized the nodes, we can then draw an edge (line) between a topic and a fieldnote, if the specific node is assigned with the topic. In this way, our topic mapping visualizes the relations between each fieldnote and the topics it covers.

In this mapping, the location of nodes indicates topical similarity. The closeness of the "Debates on Stage" and "Crowd Reactions" topics, for example, indicates that the two topics are often co-present in our fieldnotes. The fieldnotes located between the topics appear to mainly describe situations related to political debates. Conversely, the topic "Atmosphere" seems to appear more prominently with reference to specific situations in *between* debates.We can also see that fieldnotes related to the topic "Planning the Festival" are related less to the other topics and thus seem like a very separate kind of notes. When we look closer at the individual notes, it appears that they are written just before the festival was open to visitors, thus utilizing the "Time" tags. From the topical mapping, we can only see an overall thematic pattern of our fieldnotes, however, this can provide a basis for topically focused re-reading of the fieldnotes with a new analytical view.

5.4 Conclusions

Producing and processing fieldnotes do not need to be a lonely and low-tech endeavor. As we have shown in this article, significant gains can be reaped from using digital formats for making ethnographic fieldnotes more sharable between researchers, and more suited for computational text analysis and other mixed methods approaches. Indeed, the EthnoPlatform is especially suited for interdisciplinary teams with multiple data sources and members with a mix of qualitative and quantitative backgrounds. The possibility of using computational techniques for pattern discovery can help mitigate biases of individual researchers, as well as opening up what Abramson *et al.*, 2018 have described as the black box of idiosyncratic choices made by qualitative researcher when producing and processing their data. However, the tool can also be useful for individual researchers in the context of more open-ended long-term fieldwork. In this case, the researcher could use fewer pre-defined tags to retrain as must freedom as possible, but still gain the benefits of the resulting structured corpus.

As we have demonstrated using our attention study as an example, the tag-based data architecture of the EthnoPlatform allows ethnographers to log their fieldnotes in a structured format. This not only increased organizability through sorting and searching, and better integrabilitythrough merging the fieldnotes with other data sources, but also opens up for using them as sources of computational analysis in their own right. As has been suggested in recent scholarship in the interface between mixed methods and data science, quantitative analysis of qualitative data can thus be used to not only validate qualitative findings (Abramson *et al.*, 2018; Grigoropoulou and Small, 2022; Maxwell, 2010) but also to automate labor-intensive manual coding procedures (Marathe and Toyama, 2018). In many ways, the present article follows in the footsteps of and seeks to contribute to this important nascent literature. For example, in our topical mapping of fieldnotes from the People's Meeting, we made several interpretative steps to infer meaning from the network. These steps derived in a qualitative reading of the data; however, the quantitative nature of the topical mapping made these steps transparent and translatable among ourselves as well as to others. As a quantitatively based form of ethnographic data exploration, the topic map thus provided an overview of the data and revealed patterns that would have been hard to find via traditional qualitative means. As such, we again see how the EthnoPlatform as a tool for the digitization of fieldnotes allows for not just deductive but also inductive/abductive strategies for the collection, processing, and analysis of ethnographic data.

This new form of topical modelling of fieldnotes *topical mapping* shows significant promise for not just anthropological and sociological but also other fields of social scientific and digital humanities research. Indeed, it holds the potential to serve both as an instrument for analytical insights in itself and an offset to further empirical work. One could, for instance, explore the nature and intersections of topics by doing deep qualitative interpretations of selected fieldnotes or constellations of fieldnotes, or one might imagine that topical mappings can be served as codes in a qualitative coding protocol. Much as we saw in the two earlier sections on data organizing and data integration, the use of network visualizations and NLP (including topic models) can open up for new unexplored paths and patterns in ethnographic analysis difficult via existing ways of processing fieldnotes, whether manual or automated.

Despite these notably gains in terms of efficiency, transparency, and veracity, the EthnoPlatform of course has limitations. Notably, these include the inevitable trade-offs between rigor and flexibility arising from using tags to structure ethnographic fieldnotes, and, more generally, the significant methodological and epistemological pitfalls pertaining from any (over) reliance on automated techniques for quantitative text analysis (Carlsen and Ralund, 2022). Even with the recent introduction of sophisticated unsupervised machine learning methods into the social scientific study of text data (Enggaard *et al.*, 2023; Kozlowski *et al.*, 2019; Milbauer *et al.*, 2021), any attempt to make use of the EthnoPlatform's significant potential computational processing and modelling must be combined with a strong and systematic qualitative component. This spans all the way from the pre-processing to the visualization stage, including in-depth reading of and continuous revisiting of selected raw fieldnotes. Ideally, such a systematic complementarity (Blok and Pedersen, 2014) between qualitative and quantitative data, methods, and approaches will allow for an iterative process of zooming in on and out from the microdetails of fieldwork to obtain a bigger and thicker understanding of such ethnographic materials. Indeed, as we hope to have shown, this might be particularly beneficial in interdisciplinary research projects with a core ethnographic component.

Bibliography

Abramson, Corey M., J. Joslyn, Katharine A. Rendle, S. B. Garrett, and Daniel Dohan (2018). „The Promises of Computational Ethnography: Improving Transparency, Replicability, and Validity for Realist Approaches to Ethnographic Analysis". In: *Ethnography* 19.2, pp. 254–284.

Albris, Kristian, E. I. Otto, S. L. Astrupgaard, E. M. Gregersen, L. S. Jørgensen, O. Jørgensen, C. R. Sandbye, and S. Schønning (2021). „A View from Anthropology: Should Anthropologists Fear the Data Machines?" In: *Big Data & Society* 8.2, p. 20539517211043656.

Astrupgaard, S. L., E. M. Gregersen, and C. R. Sandbye (2022). „See You Later, Thick Data? How We Experimented with Doing Collaborative Fieldwork as Part of an Interdisciplinary Research Project". In: *Anthrodendum*. Available at: https://anthrodendum.org/author/distract/ (accessed 17 November 2022).

Astrupgaard, S. L., E. M. Gregersen, C. R. Sandbye, and Morten Axel Pedersen (Forthcoming). „Broad Data Ethnography: Ethnographic Methods in the Age of Interdisciplinary Collaboration". In.

ATLAS.ti, Scientific Software Development GmbH (2023). *ATLAS.ti Mac. Version 23.2.1 [Qualitative Data Analysis Software]*.

Bernard, H. Russell (2006). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. 4th. AltaMira Press.

Bjerre-Nielsen, Andreas and Kristoffer Laigaard Glavind (2022). „Ethnographic data in the age of big data: How to compare and combine". In: *Big Data & Society* 9.1.

Blei, David M. (2012). „Probabilistic Topic Models". In: *Communications of the ACM* 55.4, pp. 77–84.

Blok, Anders, Henrik B. Carlsen, Tobias Bornakke Jørgensen, Mogens Møller Madsen, Søren Ralund, and Morten Axel Pedersen (2017). „Stitching Together the Heterogeneous Party: A Complementary Social Data Science Experiment". In: *Big Data & Society* 4.2, p. 2053951717736337.

Blok, Anders and Morten Axel Pedersen (2014). „Complementary social science? Quali-quantitative experiments in a big data world". In: *Big Data & Society* 1.2.

Borgatti, Steve (1989). „Using Anthropac to Investigate a Cultural Domain". In: *CAM Newsletter* 1.2, pp. 11–11.

Boyer, Dominic and George E. Marcus (2021). *Collaborative Anthropology Today: A Collection of Exceptions*. Cornell University Press.

Brandt, Patrick and Stefan Timmermans (2021). „Abductive Logic of Inquiry for Quantitative Research in the Digital Age". In: *Sociological Science* 8, pp. 191–210.

Bunkenborg, Mikkel, Mikkel Nielsen, and Morten Axel Pedersen (2022). *Collaborative Damage: An Experimental Ethnography of Chinese Globalization*. Cornell University Press.

Carlsen, Hjalmar Bang and Snorre Ralund (Jan. 2022). „Computational grounded theory revisited: From computer-led to computer-assisted text analysis". en. In: *Big Data & Society* 9.1, p. 205395172210801.

Enggaard, T., A. Lohse, M. A. Pedersen, and S. Lehmann (2023). „Dialectograms: Machine Learning Differences Between Discursive Communities". In: *ArXiv Preprint*. eprint: 2302.05657.

EthOS (2023). *The EthOS Platform*. URL: https://ethosapp.com/platform.

Evans, James A. and Pedro Aceves (July 2016). „Machine Translation: Mining Text for Social Theory". en. In: *Annual Review of Sociology* 42.1, pp. 21–50.

Favero, P. S. and E. Theunissen (2018). „With the Smartphone as Field Assistant: Designing, Making, and Testing EthnoAlly, a Multimodal Tool for Conducting Serendipitous Ethnography in a Multisensory World". In: *American Anthropologist* 120.1, pp. 163–167.

Fielding, N. G. (2012). „Triangulation and Mixed Methods Designs". In: *Journal of Mixed Methods Research* 6.2, pp. 124–136.

Fischer, M. D. and C. R. Ember (2018). „Big Data and Research Opportunities Using HRAF Databases“. In: *Computational Social Sciences*, pp. 323–336.

FitzGerald, J. and J. Mills (2022). „The Importance of Ethnographic Observation in Grounded Theory Research“. In: *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 23.2, p. 15.

Folkemødet (2022). *About Folkemødet (in English)*. URL: https://folkemoedet.dk/om-folkemodet/about-folkemodet-in-english/.

Gerlach, Martin, Tiago P. Peixoto, and Eduardo G. Altmann (July 2018). „A network approach to topic models“. en. In: *Science Advances* 4.7. ZSCC: 0000087, eaaq1360.

Gill, T., D. H. Slater, and B. Steger (2015). „Japan Copes with Calamity“. In.

Gregersen, E. M., A. Lohse, N. B. Andersen, and M. A. Pedersen (Forthcoming"). *EthNote: An Open-Source Tool for Fieldnote Collection, Sharing, and Processing*. https://ethnote.org.

Grigoropoulou, N. and M. L. Small (2022). „The Data Revolution in Social Science Needs Qualitative Research“. In: *Nature Human Behaviour* 6.7, pp. 904–906.

Hastrup, K., J. Flora, and A. O. Andersen (2016). „Moving Facts in an Arctic Field: The Expedition as Anthropological Method“. In: *Ethnography* 17.4, pp. 559–577.

Hymes, D. H., ed. (1965). *The Use of Computers in Anthropology*. Reprint 2011. De Gruyter Mouton.

Indeemo (2018). *An Introduction to Mobile Ethnography*. www.Indeemo.com.

Karlgren, J., R. Li, and E. M. M. Milgrom (2020). „Text Mining for Processing Interview Data in Computational Social Science“. In: *ArXiv Preprint*. eprint: 2011.14037.

Kemper, R. V., R. K. Wetherington, and M. Adler (1992). „From Mainframes to Micros: Computer Applications for Anthropologists“. In: *Social Science Computer Review* 10.4, pp. 470–489.

Korsby, T. M. and A. Stavrianakis (2018). „Moments in Collaboration: Experiments in Concept Work“. In: *Ethnos* 83.1, pp. 39–57.

Kozlowski, Austin C., Matt Taddy, and James A. Evans (Oct. 2019). „The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings“. en. In: *American Sociological Review* 84.5, pp. 905–949.

Krieg, L. J., M. Berning, and A. Hardon (2017). „Anthropology with Algorithms?“ In: *Medicine Anthropology Theory* 4.3.

Lumivero (2023). *NVivo (Version 14)*. www.lumivero.com.

Lyon, S. M. (1999). „“OPEN” Ethnography and the Internet in the Field: Increased Communications, Feedback, and “Usability” Versus Technical and Ethical Issues“. In: *Journal of the Anthropology Society of Oxford* 30.1, pp. 35–50.

Marathe, M. and K. Toyama (2018). „Semi-Automated Coding for Qualitative Research: A User-Centered Inquiry and Initial Prototypes“. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, pp. 1–12.

Marda, V. and S. Narayan (2021). „On the Importance of Ethnographic Methods in AI Research“. In: *Nature Machine Intelligence* 3.3, pp. 187–189.

Maxwell, J. A. (2010). „Using Numbers in Qualitative Research“. In: *Qualitative Inquiry* 16.6, pp. 475–482.

Milbauer, Jeremiah, Adarsh Mathew, and James Evans (2021). „Aligning Multidimensional Worldviews and Discovering Ideological Differences“. en. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. ZSCC: 0000000, pp. 4832–4845.

Munk, Anders Kristian (June 2019). „Four Styles of Quali-Quantitative Analysis: Making sense of the new Nordic food movement on the web“. en. In: *Nordicom Review* 40.s1, pp. 159–176.

Nelson, Laura K. (Feb. 2020). „Computational Grounded Theory: A Methodological Framework“. en. In: *Sociological Methods & Research* 49.1, pp. 3–42.

Nelson, Laura K. (Mar. 2021). „Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century U.S. South“. en. In: *Poetics*. ZSCC: 0000006, p. 101539.

Nippert-Eng, C. E. (2015). *Watching Closely: A Guide to Ethnographic Observation*. Oxford University Press.

Paredes, P., A. Rufino Ferreira, C. Schillaci, G. Yoo, P. Karashchuk, D. Xing, C. Cheshire, and J. Canny (2017). „Inquire: Large-Scale Early Insight Discovery for Qualitative Research“. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, pp. 1562–1575.

Phillippi, J. and J. Lauderdale (2018). „A Guide to Field Notes for Qualitative Research: Context and Conversation“. In: *Qualitative Health Research* 28.3, pp. 381–388.

Sago (2023). *QualMobile*. https://sago.com/en/solutions/platforms/qualmobile/.

Sanjek, R., ed. (1990). *Fieldnotes: The Makings of Anthropology*. http://www.jstor.org/stable/10.7591/j.ctvv4124m. Cornell University Press.

Seidel, J. V. and J. A. Clark (1984). „The ETHNOGRAPH: A Computer Program for the Analysis of Qualitative Data“. In: *Qualitative Sociology* 7, pp. 110–125.

SocioCultural Research Consultants (2021). *Dedoose (Version 9.0.17), Cloud Application for Managing, Analyzing, and Presenting Qualitative and Mixed Method Research Data*. www.dedoose.com.

Spärck Jones, K. (1972). „A Statistical Interpretation of Term Specificity and its Application in Retrieval“. In: *Journal of Documentation* 28.1, pp. 11–21.

VERBI Software (2021). *MAXQDA 2022 [Computer Software]*. www.maxqda.com.

Wang, T. (2012). „Writing Live Fieldnotes: Towards a More Open Ethnography“. In.

Weinstein, M. (2006). „TAMS Analyzer: Anthropology as Cultural Critique in a Digital Age“. In: *Social Science Computer Review* 2.1, pp. 68–77.

Appendices

6

6.1 Appendix for Paper 1

6.1.1 Significance Tests

In this appendix, we present statistical significance tests for all the comparisons made in the main paper. For most of the comparisons, we use a standard Z-test for proportions, as we are comparing the share of users or people in a category across groups. When examining the average number of likes per user, we instead use a two-tailed T-test with the unequal variance correction. The first two tables present the difference between the two groups mentioned in the "Comparison" column, with the format $group1 - group2$, such that in the first comparison in table A.1, we see that there are 8.30%-points more non-politically active women on social media than women in the population. For tables A.3, A.4, and A.5, the test compares politically active users against non-politically active users.

Comparison	Diff	Stat	P-Value	Test-Type
Non-Political Women vs Population	8.30	289.65	0.0	Z-Test
Non-Political Women vs Population Likes	19.93	10203.27	0.0	Z-Test
Political Women vs Population	5.77	94.32	0.0	Z-Test
Political Women vs Population Likes	1.20	130.05	0.0	Z-Test
Political Women vs Non-Political Women	-2.53	-37.65	0.0	Z-Test
Political Women vs Non-Political Women Likes	-18.73	-2140.08	0.0	Z-Test
Political Women vs Political Men Average Likes	-8.24	-25.53	0.0	T-Test
Non-Political Women vs Non-Political Men Average Likes	90.15	125.02	0.0	T-Test

Table A.1: **Significance Test Results for Gender Differences.** The tests compare group 1 against group 2 using the test type indicated. For the T-test, a two-tailed test with unequal variance is used, as the assumption of equal variance is not satisfied.

6.1.2 Gender and Ethnicity Estimation

We estimate the gender of social media users by mapping their name to the official legal list of names for men and women (Familieretshuset, 2024). To begin with, we need to assign each user a unique username, as it is possible to change your username on Facebook. As such, we assign each user the username by which they have made the most likes. We then lowercase it, remove whitespace, and split potential double-names,

Comparison	Diff	Stat	P-Value	Test-Type
Non-Political non-Danish vs Population	36.17	1361.41	0.0	Z-Test
Non-Political non-Danish vs Population Likes	20.62	10908.87	0.0	Z-Test
Political non-Danish vs Population	18.98	350.29	0.0	Z-Test
Political non-Danish vs Population Likes	17.89	2221.93	0.0	Z-Test
Political non-Danish vs Non-Political non-Danish	-17.19	-268.01	0.0	Z-Test
Political non-Danish vs Non-Political non-Danish Likes	-2.73	-323.31	0.0	Z-Test
Political non-Danish vs Political Danish Average Likes	-2.28	-7.42	0.0	T-Test
Non-Political non-Danish vs Non-Political Danish Average Likes	-109.77	-166.68	0.0	T-Test

Table A.2: **Significance Test Results for Ethnic Minority Differences.** The tests compare group 1 against group 2 using the test type indicated. For the T-test, a two-tailed test with unequal variance is used, as the assumption of equal variance is not satisfied.

Cluster	Diff	Stat	P-Value	Test-Type
Shopping	-4.02	-93.28	0.0	Z-Test
Cars, Men and Nationalism	1.09	37.64	5.2e-310	Z-Test
Entertainment	-16.75	-264.44	0.0	Z-Test
Local Pages	-1.58	-37.16	2.8e-302	Z-Test
Horses	-1.27	-58.69	0.0	Z-Test
News, Civics and Politics	22.53	396.03	0.0	Z-Test

Table A.3: **Significance Test Results for Cluster Differences.** The tests compare the share of users belonging to the cluster in the politically active user set against the share of users in the non-politically active set.

i.e., names such as "Anne-Sofie," which are fairly typical in a Danish context. We then map the first name to the list of legal names and assign gender.

While the existence of stringent naming laws in Denmark provides us with a relatively high level of confidence in the gender associated with names, certain complexities remain. Most notably, the gender category that we assign to social media users is estimated based on the gender they would have had if they were born in Denmark and given that name at birth. Our methodology thus cannot measure the user's own gender identification, nor does it capture cases where users have a name that corresponds to a different gender category than the Danish one, such as the name "Andrea," which is only legal to name a girl in Denmark, but is a common name for men in southern Europe.

We also estimate the so-called ethnic heritage of the social media users. Ethnic heritage is a term used by Statistics Denmark, which categorizes Danes into five categories: Danish heritage, immigrants from a western country, a child of an immigrant from a western country, an immigrant from a non-western country, and a child of an immigrant from a non-western country (Elmeskov, 2019). We choose to focus on the differences between people with Danish heritage and non-Danish heritage, partly because this is a very salient political cleavage in Danish politics, and partly because it is a lot easier to label using usernames. As such, we collapse the 4 non-Danish heritage categories.

Political Characteristic	Diff	Stat	P-Value	Test-Type
Ideology	-0.03	-4.28	0.00	T-Test
Extremism	0.06	17.54	0.00	T-Test
Partisanship	-0.00	-1.04	0.30	T-Test

Table A.4: **Ideology, Extremism, and Partisanship Significance Test Results.** The tests compare the political leanings of users in the politically active user set against the population. A two-tailed T-test with unequal variance is used, as the assumption of equal variance is not satisfied.

Correlation with Like Frequency	Correlation	P-Value	Test-Type
Ideology	-0.01	0.00	T-Test
Ideological Extremity	0.17	0.00	T-Test
Partisanship	0.03	0.00	T-Test

Table A.5: **Correlation Significance Test Results.** The tests compare the correlation between political leanings of users and their (log.) like frequency against a correlation coefficient of 0. using a two-tailed T-test

We use a supervised machine learning model to estimate the heritage of the users based on their full name. For computational efficiency reasons, we focus on labeling names rather than users, and then assign each user a heritage category based on the user's name. We begin by sampling 10,000 unique names, which we then label manually. We then create an 80%/10%/10% split of the data into training, validation, and test sets, stratifying on the labels to ensure both Danish and non-Danish heritage names in each split. Using this, we train 4 different transformer models: Standard BERT, Multilingual BERT, standard RoBERTa, and finally an XLM RoBERTa model (Devlin *et al.*, 2019; Liu *et al.*, 2019; Conneau *et al.*, 2020). We train them all for 10 epochs with 500 warmup steps and standard training parameters. We pick the iterations of the models that perform best on the validation set and use the test set for unbiased performance prediction.

Model	Loss	Accuracy	F1 Score	Recall	Precision
bert-base-cased	0.474433	0.946	0.945640	0.971264	0.966819
bert-base-multilingual-cased	0.410224	0.948	0.947109	0.975862	0.964773
roberta-base	0.347106	0.937	0.938444	0.955172	0.971930
xlm-roberta-base	0.393558	0.944	0.942630	0.975862	0.960407

Table A.6: **Comparison of different models based on various performance metrics.** All measurements are done on the hold-out test set.

All the models perform satisfactorily with very good and balanced performance metrics on the test set. As such, we simply use the simplest and fastest model, the base BERT model. After fine-tuning, we predict the heritage of all names and assign an ethnic heritage category to all users.

6.1.3 Cultural Cluster Creation

To make cultural clusters, we begin by taking the 10,000 most liked pages by the non-politically active users and then create a bipartite network G. There are two sets of nodes in G, $U \in \{1, ..., i\}$ representing all the social media users, both the politically active and the non-politically active, and $V \in \{1, ..., j\}$ representing all the 10,000 most liked non-political pages. We then add all the edges between U_i and V_j to G with a weight based on how many times U_i had liked a post by the page V_j.

In order to cluster the pages, we project G onto the page set, meaning we make a new graph G_v containing only the nodes in V, and edges linking pages with a weight equal to the number of users in U who have liked both pages. As the number of users vastly outnumbers the number of pages, G_v is a very densely connected graph. We therefore perform network backboning using noise correction, which aims at removing the edges that are most likely to be noise (Coscia and Neffke, 2017). We obtain the backboned graph B_v containing 9,983 nodes (pages) and 3,033,898 edges, and perform clustering on this graph using the Louvain modularity optimization method (Blondel *et al.*, 2008).

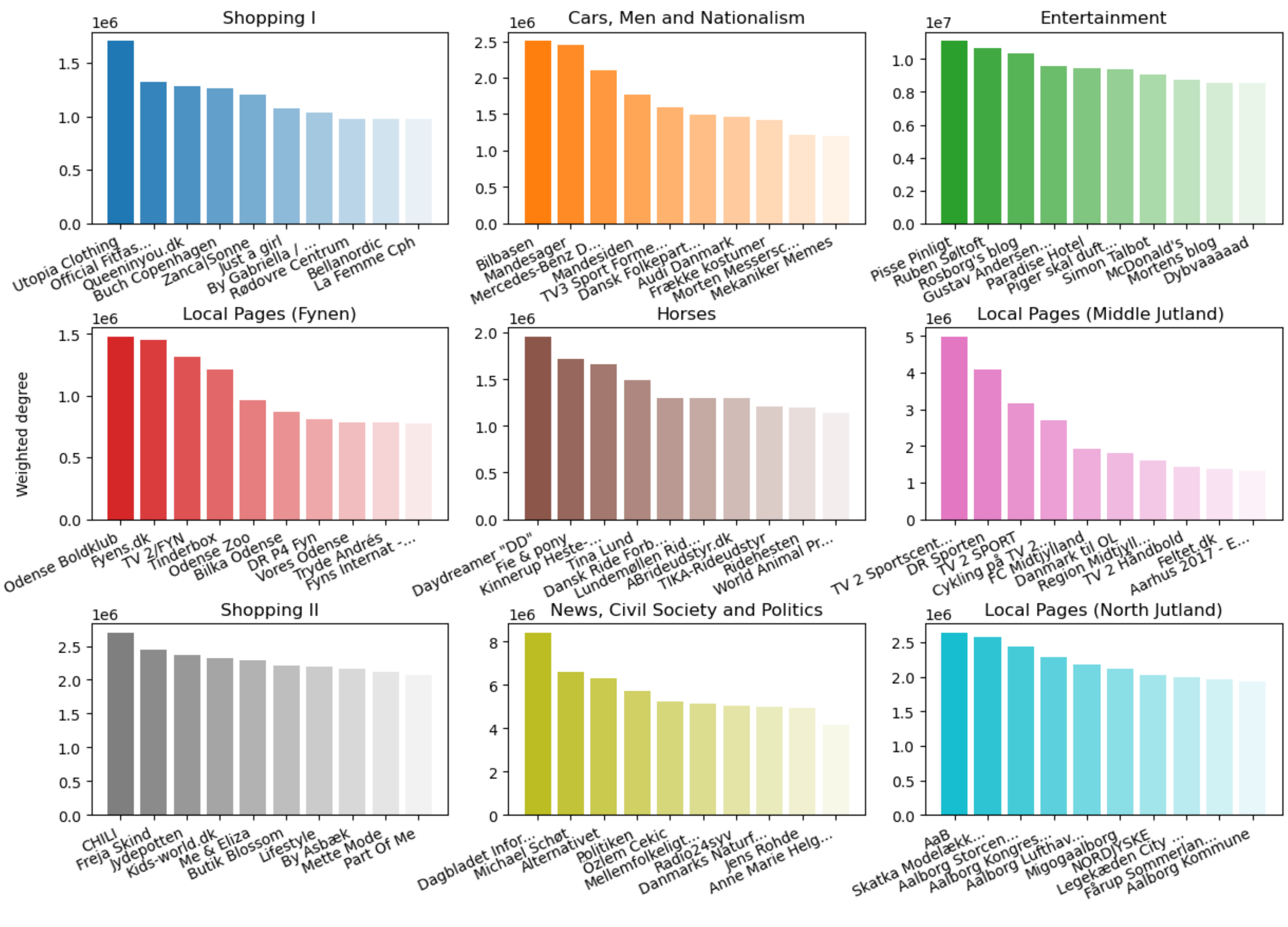


Figure A.1: All Cultural Preference Clusters

This process yields a partition of the pages into 9 clusters. We manually label these clusters by inspecting the top pages contained within each cluster based on in-degree and out-degree, in a process similar to the one typically followed when performing topic modeling. As part of this process, we combined the initial 9 clusters seen in figure A.1, into the 6 distinct clusters that we use in the main paper. Specifically, we combine the three clusters containing pages related to different local areas in Denmark, and the two shopping clusters.

6.1.4 Alternative Cultural Preference Clusters

Looking at the distribution of likes rather than users in each cultural preference cluster in figure A.2, we see that the differences shown in the main paper are more pronounced than when examining users. Similarly, we see that the right-wing nationalism cluster has become the third largest with almost 10% of the political likes.

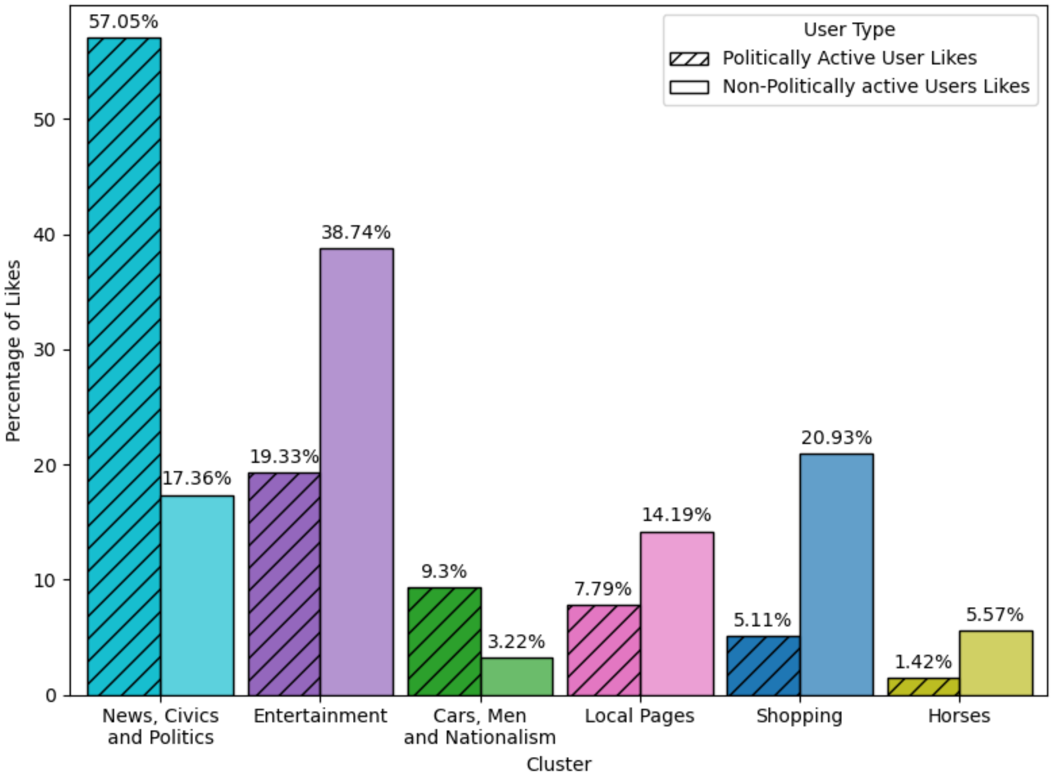


Figure A.2: Distribution of Likes in Cultural Preference Clusters. Each user is assigned the cluster where they have made the most likes.

Looking at the likes over time in figure A.3, we see results that are very similar to the results shown in the main paper. Only the differences between the politically active and the non-politically active are more pronounced.

In the main paper, we assign users to a cultural preference cluster by assigning them to the cluster with pages whose posts they have liked the most. However, this does not take into account that some pages are larger than others, and as such will get more users to like them in the first place. We therefore reproduce the results with a different assignment strategy, namely adding a weight to each page, that is $\frac{1}{d}$ where d is the degree of the page, meaning the number of people in our dataset that have liked a post by that page. We then assign each user to the cluster that they have liked the most, weighing each like by the page weight. As seen in figures A.4 and A.5, this makes almost no difference when looking at users or likes. We have reproduced this with a weight of $\frac{1}{\log(d)}$ instead, though this makes no difference, so we omit it here for brevity.

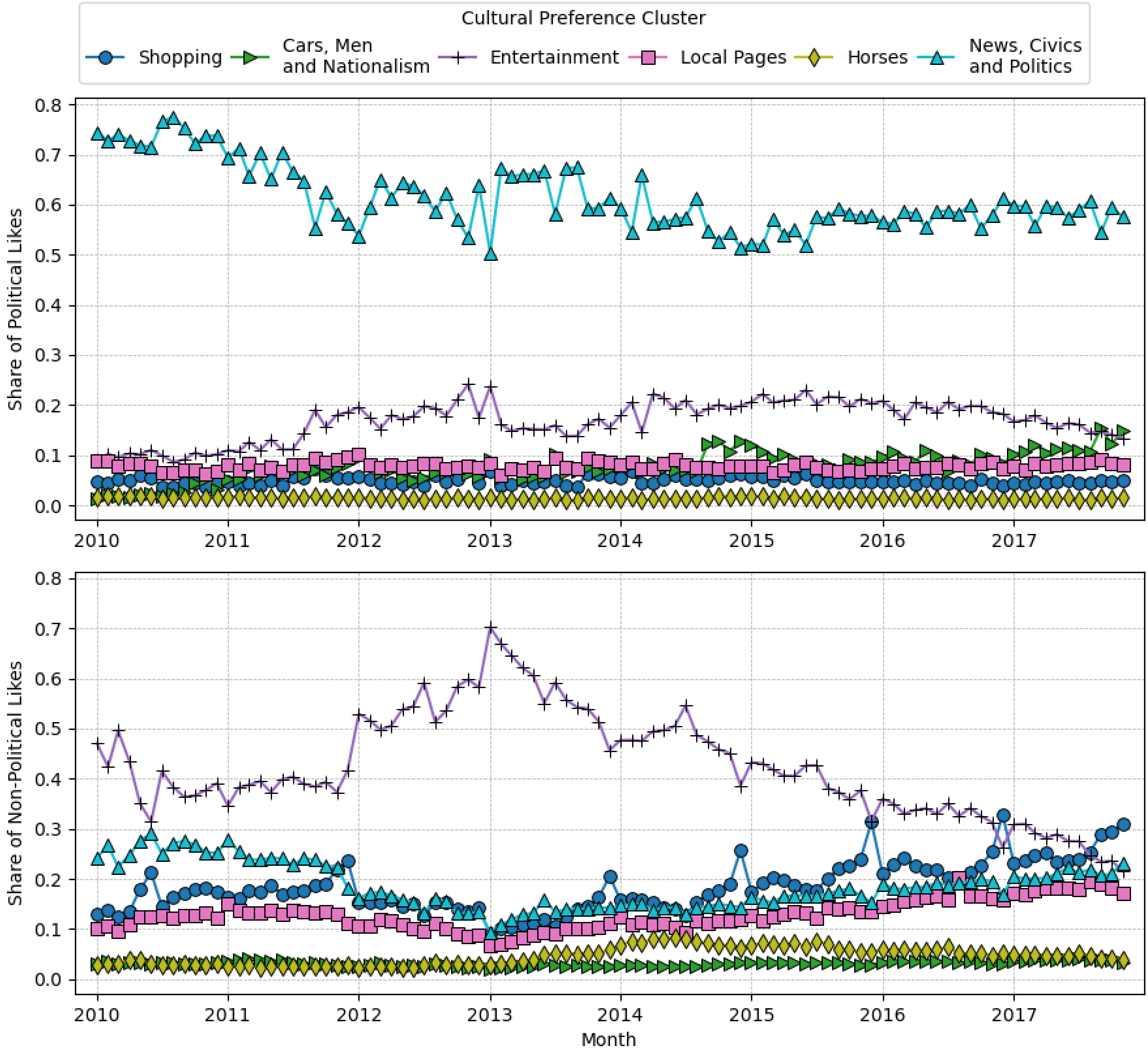


Figure A.3: Cultural Preference Cluster Likes Over Time

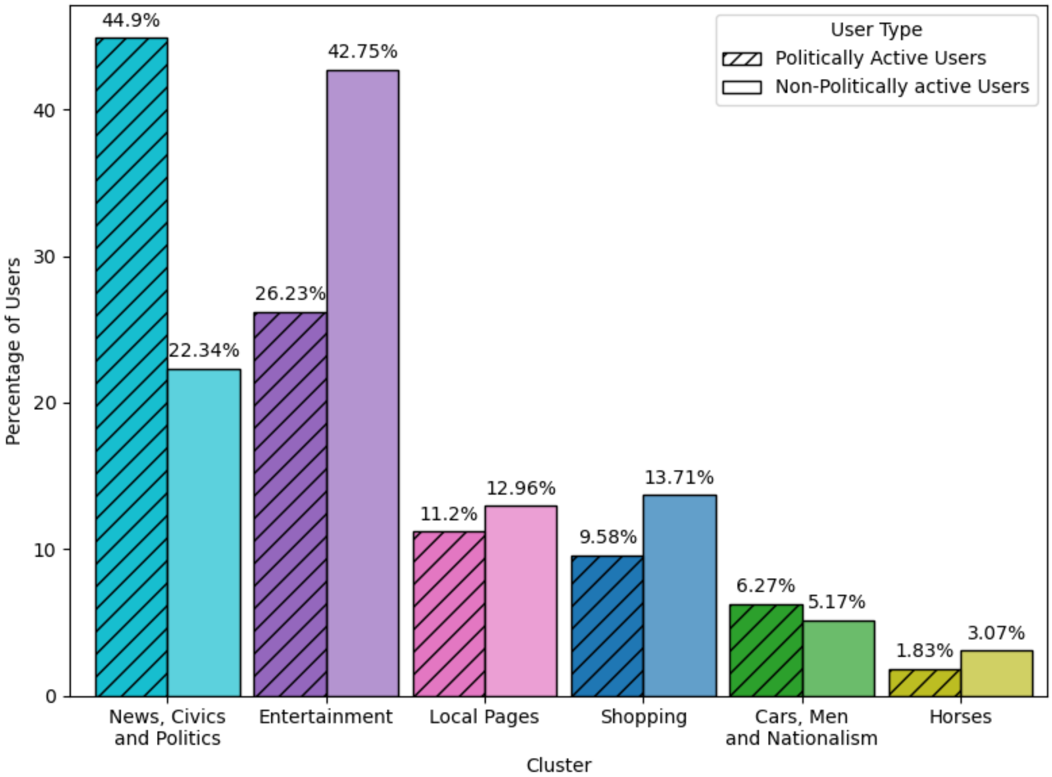


Figure A.4: Distribution of Users in Cultural Preference Clusters. Weighted assignment of users, based on inverse edge weight.

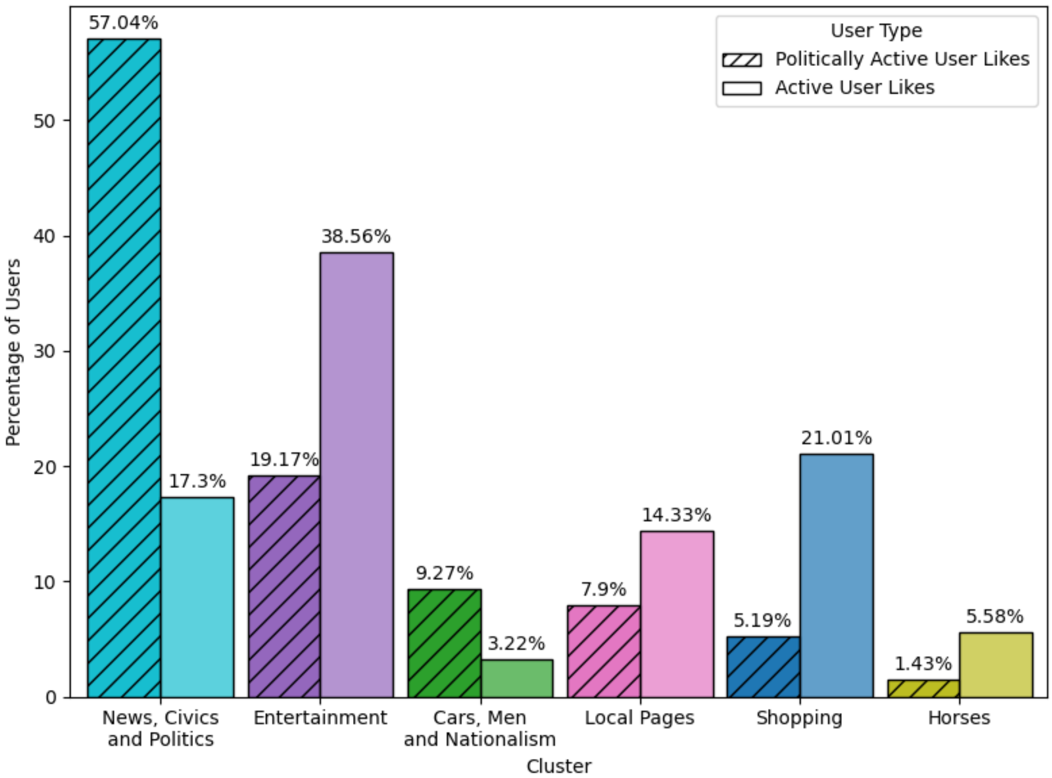


Figure A.5: Distribution of Likes in Cultural Preference Clusters. Weighted assignment of users, based on inverse edge weight.

6.1.5 Alternative Ideology Analysis

In this section, we show the results of comparing the public and the politically active social media users, if we use a different ideological measure for the politically active social media users. Here we estimate their ideology based on the parties of the politicians they have liked. Specifically, we use the average ideology scores of the parties given by the population in the election survey and take each social media user's ideology score to be the average of the parties they have liked. As such we measure ideology as:

$$I = \frac{\sum_n^N P_n}{N} \tag{6.1}$$

Where I is the ideology score, and P_n is the ideology of the post liked by like n out of all N likes made by an individual. The scores are then normalized to 0-1 for comparison with the population.

As seen in table A.7, when measuring ideology in this way, the social media public looks very different. Specifically, we can see that it is way more right-leaning than the public. They are also more ideologically extreme than the public, though this is similarly the case when we measure it using our network measure of ideology.

	Political Users	Population
Ideology	0.612	0.516
Extremism	0.280	0.220
Partisanship	0.606	0.609

Table A.7: Alternative Ideology and Extremity. The left-right scale is normalized to span 0-1, and the ideological extremity measure is the distance on this scale from the midpoint of 0.5. The left-right scale goes from left (0) to right (1), meaning a higher score indicates more right-leaning.

Political Characteristic	Diff	Stat	P-Value	Test-Type
Ideology	0.10	14.93	0.00	T-Test
Extremism	0.06	16.53	0.00	T-Test
Partisanship	-0.00	-1.04	0.30	T-Test

Table A.8: Alternative Ideology, Extremism Significance Test Results. The tests compare the political leanings of users in the politically active user set against the population. A two-tailed T-test with unequal variance is used, as the assumption of equal variance is not satisfied.

We see a similar case when examining the distributions in figure A.6. Here we see that measuring ideology in this way leads us to seeing many more highly right-wing users, with almost no moderately right-wing users.

Looking at the correlation with like frequency in A.7, we also see the correlations found using the main measure of ideology, disappear when measuring ideology using parties rather than the network-based measure.

Over time we see in figure A.8 that the pattern is fairly similar to the one we see using the network measure of ideology. Specifically, we see an ideological shift to the right from 2010-2012, and then a stabilization

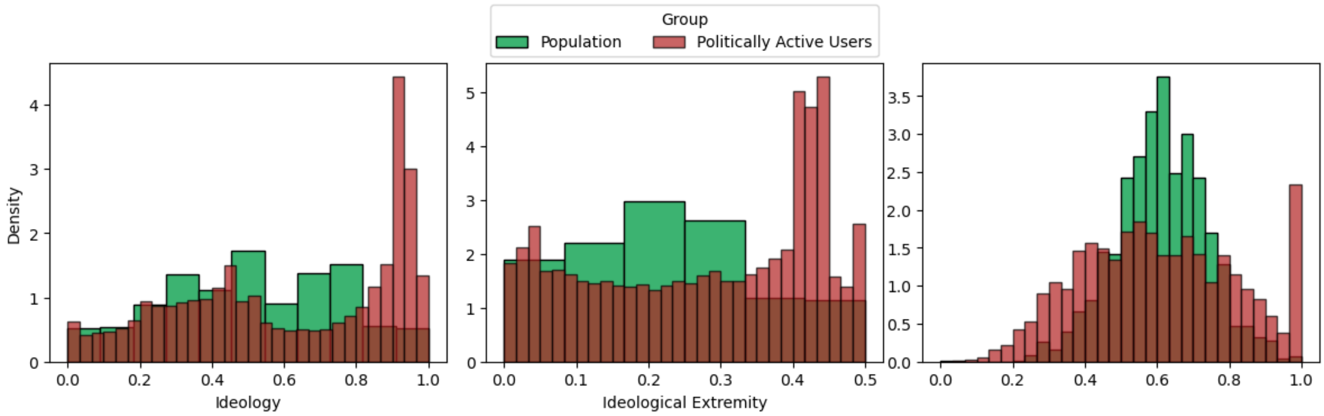


Figure A.6: **Party-based Ideology and Extremism**

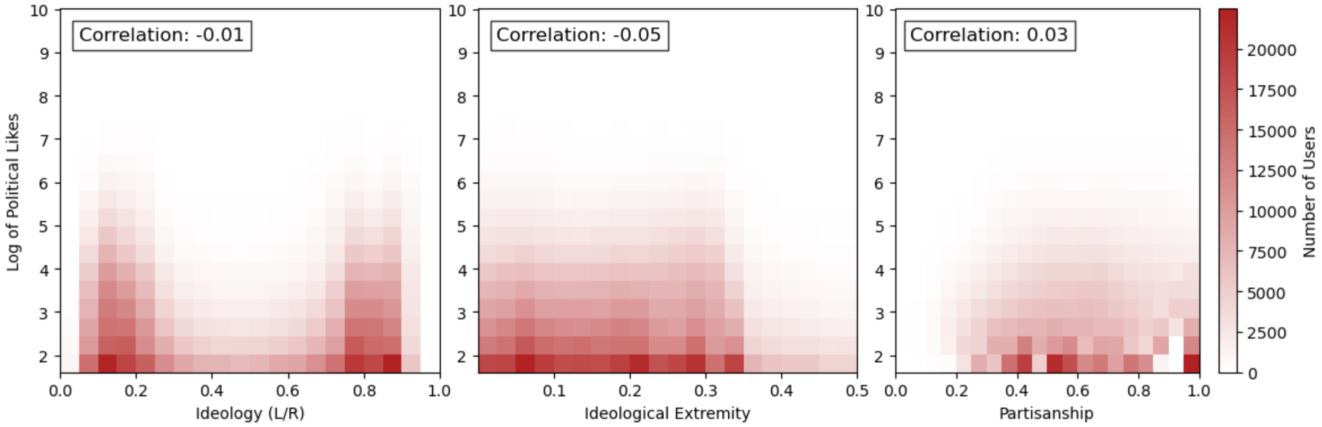


Figure A.7: **Party-based Like Frequency, Ideology, Extremism**

afterward, in this case around 0.6 on the ideological scale. For ideological extremism, we see an increase in the 2010-2012 period and then a gradual but very weak decline over time.

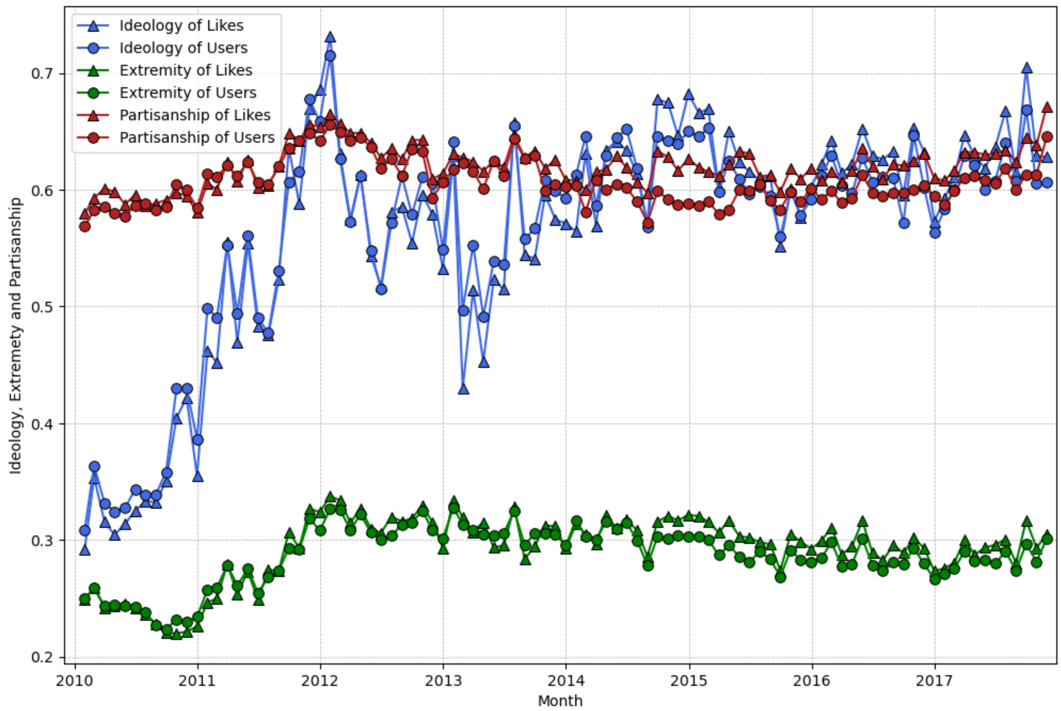


Figure A.8: Party-based Ideology, Extremism, and Partisanship over time

6.1.6 2015 Analysis

In this appendix, we show that the results shown in the main paper, concerning the comparisons between the population and the politically active social media users, are consistent if we only examine 2015 data. Specifically, in the main paper, we use data from the full data period, and here we present only the data from 2015. As we can see in figures A.9 and A.10, the results are nearly identical. We don't show results over time here, as they are identical to a snapshot of the plots in the main paper.

	Political Users	Population
Ideology	0.49	0.52
Extremism	0.28	0.22
Partisanship	0.59	0.61

Table A.9: Ideological extremity on social media and in the population in 2015. The left-right scale is normalized to span 0-1, and the ideological extremity measure is the distance on this scale from the midpoint of 0.5. The left-right scale goes from left (0) to right (1), meaning a higher score indicates more right-leaning.

Political Characteristic	Diff	Stat	P-Value	Test-Type
Ideology	-0.03	-4.71	0.00	T-Test
Extremism	0.06	17.35	0.00	T-Test
Partisanship	-0.01	-4.85	0.00	T-Test

Table A.10: Ideology, Extremism, and Partisanship Significance Test Results for 2015. The tests compare the political leanings of users in the politically active user set against the population. A two-tailed T-test with unequal variance is used, as the assumption of equal variance is not satisfied.

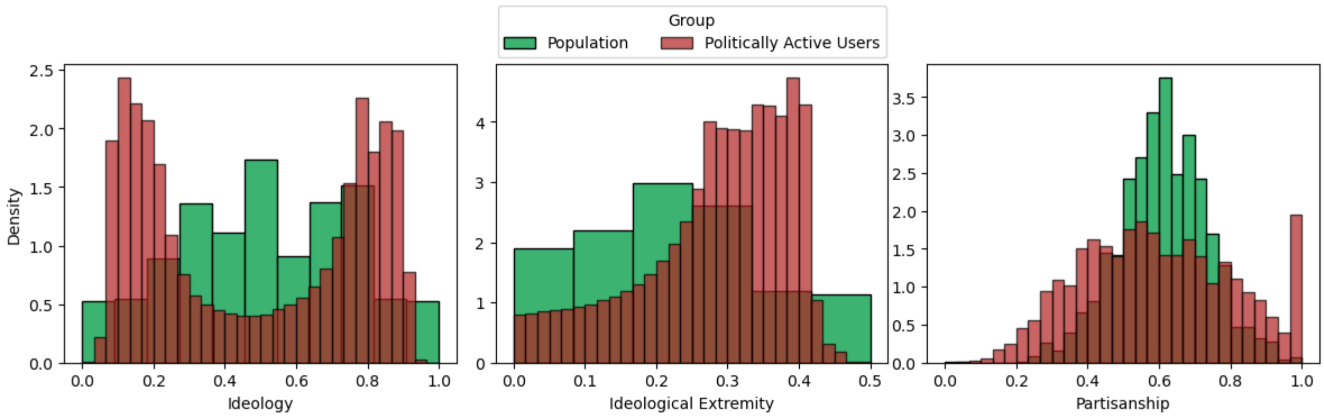


Figure A.9: Ideology and Extremism in 2015

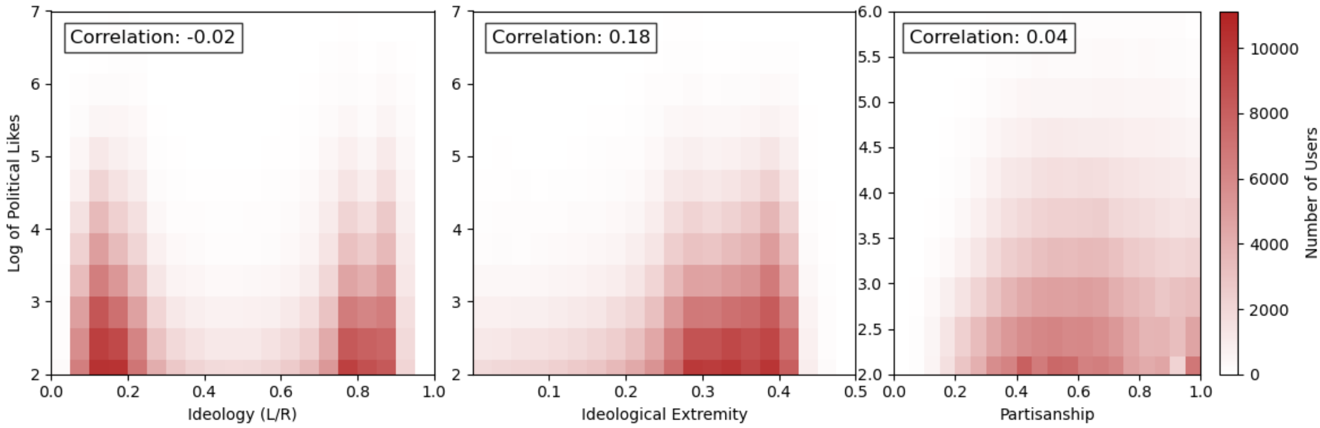


Figure A.10: Like Frequency, Ideology, Extremism, and Partisanship in 2015

Bibliography

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (Oct. 2008). „Fast unfolding of communities in large networks“. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov (2020). *Unsupervised Cross-lingual Representation Learning at Scale*. arXiv: 1911.02116 [cs.CL].

Coscia, Michele and Frank Neffke (2017). *Network Backboning with Noisy Data*. arXiv: 1701.07336 [physics.soc-ph].

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs.CL].

Elmeskov, Jørgen (2019). *Derfor Inddeler VI Verden I Vestlige og Ikke-vestlige lande*.

Familieretshuset (2024). *Godkendte Fornavne*.

Liu, Yinhan, 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: 1907.11692 [cs.CL].

6.2 Appendix for Paper 2

6.2.1 Topic creation

In this appendix we describe how our topics are related to the topics created in the comparative policy agenda projects (Green-Pedersen and Mortensen, 2019; Baumgartner *et al.*, 2019). As described in section 3.4.1, we begin our topic identification with the topics from the comparative policy agenda projects, which are made specifically for Danish politics (Green-Pedersen and Mortensen, 2019). We then use an HSBM topic model combined with a rigorous qualitative examination of the topics to create new topics as well as combine and remove topics where needed, as described in 3.4.1. Through all this, we utilized the topical subcategories of the comparative policy agenda project, to keep track of the delimitation of the different topics as described in the comparative policy agenda projects codebook (Green-Pedersen, 2019).

Our topic	**Topic in Green-Pedersen, 2019**	**Subtopics in Green-Pedersen, 2019**
Agriculture & Food	4	All except EU part of 402
Crime and justice	12	All except 1208
Culture	6, 23	607, all of 23
Defense	16	All
Economy & Business	1, 14, 15, 18	All of 1 except 103, 1411, All of 15, 1804, 1805, 1807, 1808
Education	6	All except 607 and 601
Energy	8	All
Environment & Climate	7	All except 712
EU	New topic	The EU part of all subtopics, and 1910
Foreign affairs	2, 18, 19	209, all of 18 except 1804, 1805, 1807, 1808, all of 19 except 1910
Government operations	2, 20	206, 211, All of 20
Health care	3	All
Immigration	2, 9	200, 201, all of 9
Infrastructure	7, 10, 14, 21	712, all of 10, All of 14 except 1408, 1409, 1411,2101, 2103, 2104
Labour	1, 5	103, all of 5
Religion	New topic	207 (religion part), 210, as well as all questions related to religion
Social policy & Welfare	2, 12, 13, 14	202, 204, 205, 1208, All of 13, 1408, 1409
Technology & Science	6, 17	601 and all of 17
Territories & former colonies	21	2015 and former Caribbean colonies

Table A.11: Initial topics

As seen in table A.11, we did initially have a topic list that was fairly close to the codebook of the Danish comparative policy project (Green-Pedersen, 2019), with a few editions like the EU and religion, which we based upon our topic model and qualitative reading of sampled posts.

Following the iterative process of sampling, reading, and updating topic definitions outlined in section 3.4.1, we landed on a final set of 15 topics that describe the Danish political social media sphere well. In table A.12 we show how our final topics relate to those in Green-Pedersen, 2019, and for some topics describe what possible edge cases there are with other topics. During our process from initial to final topics, we collapsed the topics of religion and immigration and removed the part about religion concerning the Danish state church, as no social media posts on the topics were found. We also combined agriculture and food with the environment, as they are predominantly discussed together, and our reading of the posts showed that often the environment is discussed in relation to agriculture (to be more precise in opposition to agriculture). We collapse climate and energy because climate policy is often discussed in relation to energy infrastructure and vice versa. There are some posts on climate that also relate to the environment, but they seem to more often relate to energy. There are also some posts on energy that only relate to energy and do not concern themselves with climate at all, but these are rare. The issue of unemployment benefits and early retirement (specifically social services like "kontanthjælp", "førtidspension", "efterløn", etc), have been moved from social policy and Welfare to labour. This change was done as it seemed to be discussed more as a labour issue than as a welfare issue. Any posts concerning state pensions are still in social policy and welfare. The prior topic of territories and former colonies has been split into foreign affairs and defense of the Arctic. Finally, we removed the topics of government operations as they did not appear to exist in Danish political social media.

Our topic	Topic in Green-Pedersen, 2019	Subtopics in Green-Pedersen, 2019	Edge-cases
Agriculture, Environment & Food	4, 7	All of 4 except EU part of 402, all of 7 except 712, 705 & 798	Some edge cases with climate
Crime and justice	12	All except 1208	
Culture	6, 23	607, all of 23	
Defense	16	All	
Economy & Business	1, 14, 15, 18	All of 1 except 103, 1411, All of 15, 1804, 1805, 1807, 1808	
Education	6	All except 607 and 601	Edgecase with students at universities, which is coded as technology and science.
Energy & Climate	7, 8	705 & 798, all of 8	
EU	New topic	The EU part of all subtopics, and 1910	
Foreign affairs	2, 18, 19, 21	209, all of 18 except 1804, 1806, 1807, 1808, all of 19 except 1910, 2105 and former Caribbean colonies	Some edge cases with defense, specifically military actions with Danish involvement is foreign affairs
Health care	3	All	
Immigration & Religion	2, 9	200, 201, 207, 210, all of 9 (religion part), as well as all questions related to religion as it concerns immigration and integration.	
Infrastructure	7, 10, 14, 21	712, all of 10, All of 14 except 1408, 1409, 1411, 2101, 2103, 2104.	Some edge cases with energy, when writing on energy infrastructure
Labour	1, 5, 13	103, all of 5, 1302	Edge cases with social policy and welfare with unemployment benefits and early retirement.
Social policy & Welfare	2, 12, 13, 14	202, 204, 205, 1208, All of 13 except (1302), 1408, 1409	Look at labour.
Technology & Science	6, 17	601, all of 17	This topic includes higher education which gives some edge cases with education.

Table A.12: Final topics

6.2.2 Topic labelling

In table A.13 we can see the topicwise labelling metrics of the best performing model on the test data. We test a number of different classification models as mentioned in section 3.4.2. For each pre-trained model chosen, we train a classifier using both a finetuned and a base model. By finetuning we here refer to further masked language training on our data, as opposed to "finetuning" the models to actually predict the labels. All the finetuned models were finetuned on a random 90% subset of the data with a 15% masked word probability for 10 epochs, where after the best model was chosen based on the model that best recovered the masked words in the 10% left out subset of the data.

Topic label	Topic name	F1	Recall	Precision
0	Agriculture, Environment & Food	0.829268	0.829268	0.829268
1	Crime & Justice	0.623377	0.533333	0.750000
2	Culture	0.695238	0.682243	0.708738
3	Defence	0.789474	0.681818	0.937500
4	Economy & Business	0.672897	0.615385	0.742268
5	Education	0.776471	0.709677	0.857143
6	Climate & Energy	0.786207	0.740260	0.838235
7	EU	0.847926	0.754098	0.968421
8	Foreign Policy	0.715447	0.590604	0.907216
9	Health Care	0.833333	0.797872	0.872093
10	Immigration & Religion	0.869010	0.824242	0.918919
11	Infrastructure	0.718563	0.740741	0.697674
12	Labour	0.789062	0.748148	0.834711
13	Social Policy & Welfare	0.717949	0.666667	0.777778
14	Science & Technology	0.508475	0.394737	0.714286
15	Non Topical	0.862682	0.824074	0.905085

Table A.13: **Topicwise classification metrics**. The table shows the test classification metrics for each topic classified with thefinetuned_dfm-encoder-larger-v1 as described in 3.4.2.

After having finetuned a version of each model, we then retrain all of them to classify the topic labels, as well as a multilingual BERT base model for comparison. All these 9 models are retrained by adding a fully connected classification layer on top of the models and using a binary cross entropy loss function to train them. We train the models for 10 epochs on the same training set using a learning rate scheduler, a weighted loss function, and other standard training steps. After training is complete, the best models are chosen based on the validation set, and them metrics for each topic and collectively are calculated based on the test set as described in section 3.4.2. In table A.14 we can see the overall metrics for each model. Here we see that the DFM encoder model performs best overall, and that the finetuned version performs slightly better at F1 score, which is why we chose this model. We also see that finetuning some of the models actually hurt performance when retraining them for classification, as seen with the NordicBERT models (nb-bert-large).

Model	Loss	Recall	Precision	F1
finetuned_danish-bert-botxo	0.005637	0.726843	0.794414	0.742939
danish-bert-botxo	0.005977	0.694334	0.773695	0.716277
finetuned_dfm-encoder-large-v1	0.005147	**0.778182**	0.824691	**0.784190**
dfm-encoder-large-v1	**0.005029**	0.776351	**0.826007**	0.783409
finetuned_nb-bert-large	0.014063	0.000000	0.000000	0.000000
nb-bert-large	0.005225	0.747104	0.813416	0.762435
finetuned_ScandiBERT	0.006059	0.749679	0.806811	0.759444
ScandiBERT	0.006059	0.749679	0.806811	0.759444
bert-base-multilingual-cased	0.006848	0.658734	0.727908	0.676580

Table A.14: **Classification metrics.** All metrics are calculated as sample averages, meaning the average of the metric calculated for each sample. The prefix "finetuned" here refers finetuning the base models to our data, using further masked language training on the full dataset. All models have been finetuned to the classification task afterward.

6.2.3 Pooled standard deviations

Topic	Likes	Comments	Shares
Agriculture, Environment & Food	0.371567	0.209619	0.178017
Crime & Justice	0.410479	0.263840	0.214796
Culture	0.439764	0.232725	0.159705
Defence	0.223037	0.132752	0.107494
Economy & Business	0.503183	0.312762	0.243658
Education	0.408421	0.235093	0.180539
Climate & Energy	0.303536	0.160614	0.135898
EU	0.360753	0.213525	0.171475
Foreign Policy	0.478505	0.261015	0.207270
Health Care	0.333612	0.195954	0.156041
Immigration and Religion	0.623845	0.406055	0.367090
Infrastructure	0.308907	0.182418	0.128714
Labour	0.515943	0.330398	0.260403
Social Policy & Welfare	0.477754	0.284318	0.242885
Science & Technology	0.218612	0.114087	0.091423

Table A.15: **Pooled std. dev. of feedback advantage by topic for Likes, Comments, and Shares.** Pooled standard deviation of feedback advantage for each topic and feedback type. The standard deviations are pooled on politicians.

6.2.4 Regression tables for main specification

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.136**	0.116**	0.146**	0.082**	0.156**
	(0.01)	(0.005)	(0.009)	(0.008)	(0.006)
Topical interest	-0.063**	-0.059*	-0.123	-0.025	-0.031
	(0.024)	(0.025)	(0.069)	(0.028)	(0.02)
Topic popularity	0.249**	0.25**	0.193**	0.383**	0.151**
	(0.032)	(0.022)	(0.014)	(0.034)	(0.024)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
$R^2 within$	0.138	0.124	0.135	0.119	0.135
F Stat.	6615.481	5556.021	6399.634	5574.016	6208.807
Note:	*p<0.05; **p<0.01				

Table A.16: **Regression models for main specification with likes (1/3)**

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.138**	0.142**	0.141**	0.124**	0.13**
	(0.007)	(0.011)	(0.008)	(0.008)	(0.01)
Topical interest	-0.026	0.011	-0.014	-0.012	-0.013
	(0.026)	(0.03)	(0.028)	(0.024)	(0.025)
Topic popularity	0.255**	0.25**	0.21**	0.244**	0.251**
	(0.024)	(0.044)	(0.021)	(0.027)	(0.029)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.145	0.123	0.141	0.158	0.13
F Stat.	6964.208	5818.621	6248.964	7448.473	6274.397
Note:	*p<0.05; **p<0.01				

Table A.17: **Regression models for main specification with likes (2/3)**

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.114**	0.136**	0.153**	0.147**	0.107**
	(0.006)	(0.009)	(0.008)	(0.006)	(0.013)
Topical interest	-0.01	-0.038	-0.056*	-0.055**	0.001
	(0.019)	(0.03)	(0.025)	(0.02)	(0.023)
Topic popularity	0.231**	0.241**	0.159**	0.174**	0.421**
	(0.023)	(0.028)	(0.018)	(0.021)	(0.052)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.12	0.128	0.122	0.131
F Stat.	6561.887	5719.188	5878.412	5725.522	6427.93
Note:	*p<0.05; **p<0.01				

Table A.18: Regression models for main specification with likes (3/3)

6.2.5 Main specification with shares

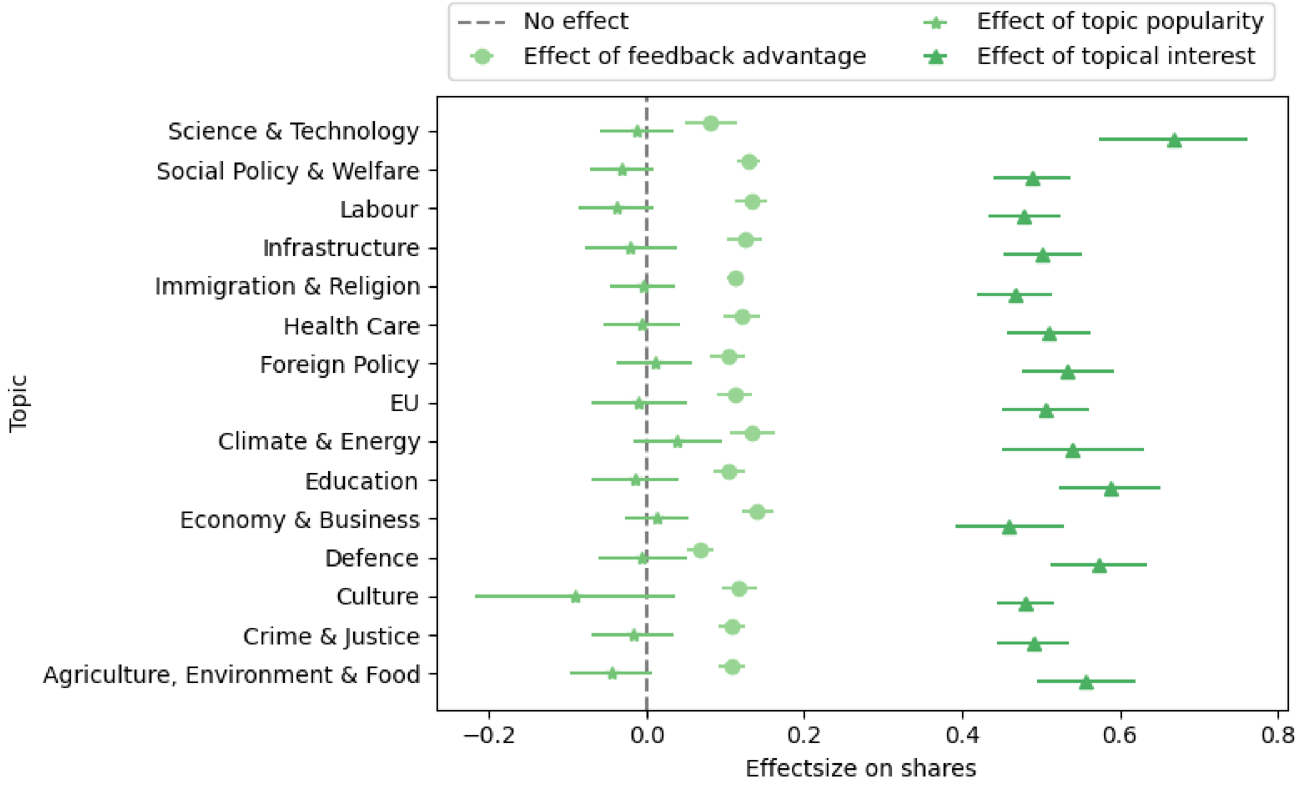


Figure A.11: Effect size of feedback advantage, topic popularity and topical interest for shares on topic probability. This figure depicts the effect size of feedback advantage for predicting topic i using linear probability models. The models are estimated with topic popularity and topical interest, fixed effects on the individual and monthly basis and clustered standard errors.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.108**	0.108**	0.118**	0.068**	0.141**
	(0.009)	(0.008)	(0.012)	(0.008)	(0.01)
Topical interest	-0.045	-0.017	-0.09	-0.006	0.013
	(0.027)	(0.026)	(0.065)	(0.028)	(0.021)
Topic popularity	0.557**	0.49**	0.48**	0.573**	0.46**
	(0.032)	(0.023)	(0.018)	(0.031)	(0.035)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.125	0.116	0.117	0.114	0.116
F Stat.	5888.555	5126.1	5423.604	5256.584	5201.345
Note:	*p<0.05; **p<0.01				

Table A.19: Regression models for main specification with shares (1/3)

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.104**	0.134**	0.112**	0.103**	0.121**
	(0.01)	(0.015)	(0.011)	(0.011)	(0.011)
Topical interest	-0.015	0.039	-0.009	0.01	-0.006
	(0.028)	(0.029)	(0.031)	(0.024)	(0.025)
Topic popularity	0.588**	0.541**	0.507**	0.534**	0.51**
	(0.033)	(0.046)	(0.028)	(0.03)	(0.027)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.131	0.111	0.127	0.144	0.118
F Stat.	6139.114	5149.814	5470.254	6637.884	5632.72
Note:	*p<0.05; **p<0.01				

Table A.20: Regression models for main specification with shares (2/3)

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.112**	0.124**	0.133**	0.128**	0.081**
	(0.005)	(0.012)	(0.011)	(0.008)	(0.017)
Topical interest	-0.005	-0.02	-0.038	-0.032	-0.013
	(0.021)	(0.029)	(0.024)	(0.02)	(0.024)
Topic popularity	0.467**	0.502**	0.479**	0.489**	0.668**
	(0.024)	(0.026)	(0.023)	(0.024)	(0.048)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.141	0.108	0.111	0.11	0.124
F Stat.	5990.252	5055.345	5025.696	5054.468	5963.978
Note:	*p<0.05; **p<0.01				

Table A.21: Regression models for main specification with shares (3/3)

6.2.6 Main specification with comments

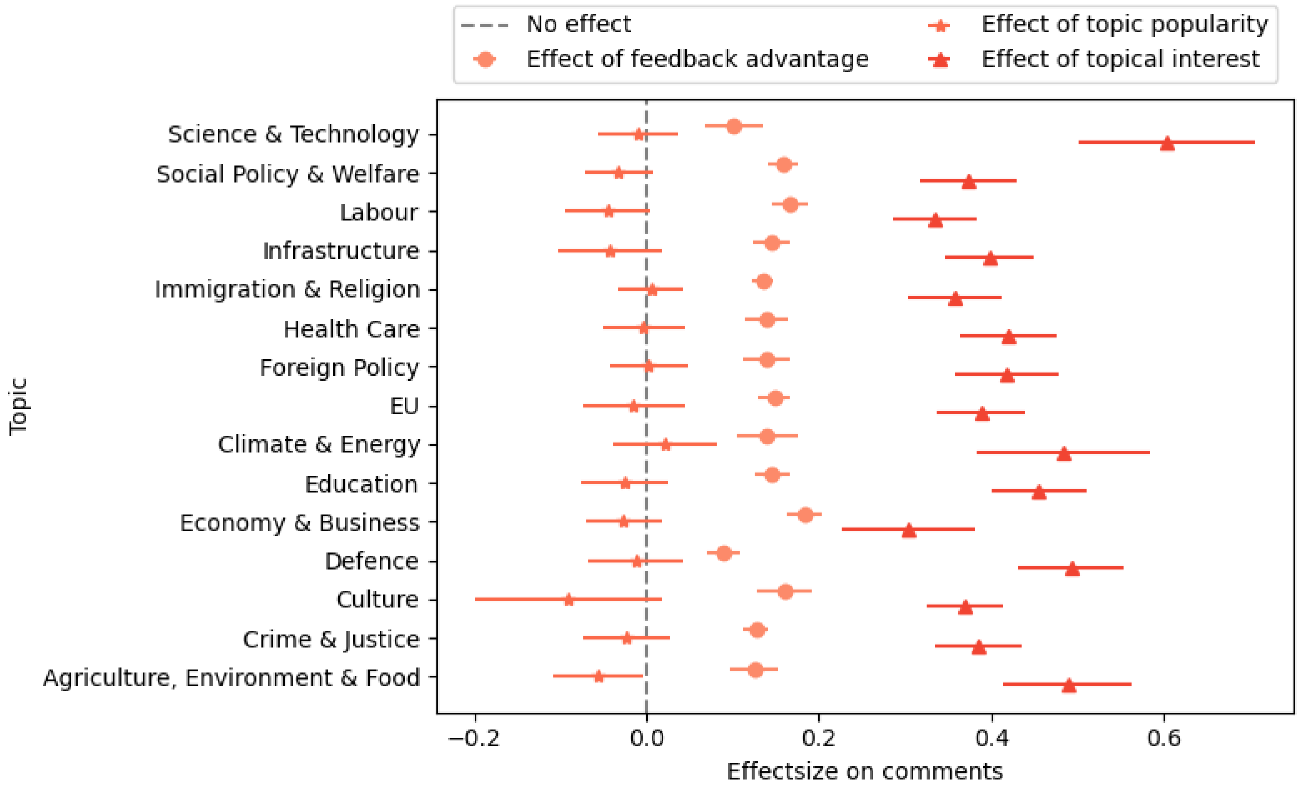


Figure A.12: Effect size of feedback advantage, topic popularity and topical interest for comments on topic probability. This figure depicts the effect size of feedback advantage for predicting topic i using linear probability models. The models are estimated with topic popularity and topical interest, fixed effects on the individual and monthly basis and clustered standard errors.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.125**	0.128**	0.16**	0.089**	0.183**
	(0.014)	(0.007)	(0.016)	(0.01)	(0.01)
Topical interest	-0.057*	-0.023	-0.091	-0.012	-0.026
	(0.027)	(0.026)	(0.055)	(0.028)	(0.023)
Topic popularity	0.489**	0.386**	0.369**	0.493**	0.304**
	(0.038)	(0.026)	(0.022)	(0.031)	(0.039)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.128	0.119	0.124	0.117	0.126
F Stat.	6030.863	5309.02	5800.446	5392.222	5700.674
Note:	*p<0.05; **p<0.01				

Table A.22: **Regression models for main specification with comments (1/3)**

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.146**	0.14**	0.148**	0.139**	0.14**
	(0.01)	(0.018)	(0.009)	(0.013)	(0.013)
Topical interest	-0.026	0.021	-0.015	0.003	-0.003
	(0.026)	(0.031)	(0.03)	(0.023)	(0.024)
Topic popularity	0.455**	0.483**	0.388**	0.418**	0.42**
	(0.028)	(0.051)	(0.026)	(0.031)	(0.029)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.136	0.112	0.133	0.149	0.122
F Stat.	6472.492	5222.026	5752.0	6920.25	5850.531
Note:	*p<0.05; **p<0.01				

Table A.23: **Regression models for main specification with comments (2/3)**

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.135**	0.145**	0.166**	0.158**	0.101**
	(0.006)	(0.011)	(0.011)	(0.009)	(0.018)
Topical interest	0.005	-0.043	-0.045	-0.032	-0.01
	(0.019)	(0.03)	(0.025)	(0.021)	(0.023)
Topic popularity	0.357**	0.398**	0.335**	0.374**	0.603**
	(0.028)	(0.026)	(0.024)	(0.028)	(0.052)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.145	0.113	0.119	0.114	0.126
F Stat.	6229.014	5319.162	5400.75	5274.524	6076.537
Note:	*p<0.05; **p<0.01				

Table A.24: **Regression models for main specification with comments (3/3)**

6.2.7 Heterogenous effects of election

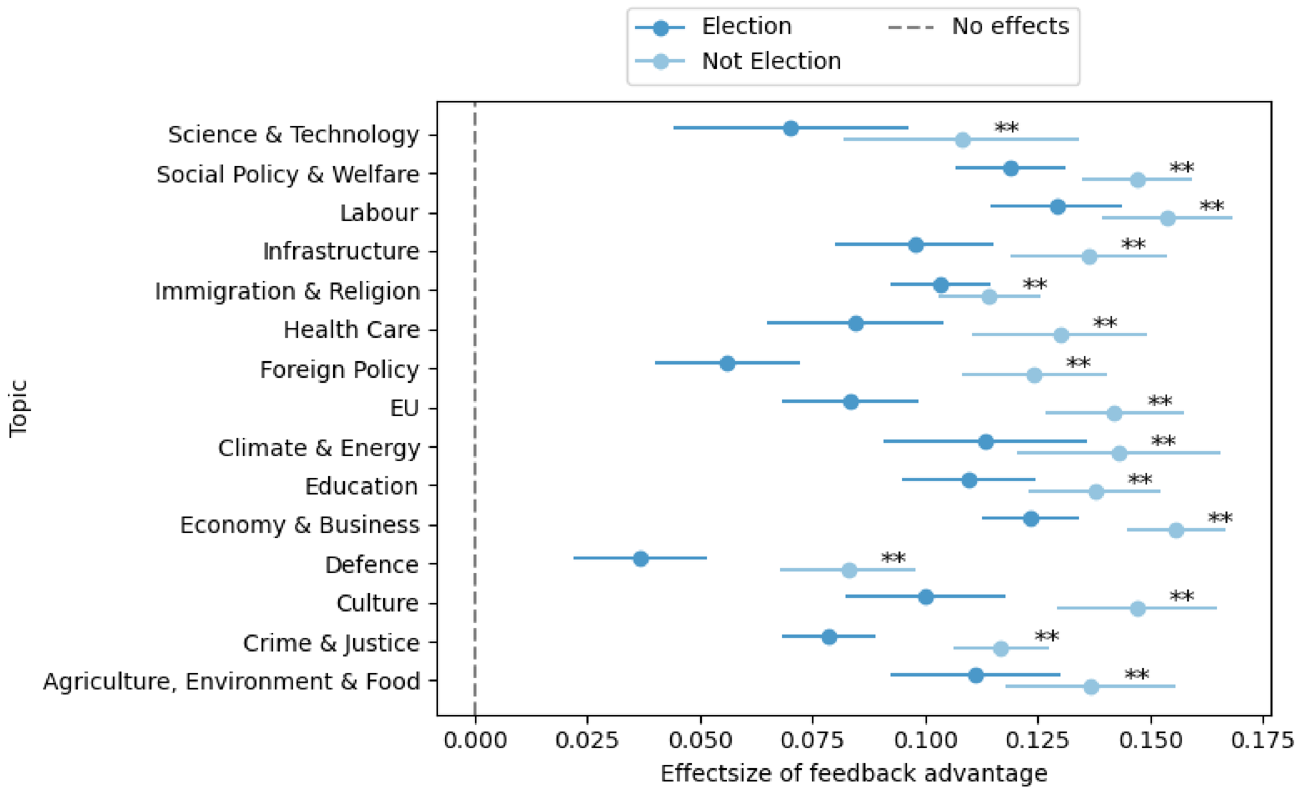


Figure A.13: Heterogeneous effects of of elections (likes). Effect size of feedback advantage (Likes) for predicting topic i during election periods. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.137**	0.117**	0.147**	0.083**	0.156**
	(0.01)	(0.005)	(0.009)	(0.008)	(0.006)
Topical interest	-0.064**	-0.063*	-0.126	-0.026	-0.028
	(0.024)	(0.025)	(0.067)	(0.028)	(0.02)
Topic popularity	0.251**	0.253**	0.197**	0.384**	0.157**
	(0.032)	(0.022)	(0.013)	(0.034)	(0.024)
Election	-0.006*	-0.008*	-0.01**	-0.011**	-0.01*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Feedback advantage * Election	-0.026**	-0.038**	-0.047**	-0.046**	-0.032**
	(0.009)	(0.006)	(0.008)	(0.013)	(0.007)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.138	0.124	0.135	0.12	0.135
F Stat.	3974.618	3343.302	3850.841	3355.353	3733.035
Note:	*p<0.05; **p<0.01				

Table A.25: Regression models with heterogeneous effects of election (likes 1/3)

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.138**	0.143**	0.142**	0.124**	0.13**
	(0.007)	(0.012)	(0.008)	(0.008)	(0.01)
Topical interest	-0.027	0.014	-0.017	-0.013	-0.014
	(0.026)	(0.03)	(0.028)	(0.024)	(0.025)
Topic popularity	0.258**	0.254**	0.208**	0.246**	0.259**
	(0.024)	(0.045)	(0.021)	(0.027)	(0.028)
Election	-0.008**	-0.012**	-0.011**	-0.011**	-0.011**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Feedback advantage * Election	-0.028**	-0.03**	-0.059**	-0.068**	-0.045**
	(0.007)	(0.01)	(0.011)	(0.01)	(0.012)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.146	0.124	0.142	0.159	0.131
F Stat.	4182.964	3499.809	3759.41	4480.857	3781.027
Note:	*p<0.05; **p<0.01				

Table A.26: Regression models with heterogeneous effects of election (likes 2/3)

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.114**	0.137**	0.154**	0.147**	0.108**
	(0.006)	(0.009)	(0.007)	(0.006)	(0.013)
Topical interest	-0.009	-0.04	-0.056*	-0.045*	-0.001
	(0.019)	(0.03)	(0.025)	(0.02)	(0.023)
Topic popularity	0.233**	0.247**	0.162**	0.181**	0.421**
	(0.023)	(0.029)	(0.018)	(0.021)	(0.051)
Election	0.001	-0.008**	0.008*	-0.012**	-0.01*
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Feedback advantage * Election	-0.011**	-0.039**	-0.025**	-0.028**	-0.038**
	(0.003)	(0.005)	(0.005)	(0.005)	(0.014)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.121	0.128	0.123	0.132
F Stat.	3938.454	3443.567	3532.38	3443.688	3862.489
Note:	*p<0.05; **p<0.01				

Table A.27: Regression models with heterogeneous effects of election (likes 3/3)

6.2.8 Heterogenous effects of government

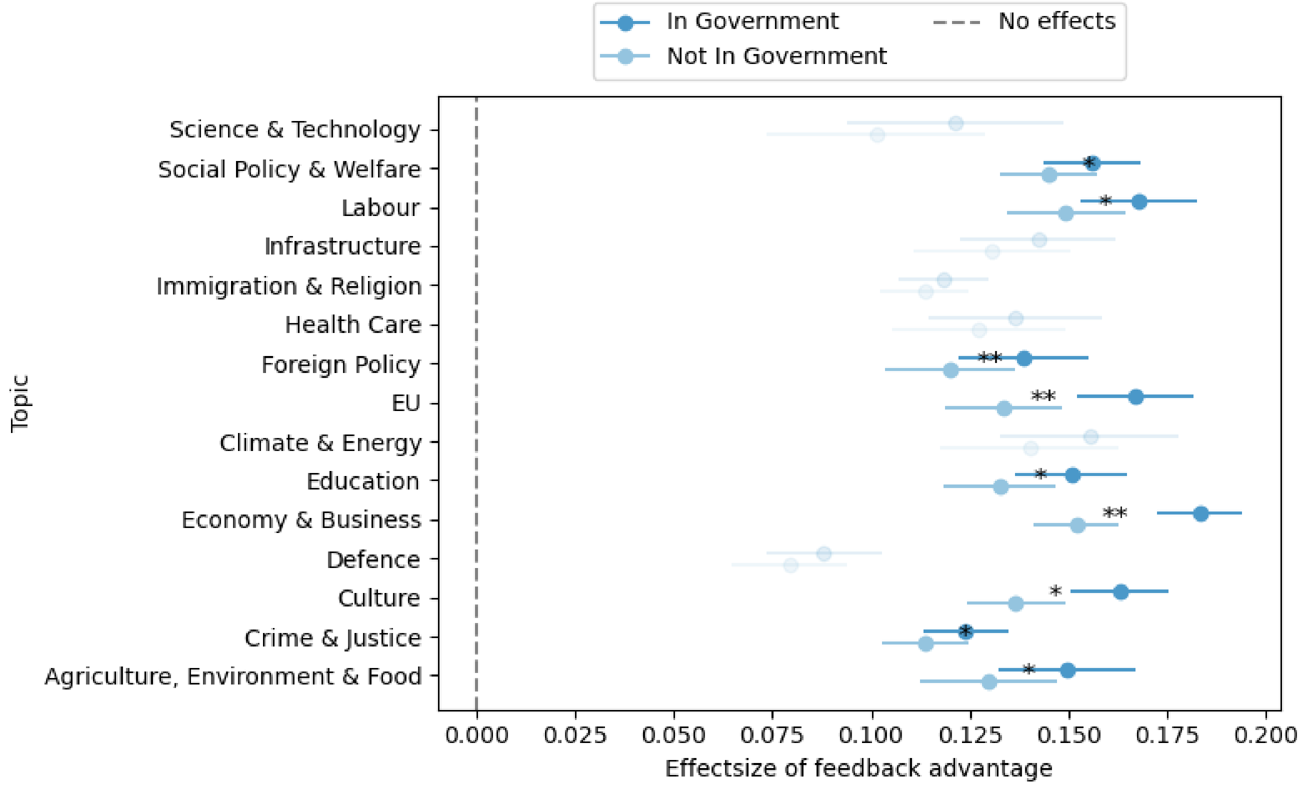


Figure A.14: Heterogeneous effects of government periods (likes). Effect size of feedback advantage (Likes) for predicting topic i during periods when politicians are part of the government or not. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.13**	0.114**	0.137**	0.079**	0.152**
	(0.009)	(0.006)	(0.006)	(0.007)	(0.005)
Topical interest	-0.052*	-0.058*	-0.109*	-0.025	-0.034
	(0.022)	(0.024)	(0.054)	(0.028)	(0.021)
Topic popularity	0.248**	0.248**	0.189**	0.38**	0.137**
	(0.031)	(0.022)	(0.015)	(0.035)	(0.024)
Government	0.003	0.001	0.001	0.002	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)
Feedback advantage * Government	0.02*	0.01*	0.027*	0.009	0.031**
	(0.009)	(0.005)	(0.012)	(0.005)	(0.009)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.138	0.124	0.135	0.12	0.136
F Stat.	3982.038	3337.057	3862.22	3347.522	3752.011
Note:	*p<0.05; **p<0.01				

Table A.28: Regression models with heterogeneous effects of government (likes 1/3)

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.133**	0.14**	0.134**	0.12**	0.127**
	(0.007)	(0.012)	(0.008)	(0.008)	(0.011)
Topical interest	-0.023	0.012	-0.001	-0.007	-0.014
	(0.026)	(0.03)	(0.026)	(0.023)	(0.025)
Topic popularity	0.25**	0.246**	0.198**	0.236**	0.249**
	(0.021)	(0.044)	(0.019)	(0.028)	(0.028)
Government	0.003	0.004*	0.004*	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Feedback advantage * Government	0.018*	0.015	0.033**	0.018**	0.009
	(0.008)	(0.008)	(0.01)	(0.007)	(0.009)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.146	0.123	0.142	0.159	0.13
F Stat.	4192.425	3496.543	3778.3	4482.914	3767.535
Note:	*p<0.05; **p<0.01				

Table A.29: Regression models with heterogeneous effects of government (likes 2/3)

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.114**	0.131**	0.149**	0.145**	0.101**
	(0.006)	(0.01)	(0.008)	(0.006)	(0.014)
Topical interest	-0.01	-0.037	-0.056*	-0.056**	0.005
	(0.019)	(0.03)	(0.025)	(0.02)	(0.023)
Topic popularity	0.231**	0.24**	0.153**	0.171**	0.413**
	(0.024)	(0.027)	(0.018)	(0.02)	(0.051)
Government	-0.0	0.001	-0.001	0.001	0.005
	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
Feedback advantage * Government	0.005	0.012	0.018*	0.011*	0.02
	(0.008)	(0.006)	(0.008)	(0.005)	(0.013)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.12	0.128	0.122	0.132
F Stat.	3937.96	3435.458	3536.312	3438.43	3866.668
Note:	*p<0.05; **p<0.01				

Table A.30: Regression models with heterogeneous effects of government (likes 3/3)

6.2.9 Heterogenous effects of gender

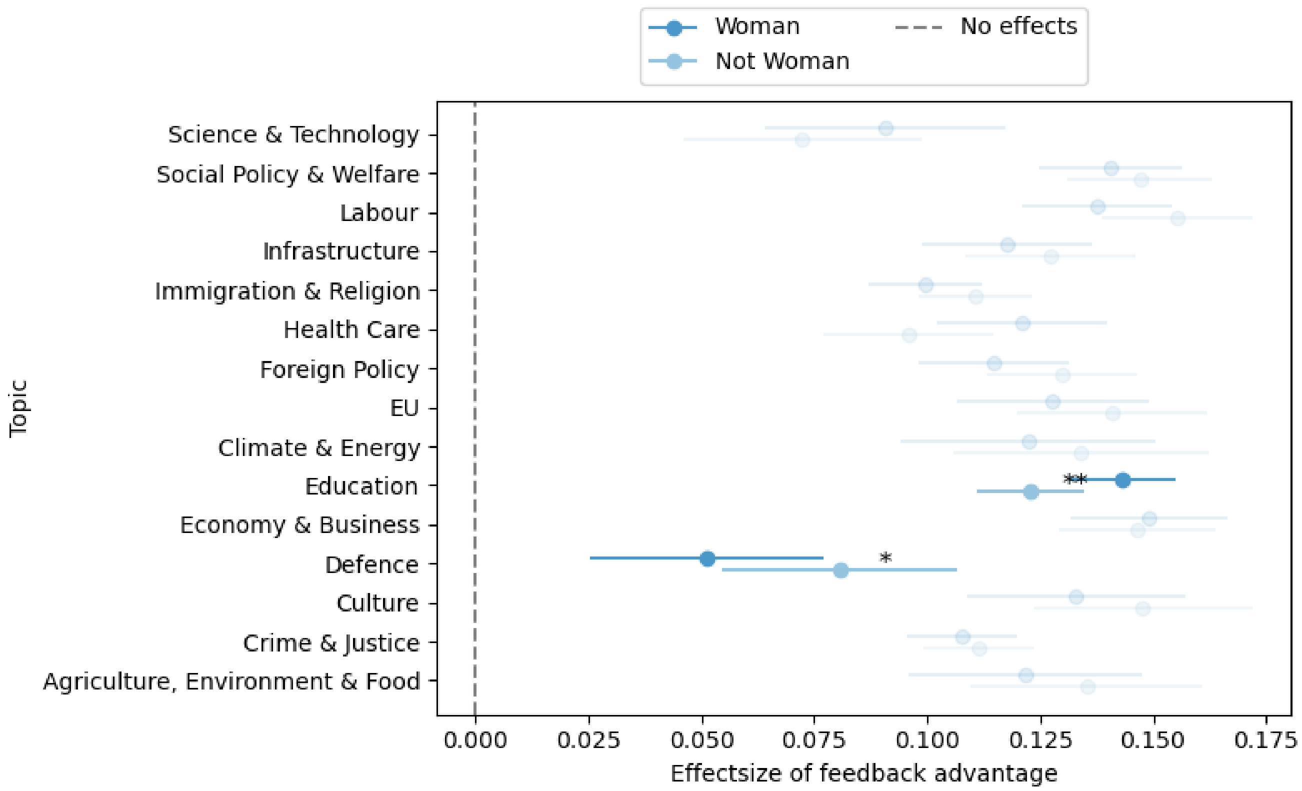


Figure A.15: Heterogeneous effects of gender (likes). Effect size of feedback advantage (Likes) for predicting topic i based on gender. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.135**	0.112**	0.148**	0.081**	0.146**
	(0.013)	(0.006)	(0.012)	(0.013)	(0.009)
Topical interest	-0.086**	-0.06*	-0.124	-0.038	-0.025
	(0.024)	(0.026)	(0.065)	(0.029)	(0.023)
Topic popularity	0.312**	0.278**	0.207**	0.448**	0.174**
	(0.04)	(0.023)	(0.016)	(0.039)	(0.027)
Gender	0.002	0.003	0.0	-0.005	0.0
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Feedback advantage * Gender	-0.014	-0.004	-0.015	-0.029*	0.003
	(0.011)	(0.007)	(0.015)	(0.012)	(0.014)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.137	0.124	0.135	0.119	0.135
F Stat.	7372.376	4678.747	5412.539	5284.174	5470.662
Note:	*p<0.05; **p<0.01				

Table A.31: Regression models with heterogeneous effects of gender (likes 1/3)

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.123**	0.134**	0.141**	0.13**	0.096**
	(0.006)	(0.014)	(0.011)	(0.008)	(0.01)
Topical interest	-0.039	-0.001	-0.035	-0.024	-0.025
	(0.027)	(0.03)	(0.028)	(0.026)	(0.026)
Topic popularity	0.288**	0.323**	0.252**	0.269**	0.341**
	(0.027)	(0.053)	(0.02)	(0.034)	(0.033)
Gender	0.006*	0.001	0.001	0.003	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Feedback advantage * Gender	0.02**	-0.012	-0.013	-0.015	0.025
	(0.008)	(0.012)	(0.013)	(0.009)	(0.013)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.145	0.123	0.141	0.158	0.129
F Stat.	8003.291	5583.095	5740.099	7332.851	6602.699
Note:	*p<0.05; **p<0.01				

Table A.32: Regression models with heterogeneous effects of gender (likes 2/3)

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.111**	0.127**	0.155**	0.147**	0.072**
	(0.006)	(0.01)	(0.009)	(0.008)	(0.014)
Topical interest	-0.012	-0.041	-0.069**	-0.063**	-0.01
	(0.022)	(0.031)	(0.025)	(0.02)	(0.024)
Topic popularity	0.271**	0.272**	0.187**	0.198**	0.505**
	(0.024)	(0.031)	(0.022)	(0.026)	(0.055)
Gender	0.003	-0.0	0.001	0.005	0.006
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Feedback advantage * Gender	-0.011	-0.01	-0.018	-0.007	0.018
	(0.008)	(0.01)	(0.01)	(0.009)	(0.014)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.12	0.128	0.122	0.131
F Stat.	8282.497	4775.883	5440.325	4679.94	4538.125
Note:	*p<0.05; **p<0.01				

Table A.33: Regression models with heterogeneous effects of gender (likes 3/3)

6.2.10 Heterogenous effects of vote count

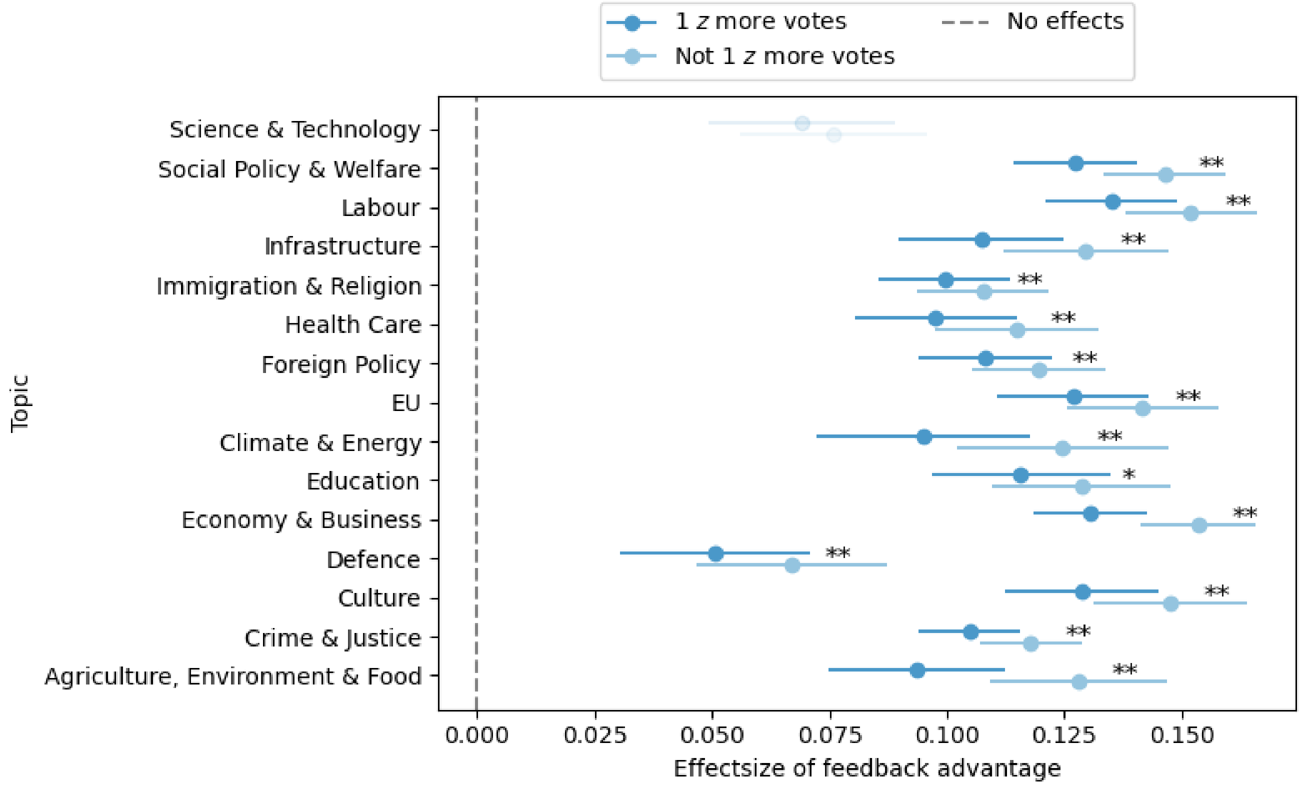


Figure A.16: Heterogeneous effects vote count (likes). Effect size of feedback advantage (Likes) for predicting topic i based on vote counts at last election. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.128**	0.118**	0.148**	0.067**	0.153**
	(0.010)	(0.006)	(0.008)	(0.010)	(0.006)
Topical interest	-0.083**	-0.057*	-0.116*	-0.035	-0.010
	(0.024)	(0.027)	(0.053)	(0.031)	(0.023)
Topic popularity	0.293**	0.258**	0.209**	0.483**	0.160**
	(0.034)	(0.026)	(0.019)	(0.039)	(0.035)
Vote count	-0.007**	-0.003**	-0.003**	-0.003**	-0.005**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Feedback advantage * Vote count	-0.034**	-0.013**	-0.019**	-0.016**	-0.023**
	(0.005)	(0.002)	(0.003)	(0.004)	(0.005)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	121133	121133	121133	121133	121133
R^2 within	0.139	0.125	0.136	0.120	0.136
F Stat.	7225.647	4500.718	5296.426	5117.499	5310.017
Note:	*p<0.05; **p<0.01				

Table A.34: **Regression models with heterogeneous effects of vote count (likes 1/3)**

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.129**	0.125**	0.142**	0.119**	0.115**
	(0.010)	(0.012)	(0.008)	(0.007)	(0.009)
Topical interest	-0.042	0.004	-0.042	-0.052	-0.017
	(0.028)	(0.031)	(0.028)	(0.027)	(0.034)
Topic popularity	0.307**	0.334**	0.248**	0.305**	0.343**
	(0.032)	(0.047)	(0.026)	(0.030)	(0.037)
Vote count	-0.002*	-0.006**	-0.002*	-0.002	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Feedback advantage * Vote count	-0.013*	-0.030**	-0.015**	-0.011**	-0.017**
	(0.005)	(0.005)	(0.003)	(0.004)	(0.002)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	121133	121133	121133	121133	121133
R^2 within	0.147	0.124	0.143	0.160	0.130
F Stat.	7613.331	5466.427	5526.979	7169.332	6397.917
Note:	*p<0.05; **p<0.01				

Table A.35: **Regression models with heterogeneous effects of vote count (likes 2/3)**

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.108**	0.130**	0.152**	0.146**	0.076**
	(0.007)	(0.009)	(0.007)	(0.007)	(0.010)
Topical interest	-0.011	-0.037	-0.059*	-0.060**	-0.017
	(0.023)	(0.031)	(0.026)	(0.021)	(0.027)
Topic popularity	0.268**	0.251**	0.176**	0.193**	0.540**
	(0.029)	(0.034)	(0.024)	(0.027)	(0.041)
Vote count	-0.004**	-0.005**	-0.004**	-0.004**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Feedback advantage * Vote count	-0.008**	-0.022**	-0.017**	-0.019**	-0.006
	(0.003)	(0.004)	(0.003)	(0.002)	(0.008)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	121133	121133	121133	121133	121133
R^2 within	0.152	0.122	0.128	0.123	0.132
F Stat.	7981.343	4594.193	5214.986	4471.706	4360.900
Note:	*p<0.05; **p<0.01				

Table A.36: Regression models with heterogeneous effects of vote count (likes 3/3)

6.2.11 Heterogenous effects of ideology

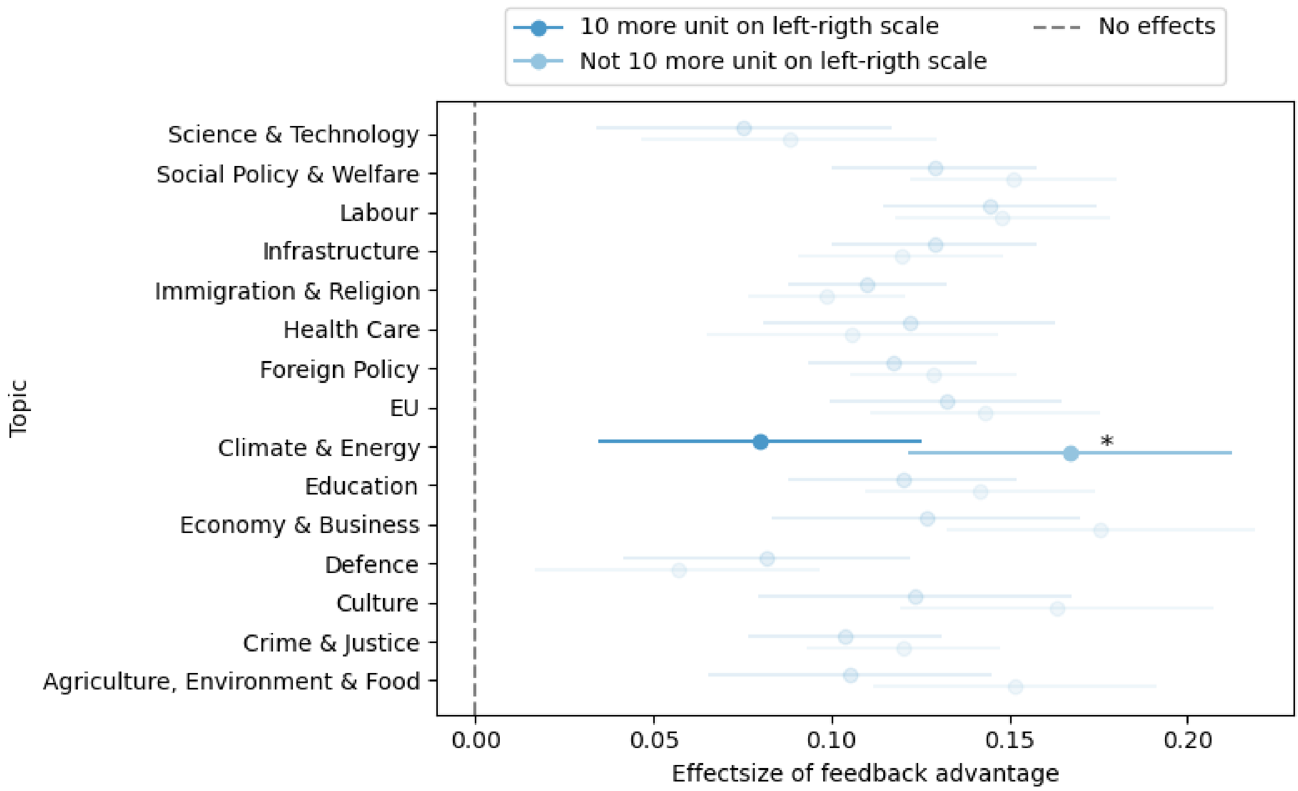


Figure A.17: Heterogeneous effects of ideology (likes). Effect size of feedback advantage (Likes) for predicting topic i based on ideology. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	**Agriculture, Environment & Food**	**Crime & Justice**	**Culture**	**Defence**	**Economy & Business**
Feedback advantage	0.152**	0.120**	0.163**	0.057**	0.176**
	(0.020)	(0.014)	(0.022)	(0.020)	(0.022)
Topical interest	-0.086**	-0.064*	-0.125	-0.041	-0.034
	(0.026)	(0.026)	(0.068)	(0.029)	(0.022)
Topic popularity	0.311**	0.277**	0.209**	0.458**	0.177**
	(0.038)	(0.024)	(0.016)	(0.039)	(0.029)
Ideology	-0.002	-0.001	-0.002*	0.000	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Feedback advantage * Ideology	-0.005	-0.002	-0.004	0.002	-0.005
	(0.004)	(0.002)	(0.004)	(0.004)	(0.004)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.137	0.124	0.135	0.119	0.135
F Stat.	7386.937	4679.763	5415.302	5255.470	5495.594
Note:	*p<0.05; **p<0.01				

Table A.37: **Regression models with heterogeneous effects of Ideology (likes 1/3)**

Topic	**Education**	**Climate & Energy**	**EU**	**Foreign Policy**	**Health Care**
Feedback advantage	0.142**	0.167**	0.143**	0.129**	0.106**
	(0.016)	(0.023)	(0.017)	(0.012)	(0.021)
Topical interest	-0.042	0.004	-0.039	-0.029	-0.035
	(0.027)	(0.030)	(0.028)	(0.025)	(0.029)
Topic popularity	0.303**	0.314**	0.254**	0.273**	0.343**
	(0.029)	(0.052)	(0.023)	(0.035)	(0.029)
Ideology	-0.001	-0.002**	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Feedback advantage * Ideology	-0.002	-0.009*	-0.001	-0.001	0.002
	(0.003)	(0.004)	(0.003)	(0.002)	(0.005)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.145	0.123	0.141	0.158	0.129
F Stat.	7977.254	5616.391	5737.559	7321.260	6575.462
Note:	*p<0.05; **p<0.01				

Table A.38: **Regression models with heterogeneous effects of Ideology (likes 2/3)**

Topic	**Immigration & Religion**	**Infrastructure**	**Labour**	**Social Policy & Welfare**	**Science & Technology**
Feedback advantage	0.099**	0.119**	0.148**	0.151**	0.088**
	(0.011)	(0.015)	(0.015)	(0.015)	(0.021)
Topical interest	-0.013	-0.042	-0.069**	-0.065**	-0.013
	(0.021)	(0.031)	(0.025)	(0.020)	(0.024)
Topic popularity	0.277**	0.273**	0.196**	0.208**	0.507**
	(0.025)	(0.031)	(0.020)	(0.026)	(0.055)
Ideology	-0.001	-0.001	0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Feedback advantage * Ideology	0.001	0.001	0.000	-0.002	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.120	0.128	0.122	0.131
F Stat.	8270.668	4775.074	5426.819	4683.336	4525.779
Note:	*p<0.05; **p<0.01				

Table A.39: Regression models with heterogeneous effects of Ideology (likes 3/3)

6.2.12 Heterogenous effects of ideological extremety

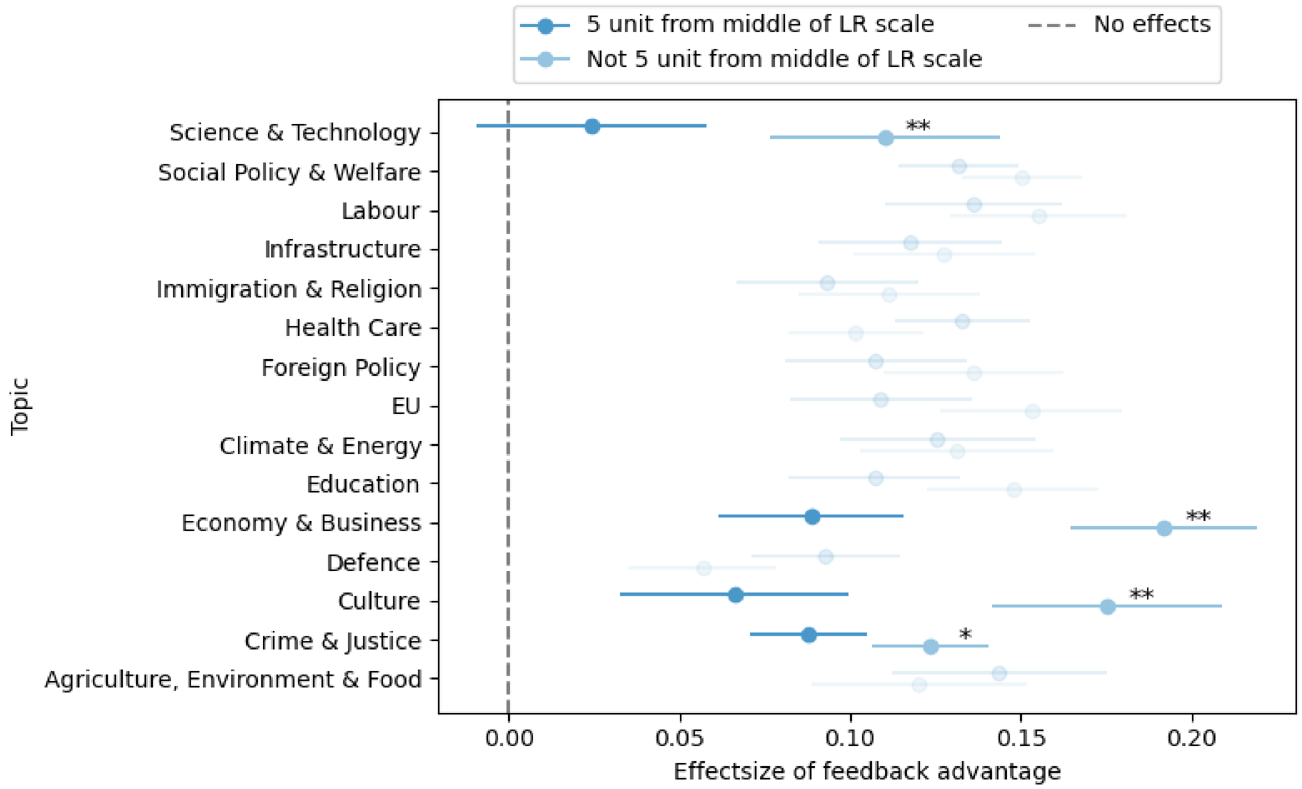


Figure A.18: Heterogeneous effects of ideological extremity (likes). Effect size of feedback advantage (Likes) for predicting topic i based on ideological extremity. The stars indicate the statistical significance of the interaction term between feedback advantage and the heterogeneous effect variable. $* = P <= 0.05$, $** = P <= 0.01$.

Topic	Agriculture, Environment & Food	Crime & Justice	Culture	Defence	Economy & Business
Feedback advantage	0.120**	0.124**	0.175**	0.057**	0.192**
	(0.016)	(0.009)	(0.017)	(0.011)	(0.014)
Topical interest	-0.087**	-0.062*	-0.104*	-0.040	-0.029
	(0.025)	(0.026)	(0.050)	(0.030)	(0.022)
Topic popularity	0.316**	0.274**	0.200**	0.461**	0.154**
	(0.041)	(0.024)	(0.017)	(0.039)	(0.031)
Ideology extremity	0.002	0.000	-0.002	0.003	-0.001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Feedback advantage *	0.005	-0.007*	-0.022**	0.007	-0.021**
Ideology extremity	(0.009)	(0.003)	(0.008)	(0.006)	(0.006)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.137	0.124	0.136	0.119	0.136
F Stat.	7368.436	4682.637	5459.183	5259.977	5525.627
Note:	*p<0.05; **p<0.01				

Table A.40: Regression models with heterogeneous effects of Ideology extremity (likes 1/3)

Topic	Education	Climate & Energy	EU	Foreign Policy	Health Care
Feedback advantage	0.148**	0.131**	0.153**	0.136**	0.102**
	(0.013)	(0.015)	(0.013)	(0.013)	(0.010)
Topical interest	-0.044	-0.004	-0.035	-0.030	-0.033
	(0.027)	(0.030)	(0.027)	(0.025)	(0.028)
Topic popularity	0.295**	0.325**	0.247**	0.262**	0.347**
	(0.026)	(0.053)	(0.022)	(0.034)	(0.036)
Ideology extremity	0.000	0.002	0.001	0.000	0.003*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Feedback advantage *	-0.008	-0.001	-0.009	-0.006	0.006
Ideology extremity	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.145	0.123	0.141	0.159	0.129
F Stat.	7984.434	5580.600	5744.036	7328.195	6580.462
Note:	*p<0.05; **p<0.01				

Table A.41: Regression models with heterogeneous effects of Ideology extremity (likes 2/3)

Topic	Immigration & Religion	Infrastructure	Labour	Social Policy & Welfare	Science & Technology
Feedback advantage	0.111**	0.128**	0.155**	0.150**	0.110**
	(0.014)	(0.014)	(0.013)	(0.009)	(0.017)
Topical interest	-0.017	-0.042	-0.071**	-0.064**	-0.012
	(0.022)	(0.030)	(0.025)	(0.020)	(0.024)
Topic popularity	0.277**	0.272**	0.190**	0.199**	0.494**
	(0.027)	(0.031)	(0.023)	(0.026)	(0.054)
Ideology extremity	0.000	0.002	0.000	0.001	-0.003
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Feedback advantage *	-0.004	-0.002	-0.004	-0.004	-0.017**
Ideology extremity	(0.005)	(0.007)	(0.005)	(0.004)	(0.007)
Time Fixed Effects (month)	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓
Obs.	128204	128204	128204	128204	128204
R^2 within	0.151	0.120	0.128	0.122	0.132
F Stat.	8273.571	4776.083	5429.172	4678.330	4552.539
Note:	*p<0.05; **p<0.01				

Table A.42: Regression models with heterogeneous effects of Ideology extremity (likes 3/3)

Bibliography

Baumgartner, Frank R., Christian Breunig, and Emiliano Grossman (Mar. 2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press. eprint: https://academic.oup.com/book/11237/book-pdf/50004199/9780192572226_web.pdf.

Green-Pedersen, Christoffer (2019). *Danish Policy Agendas Codebook (New Version)*.

Green-Pedersen, Christoffer and Peter B. Mortensen (Mar. 2019). „82The Danish Agendas Project". In: *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press. eprint: https://academic.oup.com/book/0/chapter/159764743/chapter-pdf/50004241/oso-9780198835332-chapter-8.pdf.

6.3 Appendix for Paper 3

6.3.1 Materials

To illustrate the use of dialectograms, we collected text from supporters of the Democratic and the Republican party on the social media platform Reddit. Reddit is a platform structured around numerous topical message boards, so-called subreddits, where users post and comment on messages related to the subreddit's topic. On Reddit, support for the Democratic and the Republican party is centred around the two openly partisan subreddits r/Democrats and r/Republican (r/Republican, 2022; r/Democrats, 2022). We picked these subreddits, as we expected them to contain English-language comments on similar topics, but based on different opinions and conceptions of politics.

We collect all comments made in the two subreddits from the Pushshift Reddit archive (Jason Micheal Baumgartner, 2022), using the PMAW API wrapper (Matthew Podolak, 2022). The initial dataset contains 903.024 comments from r/Democrats and 1.038.151 comments from r/Republican, covering the period from January 2011 to September 2022. We drop 142.909 comments from r/Democrats and 106.239 comments from r/Republican, that were either marked as deleted by the user or removed by a moderator.

Having collected the comments, our aim is to create two corpora, based on which we can assess how the two different subreddits use the same words differently. While part of the differences between speech communities have to do with discourse, that is specific to one or the other, that is not our focus here. In fact, we will focus on the exact opposite; how the two communities put the same vocabulary to different uses. Hence, we take four overall steps to clean the comments and make the vocabularies identical: we remove repetitive comments; we process the comments to remove minor lexical variation; we restrict each corpus to users who are active in mainly one of the subreddits; and finally, we limit the vocabularies to words that appear at least a 100 times in each corpus. Below, we cover each in turn.

Removing repetitive comments

First, our embedding-based approach is sensitive to the repetition of specific word associations, so we take steps to remove repetitive comments from our corpora. Overall, these comments can be split into two types. On one hand, Reddit has an eco-system of bots, that comment on the subreddits, using almost exactly the same language across different interactions. In many cases, these bots leave a signature in their comments, declaring themselves as such. On another hand, certain users may repeatedly post completely or near-identical comments; this is e.g. the case of moderators as well as some users posting information about how and where to vote during election campaigns (particularly evident on r/Democrats).

To identify such comments, we employed two approaches. We first read through comments containing certain keywords (starting with 'bot' and then branching out to other words and characters often appearing in the identified bot signatures, such as 'beep' and '^^'). In these comments, we identified highly distinct substrings (e.g. "*beep. boop.* I'm a bot") and removed all comments containing any of these substrings. In addition, we use co-occurrence information to identify words that exhibit unusual distributions. Specifically, we plot the frequency of words against the share of the vocabulary with which they co-occur. Words that co-occur with a relatively low share of the vocabulary compared to other words of similar frequency are targets for being repeated. By subsequently reading through comments including these words, we further identify repeated comments. With this procedure, we further remove 28.449 comments from r/Democrats and 69.231 comments from r/Republican.

Comment preprocessing

Second, we then perform a range of preprocessing steps to 'clean' the data further and to make the vocabularies for each of the two corpora reasonably simplified and similar: we remove a few duplicate comments based on their id; convert emojis to text[1]; remove the parts of comments that quote other comments[2]; replace URLs with their domain[3]; remove usernames; align the use of quotation marks; expand contractions[4]; and convert hyphens, brackets, and parentheses to whitespace. Then we use spaCy to tokenize and lemmatise comments, as well as to obtain the part-of-speech for each token. Finally, we convert all tokens to lowercase, remove any left-over possessive "'s" suffixes from tokens, as well as tokens containing whitespaces or non-alphabetic characters, and remove comments with only one token, as no word associations can be learned from this.

While it is computationally expensive to assess how each of these preprocessing steps affects our results, we believe they all serve the purpose of removing minor lexical variation and ensuring that our corpora are maximally comparable at the level of the vocabulary, which is particularly important for the embedding models that do not rely on subword tokens, as they treat words as equally distinct regardless of their lexical similarity.

Active and distinct users

Third, our inductive analysis relies crucially on the sorting of texts into two distinct corpora. The open nature of Reddit however implies that the scraped comments do not necessarily make up closed, homogeneous strands of republican and democratic discourse. In fact, against such an echo-chamber-based view, scholars are increasingly considering how user-based polarisation might be driven by the increased exposure to foreign opinions, that social media and forums like Reddit can generate (Törnberg, 2022). Based on sampling and reading comments made by users who post in both subreddits, we indeed confirm that some degree of cross-partisan debate takes place. Regardless of its polarising effects, such cross-posting distorts our comparison to the extent that partisan users deploy the same partisan discourse in both subreddits, as it makes the use of words appear more similar across subreddits.

[1]Based on the emoji package.
[2]These quotes can be distinguished from ordinary quotes, as they are represented differently in the scraped text. We remove them as they too make comments repeat, in some cases with the reply being significantly shorter than the quoted text.
[3]Based on the tldeextract package.
[4]To do so, we rely on a range of custom-made regex-expressions.

To make the corpora more homogeneous we partition users based on their activity and generate the final corpora based on this user partition. Figure A.19 shows how total user activity across the two subreddits varies with the share of the comments users post in one or the other. Based on this, remove comments from users, that have either posted less than 10 times or posted less than 90% of their comments in one of the two subreddits. This amounts to excluding 219.706 comments from r/Democrats and 285.802 from r/Republicans, which means that we conduct the analysis on a final preprocessed dataset of 511.906 comments from r/Democrats and 576.872 comments from r/Republican.

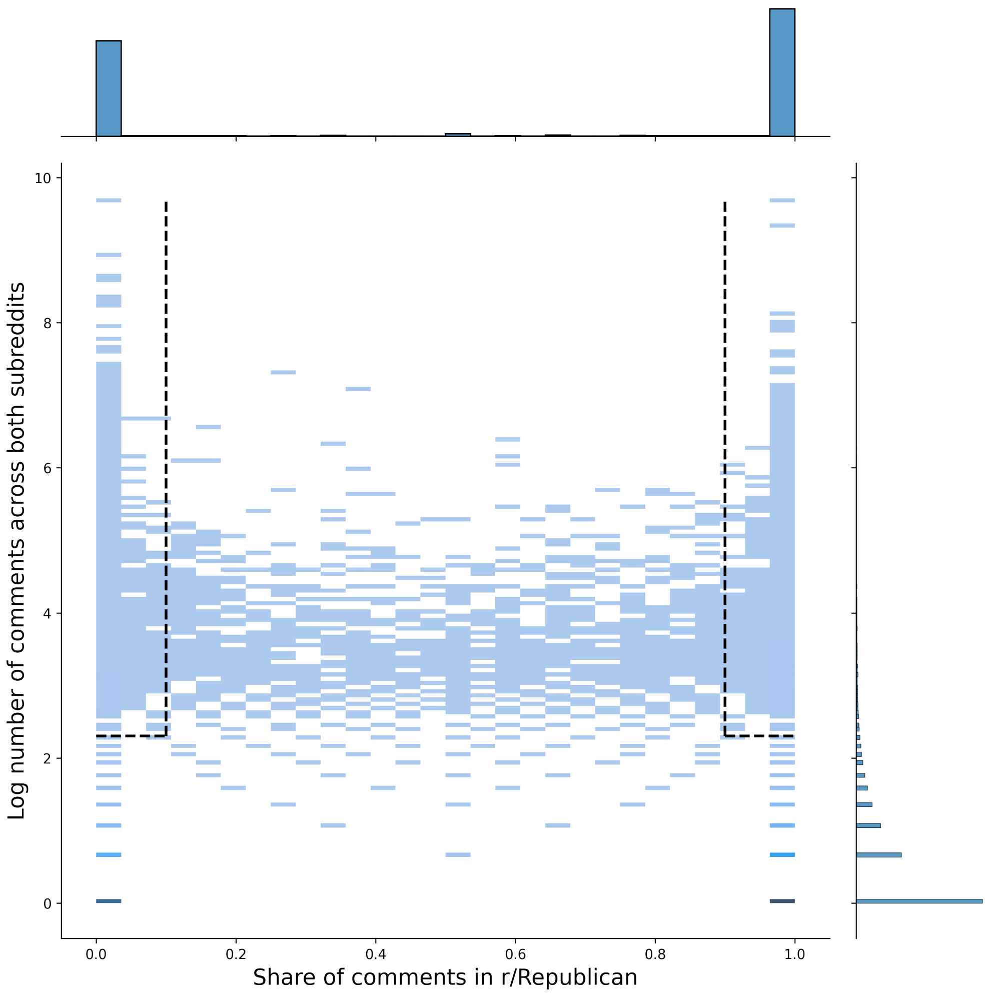


Figure A.19: The log number of times a user posts in the two subreddits combined (y-axis) versus the share of these comments posted in r/Republican (x-axis). Most users post once, and (partially for that reason) most users post only in one of the two subreddits. There is however some degree of activity by users posting moderately and in both subreddits, indicating the kind of cross-commenting that can distort the comparison of the embedding spaces. The dashed lines indicate the users, whose comments make up the final dataset.

Limit the vocabulary

The fourth and final step is to make two identical vocabularies based on the preprocessed datasets. To ensure identical vocabularies and that our interpretations of differences are based on a reasonable number of observations, we remove words from both corpora, that do not occur at least 100 times in each corpus. This leaves us with the same vocabulary of 5.738 unique words in each corpus.

6.3.2 Methods

Here, we describe the dialectograms and the sense separation measure in greater detail. In particular, we describe the GloVe embedding, including the hyperparameters we use to train them; the alignment methods we considered; the measures of difference in use, including the sense separation measure; and our attempts to remove the correlation between aligned cosine distance and word frequency.

GloVe Embeddings

GloVe embeddings are based on minimising the objective

$$\sum_{i=1}^{N}\sum_{j=1}^{N} f(X_{ij}) \cdot [w_i \cdot \tilde{w}_j - [\log(X_{ij}) - b_i - \tilde{b}_j]] \tag{6.2}$$

where X_{ij} is the co-occurrence count between word i and j; $w_i, \tilde{w}_j$ are the word and context vectors of word i and j, each of a pre-specified dimension D; $b_i, \tilde{b}_j$ are the word and context biases of word i and j; N the number of words in the vocabulary; and $f(\cdot)$ a weighting function, which by default is

$$f(x) = \begin{cases} (x/x_{\text{max}})^{\alpha} & \text{if } x < x_{\text{max}} \\ 1 & \text{otherwise} \end{cases} \tag{6.3}$$

for some specified values of x_{max} and α. The final embedding of word i is taken to be the sum of its word and context vector, $w_i + \tilde{w}_i$ (Pennington *et al.*, 2014).

Intuitively vectors and biases are trained, such that the inner product between word and context vectors, $w_i \cdot \tilde{w}_j$, represent the degree to which the words co-occur beyond what would be expected based on their marginal frequencies, $\log(X_{ij}) - b_i - \tilde{b}_j$.[5] However, for word-pairs that co-occur less than x_{max} times, the error in representation is down-weighted according to $(x/x_{\text{max}})^{\alpha} < 1$, so that in the limit of word-pairs that do not co-occur, $X_{ij} = 0$, no loss is obtained. As such, GloVe embeddings are by default not trained to represent non-occurrence.

We use the GloVe implementation made available by the authors of the model[6]. For all embeddings, we construct co-occurrence statistics based on a window size of 10 and train embeddings with 300 dimensions for 30 epochs. All other parameters are set to their default. We normalise the embedding of each word to have unit length, both before and after performing the alignment of the embeddings, equivalent to using cosine distance as opposed to euclidean distance for aligning and translating. We also apply the frequency adjustment described in the section on frequency correlation below, before we align the embeddings.

[5] We expect the biases to capture the marginal word frequencies, which we confirmed is the case, as they exhibit perfect positive correlation.

[6] https://nlp.stanford.edu/projects/glove/

Aligning Embeddings

The dialectograms rely on two aligned embedding spaces. To obtain aligned embedding spaces, several options exist; it is e.g possible to use the first embedding space to initialise the second. Here, we instead consider ways of aligning independently trained embeddings, E_1 and E_2, which we assume are both $N \times D$ matrices, sorted such that each row corresponds to the different embeddings of the same word.

Aligning word embeddings was already addressed in relation to the Word2Vec model, where the authors in a follow-up suggested learning a translation matrix, $W \in \mathbb{R}^{D \times D}$, from one embedding to the other (Mikolov, Le, *et al.*, 2013). Specifically, they solve the following optimisation problem by gradient descent:

$$W_{GD} = \arg\min_{W} \parallel E_1 - E_2 W \parallel_F^2 \tag{6.4}$$

where $\| \cdot \|_F^2$ is the squared Frobenius matrix norm.

The unrestricted nature of the optimisation implies, however, that W_{GD} not only aligns E_2 to the space of E_1 but also distorts the internal relations between the embeddings of words in the E_2 space itself. To see this, consider the matrix of word-to-word inner products, given by

$$E_2 W_{GD} [E_2 W_{GD}]^T = E_2 W_{GD} W_{GD}^T E_2^T \neq E_2 E_2^T \tag{6.5}$$

as long as $W_{GD} W_{GD}^T \neq I$, which is not guaranteed by the optimisation.

One way to impose this restriction is to require that W is orthogonal, i.e. that $WW^T = W^T W = I$. This amounts to requiring that W is a (potentially improper) rotation matrix. In this case, the optimisation problem is known as the Orthogonal Procrustes Problem, and referred to as Procrustes Analysis (PA). The problem was solved analytically in 1966 using Singular Value Decomposition (SVD) Schönemann, 1966 (see also Artetxe *et al.*, 2016). In general, the SVD of a real-valued matrix X is the decomposition $X = USV^T$, where the left and right-singular matrices, U and V are orthogonal. Given this, the solution to the Orthogonal Procrustes Problem is:

$$\begin{aligned} E_2^T E_1 &= USV^T \\ W_{PA} &= \arg \min_{W, WW^T = I} \parallel E_1 - E_2 W \parallel_F^2 \\ &= UV^T \end{aligned} \tag{6.6}$$

where the first line is the SVD of the matrix of inner products between the dimensions of the two embedding spaces. In this case,

$$E_2 W_{PA}[E_2 W_{PA}]^T = E_2 W_{PA} W_{PA}^T E_2^T = E_2 U V^T V U^T E_2^T = E_2 E_2^T \tag{6.7}$$

due to the orthogonality of U and V. The constrained nature of the optimisation implies that W_{PA} provides a worse fit than W_{GD}, but it keeps the structure of the transformed embedding intact.

In performing the alignment, PA will put greater emphasis on aligning dimensions with greater variation or length, as measured by the L2 norm of each dimension j, $\parallel E_{\cdot,j} \parallel_2$. For this reason, it is generally common to demean and unit-normalise the variation of each column before proceeding. As we observe our embeddings to have fairly similar lengths, this is unlikely to play a major role, but it does motivate a third suggestion for how to align embeddings spaces, Canonical Correlation Analysis (CCA).

Whereas PA aligns dimensions based on their scale-dependent inner products, CCA aligns dimensions based on their correlation. As such, CCA can be performed using SVD on the correlation matrix (instead of the inner product matrix, $E_2 E_1^T$) between dimensions. However, calculating the correlation matrix is not necessary - it is instead possible to perform CCA as a two-step procedure based on E_1 and E_2 directly, which in turn clarifies the relation between PA and CCA William, 2011.

Unlike the PA procedure presented above, CCA is typically done by transforming both embeddings, that is, we are looking for two transformation matrices, W_{CCA_1} and W_{CCA_2}, transforming E_1 and E_2 respectively. First, note that based on the SVD of each embedding $E_k = U_k S_k V_k^T$, we can rewrite the SVD of $E_2^T E_1$ used in PA as:

$$E_2^T E_1 = V_2 S_2^T U_2^T U_1 S_1 V_1^T \tag{6.8}$$

While PA aligns embeddings based on the inner product of the embedding dimensions, $E_2^T E_1$, CCA aligns embeddings based on the inner products of their left singular vectors, $U_2^T U_1$. The left singular vectors are orthogonal matrices, i.e. uncorrelated and scaled to unit length. As such, CCA can be interpreted as aligning the unweighted principal components of the two embedding spaces, with transformation matrices given as:

$$\begin{aligned} U_2^T U_1 &= U_* S_* V_*^T \\ W_{CCA_1} &= V_1 S_1^{-1} V_* \\ W_{CCA_2} &= V_2 S_2^{-1} U_* \end{aligned} \tag{6.9}$$

The simultaneous transformation of both embeddings spaces has the additional benefit that the dimensions of the transformed spaces have the interpretation of being sorted according to their correlation - the embeddings have the highest correlation along the first dimension, the second highest along the second dimension and so forth. This is, however, not a feature unique to CCA, as it is possible to reformulate PA to have a similar interpretation, in which the two transformation matrices are

$$\begin{aligned} E_2^T E_1 &= USV^T \\ W_{PA_1} &= V \\ W_{PA_2} &= U \end{aligned} \tag{6.10}$$

In this version of PA, the transformed dimensions are not sorted by their correlation, but by their inner product. Inspecting this further, V has the interpretation of projecting from the E_1 space to the shared space sorted by inner product, while $V^{-1} = V^T$ projects the opposite way (likewise for U and E_2). Comparing this simultaneous version of PA to the single $W_{PA} = UV^T$ shows that W_{PA} performs two projections, first projecting E_2 to the sorted inner product space by U, after which E_2U is further projected to the E_1 space by V^T. This twofold version of PA alignment is what we use here.

In some applications, the embeddings are aligned based on a subset of words, such as high-frequent stop words, that are assumed to be used similarly across the corpora. In our case, we align the embeddings based on all words, to avoid making assumptions about the similarity prior to the analysis.

Identifying words used most differently

In the empirical application, we consider three measures of difference in use. The most common measure of embedding distance is cosine distance (CD), which we also find to work well in our validation task:

$$\text{CD}(w_{1,i}, w_{2,i}) = 1 - \frac{w_{1,i} \cdot w_{2,i}}{\| w_{1,i} \|_2 \| w_{2,i} \|_2} \tag{6.11}$$

In addition to cosine distance, prior research on inductively discovering differences between contemporary speech communities has focused on mistranslations (KhudaBukhsh *et al.*, 2021; Milbauer *et al.*, 2021). A word mistranslate from C_1 to C_2 if the nearest neighbour of $w_{1,i}$ in the aligned embedding of C_2 is not $w_{2,i}$, and similarly from C_2 to C_1; here, we consider a word to mistranslate, if either of the directions does not translate to the word itself.

Sorting mistranslating words by descending frequency partially counters the tendency for cosine distance to identify low-frequent words. Mistranslation is, however, sensitive to highly co-occurring words, such as in the case of first and last names (in our application, *warren* e.g. mistranslates to *elizabeth*). One remedy could be to instead consider a word as mistranslating if it is not among its own k-nearest-neighbours for some value of $k > 1$. Yet, in this case, few or no words will likely mistranslate at all; word embeddings attempt to compress the full variety of discourse into a dense space, and against this background, words are overall likely to be used much more similarly than differently.

Further, by inspecting the dialectograms for words, that exhibit high cosine distance or are among the highest-frequent mistranslations, we observe several cases of polysemy; examples include echo *chamber* versus *chamber* of commerce, the *bell* company versus an alarm, the *viral* load versus a *viral* video, and *turkey* as a country versus an animal. We also observe cases where two persons share part of a name, such as Joy and Harry *reid*. In these cases, both subreddits are likely to use such polysemous words in both senses, but to varying degrees. This observation is what motivates our sense separation measure, which attempts to identify words, where the characteristics use of each corpus is also relatively unique to it.

In particular, for any focal word i, we identify the words in each corpus, that co-occur more with the focal word than the product of their marginal frequencies would suggest, i.e. more than we would expect if co-occurrence was independent. If $C^k_{i,j}$ is the co-occurrence count between focal word i and context word j in a corpus k, N^k_c is the sum of all co-occurrence counts in that corpus and N_w is the number of words, the criteria is:[7]

$$EC^k_{i,j} = \frac{C^k_{i,j} \cdot N^k_c}{\sum_{h=1}^{N_w} C^k_{i,h} \cdot \sum_{h=1}^{N_w} C^k_{h,j}} > 1 \tag{6.12}$$

Given the sets of words co-occurring highly with the focal word in each corpus, we identify the subset of these words, that only co-occur highly in one of the corpora, but not both, $HC^1_i = \{j \in \{1, .., N_w\} \mid EC^1_{i,j} > 1 \wedge EC^2_{i,j} \leq 1\}$ for the first corpus and likewise for the second. Intuitively, these words mark the associations with the focal word that are distinct for each corpus.

For each set of words, HC^1_i and HC^2_i, we calculate the mean of their scalar projections ($\alpha^1_{i,j}$ and $\alpha^2_{i,j}$) on the offset for the focal word - in the case of HC^1_i,

$$\overline{HC^1_i} = \frac{1}{|HC^1_i|} \sum_{j \in HC^1_i} \frac{\alpha^1_{i,j} + \alpha^2_{i,j}}{2} \in [-1, 1] \tag{6.13}$$

and likewise for HC^2_i. Taking the average of the scalar projections is equivalent to projecting the words onto the dashed line running along the main diagonal in the dialectogram; taking the mean of these averages hence captures the overall position of the uniquely high co-occurring words along the main diagonal. Finally, to get our measure of sense separation we subtract the two means,

[7] This is equivalent to cases where the pointwise mutual information (PMI) between the words is positive.

$$S_i = \overline{HC_i^1} - \overline{HC_i^2} \tag{6.14}$$

Alignment and frequency

As Figure 4.3a shows, CD is negatively correlated with frequency. Controlled experiments suggest that this is a feature of the embedding models; they find a large share of the correlation between semantic change and frequency to reappear in settings, where none would be expected (Dubossarsky *et al.*, 2017). The authors argue that this is a general feature of count-based models, as the expected cosine distance between two samples drawn from the same multinomial distribution (representing a given co-occurrence distribution) decreases with the size of the samples.

Experimenting with the frequency weight of the GloVe objective function, we also conclude that the correlation is an artefact of the models. In particular, by either removing the weighting function, (equivalent to weighting all pairs equally) or by inversing the weighting so that lower-frequent pairs receive a higher weight than high-frequent pairs, we observe that the correlation disappears. This, however, come at the cost of drastically increasing the share of mistranslations (from 11% to 67%), why we stick to the default weighting scheme.

In addition to adjusting the Glove model, we also attempted to remove frequency information from the subsequently obtained embedding spaces, before they are aligned. Here, methods have been developed to identify a subspace, that correlates with a semantic feature of interest (Rothe *et al.*, 2016; Rothe and Schütze, 2016; Dufter and Schütze, 2019). After identifying such a subspace, the embedding can be projected to the complement of the subspace, to 'remove' variation corresponding to the corresponding semantic feature.

To obtain a subspace that should capture variation in frequency, we calculate the inner product between the log frequency of words in corpus k, $f_k \in \mathbb{R}^N$ and the dimensions of the embedding space,

$$w_k = E_k^T f_k \in \mathbb{R}^D \tag{6.15}$$

Intuitively, this (row) vector should point in the direction that overall varies most with frequency. To remove this direction from the embedding space, we project the embedding to the orthogonal complement of this direction,

$$E_k^* = E_k[I - w_k^T w_k] \tag{6.16}$$

where E_k^* is the adjusted embedding.

6.3.3 Model and Measure Selection

Over the last 10 years, the NLP community has produced a long range of neural network architectures for learning word embeddings. At a high level, an important distinction is between static and contextual embedding models; whereas static embeddings provide a single vector for each word (type), contextual embeddings provide a numerical representation for each time a word (token) appears. Overall, contextual embeddings provide richer representations, allowing for comparison of how a word is differently distributed as opposed to only differently positioned. However, the contextual embedding models are typically more complex and trained on corpora larger than many of social scientific interest.

For this reason, it is advised to use a contextual model pre-trained on a larger corpus, instead of training one the corpus of interest from scratch (Rodriguez and Spirling, 2022). Although a pre-trained model can be calibrated to a given corpus by continuing the language modelling, it still creates a trade-off; either using the simpler representations and measures of alignment from the corpus-specific static embeddings or using the richer representations and measures from the only partly-attuned contextual embedding (Manjavacas and Fonteyn, 2022).

In addition, several ways of measuring differences between the resulting embeddings also exist. This raises the question of a) what embedding models and measure works best and b) whether such an optimal approach is reliable enough to serve as a computational base for the interpretation of the corpora. This is a particularly pressing challenge for computational inductive approaches, as they cannot easily rely on the train-validate-test partition behind many supervised approaches.

Computational models and measures are typically evaluated by assessing how well they perform in a controlled setting, which establishes how 'correct' or 'true' performance looks. In the case of word embedding models, such evaluation can at a high level be divided into extrinsic and intrinsic approaches. Extrinsic (or downstream) evaluation consists in using the embedding model to perform a different, but related task (typically a classification task such as part-of-speech tagging or named entity recognition). This gives an indirect assessment of the quality of the word representations. Alternatively, it is also possible to devise an intrinsic task, in which the properties of the embedding space are more directly probed. Given our corpora-focused approach, in which our interpretations rely crucially on the properties of the embedding model and measures of difference, we opt for an intrinsic evaluation.

Most embedding models are based on optimising a loss function, which serves as an intrinsic measure of the quality of the embedding. This is however not always the case (such as the SVD embeddings) and loss levels are often not comparable across models, why embedding models are often evaluated intrinsically by other means. One intrinsic option is to compare a measure of how confident the embeddings are in predicting the co-occurrences of the training data, such as perplexity. Another intrinsic option is to compare the associations of the embedding model to equivalent associations made by humans, such as in word intrusion tasks or from surveys (Chang *et al.*, 2009; Murphy, Brian *et al.*, 2012; Kozlowski *et al.*, 2019). A third intrinsic option, akin to an extrinsic evaluation, is to evaluate embedding models by how well they perform on word analogy tasks, such as the Google Analogy Test Set or Bigger Analogy Test Set (BATS) Mikolov, Chen, *et al.*, 2013; Gladkova *et al.*, 2016.

The last two approaches conceptually mix up a corpora-focused evaluation - how well does the embedding model represent the associations of the corpora at hand - with a model-focused evaluation - how well does the model represent general or objective associations, not tied to a particular corpus. To the extent the two evaluation objectives correlate, such as both improving when working with larger corpora, the model-focused evaluation provides an efficient evaluation method, but it does not constitute an ideal evaluation from a corpora-focused perspective.

In line with the basic intuition of creating a validation test, that resembles the purported use of the models and measure, while establishing a basis for adjudicating how well the models and measures perform, we create a synthetical corpus, that resembles an actual corpus in most regards, except that it has certain words swapped. By controlling which words are swapped, as well as the degree to which they are swapped, we can evaluate to what extent the embeddings and measures correctly identify the degree to which a given word is swapped, as well as which word it was swapped with.

Specifically, we create a partially swapped copy of a smaller version of the r/Republican corpus, in which 600 words are swapped (∼11% of vocabulary).[8] To ensure words are swapped in a somewhat syntactically sound manner while allowing us to evaluate how our models and measures perform at different word frequency levels, we swap words based on the following procedure:

- Words are partitioned into 10 frequency deciles.
- For each decile, we randomly draw 30 pairs of words without replacement, such that the two words making up each pair belong to the same part-of-speech.
- To each word pair we assign one of ten degrees to which they will be swapped (10%, 20%, ..., 90%, 100%), such that for each frequency decile and swap degree, we obtain three word pairs, where the words in each pair belong to the same part-of-speech.
- Based on this, we create a swapped version of the small r/Republican corpora, in which each word in each pair is swapped with its paired word according to the allotted degree.[9]

Given the original and swapped r/Republican corpora, we apply three word embedding models: the skip-gram with negative sampling (SGNS) version of Word2Vec; Global Vectors for Word Representation (GloVe); and a distilled version of a pre-trained RoBERTa model (Mikolov, Chen, *et al.*, 2013; Pennington

[8] The smaller dataset was collected at the end of 2020. We later updated the dataset to have a wider coverage. The smaller r/Republican corpus is roughly half the size of the final.

[9] While both words in a pair belong to the same frequency decile, their frequency still differs. Instead of opting for a fixed number of swaps for each pair, in which case we could only attain the allotted swap degree on average, we swap each word in the pair to the allotted degree, with the implication that frequencies of the swapped words change.

et al., 2014; Liu *et al.*, 2019; Sanh *et al.*, 2020). In addition to separate DistilBERT embeddings, we also train a joint DistilBERT model, by continuing the masked language modelling on both corpora at once, inserting a corpus-specific token at the beginning of each comment.

We evaluate how well each of a range of applicable measures described below performs in sorting words according to their degree of swap, by measuring the Spearman's (rank) correlation between the distance measures and the swap degrees.[10,11]

Before we proceed, we note that this validation task is not perfect. Most importantly, real-world corpora will likely be much more different than the corpora considered here, where most word associations are the same. Further, while we have not done any hyperparameter tuning based on the validation task, this is partly because it is likely not suitable for certain hyperparameters. In the case of window size, for instance, smaller windows will likely make it easier to identify differences in this simplified setting, without necessarily generalising to our use case. All in all, we still believe it provides a useful indication of the relative performance of the various approaches.

6.3.4 Training embeddings

For training the SGNS embeddings, we use the implementation made available by Gensim: https://radimrehurek.com/gensim/models/word2vec.html. For training GloVe embeddings, we use the implementation made available by the authors of the model: https://nlp.stanford.edu/projects/glove/. In both cases, we train 300-dimensional embeddings for 30 epochs with a window size of 10. All other parameters are set to the default of the implementations.

For DistilRoBERTa, we rely on the pretrained distilroberta-base checkpoint made available on Huggingface: https://huggingface.co/distilroberta-base, as well as the various modules for training models (AutoModelForMaskedLM, DataCollatorForLanguageModeling, Trainer, TrainingArguments). To extract embeddings for words in comments, that are longer than the token limit of 512, we split up the tokenized comment into separate parts, such that each part except the last has 512 tokens, but with an overlap of 50 tokens in each end of the parts; the 50 last tokens of the first part are identical to the first 50 tokens of the second part, the last 50 tokens of the second part are identical to the first 50 tokens of the third part and so on. For tokens that end up in the overlaps, and hence appear twice, we take the average of the embeddings from each part. For the joint model, we add the corpus-specific tokens to the AutoTokenizer of the pretrained checkpoint.

[10] We use the spearmanr function from the scipy.stats module

[11] In addition to the swap degree, the cosine distance between the swapped words (from the original embedding) is likely to affect the expected distance - the further apart two words are, the more should swapping them induce distance between their aligned representations in the original and swapped corpora. We hence also correlated the measures and models with the product of swap degree and initial cosine distance, which gave similar results.

6.3.5 Measures of difference

Given two (aligned) embedding spaces, we consider several approaches to measure their differences. At a high level, we distinguish between methods applicable to static representations (including the static representation created based on contextual embeddings) and methods applicable to contextual embeddings.

Static measures

For the static representations (including contextual centroids), we consider five measures of difference. In addition to cosine distance and our suggested measure of sense separation described above, these include:

Nearest neighbour overlap:

It is possible to measure similarity without aligning the embedding spaces, by comparing the local structure of the embeddings. Here, we measure the similarities of the words neighbourhood in the two embeddings by comparing their overlap:

$$d_w^{k-NN} = 1 - \frac{|v_{C1,w}^{k-NN} \cap v_{C2,w}^{k-NN}|}{k} \tag{6.17}$$

where $v_{C1,w}^{k-NN}$ is the set of k nearest neighbours for w in C_1, and $|v_{C1,w}^{k-NN} \cap v_{C2,w}^{k-NN}|$ it the size of the intersection of the two neighbourhoods.

PCA of word offsets: In addition to the distance between the embeddings of a word, such as captured by d_{cos}, there could also be information in the direction of the difference between the embeddings. In the naive case of a single, repeated discursive difference, we should ideally be able to identify a single direction corresponding to this difference. While this is an improbable assumption, it motivates the idea of identifying the direction, that represents most of the differences between the two embeddings. One way to do so is to perform Principal Component Analysis (PCA) of the vector offsets, $O = E_1 - E_2 \in \mathbb{R}^{N \times D}$ and use the absolute value of words' principal score on the first principal component as a measure of their difference:

$$\begin{aligned} O &= USV^T \\ d_{w_i}^{PCA} &= \mid US_{[i,1]} \mid \end{aligned} \tag{6.18}$$

where $US_{[i,1]}$ denote the ith row and 1st column of the principal scores, US and $|\cdot|$ is the absolute value.

Distance to SVM boundary: Another way to identify any shared differences between the embeddings is to train a Support Vector Machine (SVM) to separate them. SVM identifies the hyperplane that best separates the two embeddings (maximum margin hyperplane). Intuitively, the easier it is two separate a given word, the more likely is it used differently. As a measure of separability, we rely on the (euclidean) distance between the embedding of a word and the SVM hyperplane, which is equivalent to the length of the rejection vector between the embedding and the hyperplane, $r^{SVM}_{C_j,w}$ - for a given word, we obtain two such measures, corresponding to each of its two embeddings, which we sum:

$$d^{SVM}_w = \| r^{SVM}_{C_1,w} \|_2 + \| r^{SVM}_{C_2,w} \|_2 \tag{6.19}$$

Contextual measures

Extracting static embeddings from contextual embeddings amounts to collapsing the multiple contextual representations of a word across different contexts to a single point in the discursive space. Instead, we might consider comparing the full distribution of a word's contextual representations between the two corpora. Since the shared embedding space lends itself so easily to measures of distance, the most obvious candidate for comparing the difference in distribution is the Wasserstein distance (WD), which in the case of optimal transport is also known as Earth Mover Distance.

Earth Mover Distance suggests the perhaps most intuitive understanding of this measure; by considering the two probability distributions as two piles of earth, the distance between the distributions is then defined as the cost of rearranging the earth from one pile so that it is identical to the other pile, with the cost of moving earth from one point to another point defined as the distance between the points. The optimal transportation of earth (or probability mass) amounts to incurring the lowest cost.

In the case of text analysis, this measure was introduced as Word Mover Distance (WMD), aimed at measuring the similarity of documents (Kusner *et al.*, 2015). By considering documents as a distribution of its words in the embeddings space, the similarity of two documents can be measured as the least cost of moving the points from one document to the other. Subsequently, WMD was also used to measure the extent to which a document engages with one or more focal concepts (Concept Mover Distance), by measuring the cost of moving the embedded words of a document to the embedding of the focal concepts(Stoltz and Taylor, 2019).

In our case, we can use WD in a slightly different way, where we treat a word as two distributions, based on its contextual representations in each embedding, and calculate the WD between these two distributions, based on the cosine distance between the representations. Formally, if $N^{C_1}_w$ and $N^{C_2}_w$ denote the number of contextual representations of w in C_1 and C_2 respectively, the WD distance is defined as:

$$
\begin{aligned}
\text{CD}\left(v_{w_i}, v_{w_j}\right) &= 1 - \frac{\langle v_{w_i}, v_{w_j}\rangle_2}{\| v_{w_i} \|_2 \| v_{w_j} \|_2} \\
T^* &= \arg \min_{T \in \mathbb{R}^{N_w^{C_1}} \times \mathbb{R}^{N_w^{C_2}}} \sum_{i=1}^{N_w^{C_1}} \sum_{j=1}^{N_w^{C_2}} T_{i,j} \cdot \text{CD}\left(v^i_{C_1,w}, v^j_{C_2,w}\right) \\
\text{s.t} \sum_{i=1}^{N_w^{C_1}} T_{i,j} &= \frac{1}{N_w^{C_2}}, \sum_{j=1}^{N_w^{C_2}} T_{i,j} = \frac{1}{N_w^{C_1}} \\
d_w^{WD} &= \sum_{i=1}^{N_w^{C_1}} \sum_{j=1}^{N_w^{C_2}} T^*_{i,j} \cdot \text{CD}\left(v^i_{C_1,w}, v^j_{C_2,w}\right)
\end{aligned}
\tag{6.20}
$$

Experimenting with the WD measure, we noticed that increasing the size of the distributions to be matched leads to lower WD. To avoid a direct correlation between frequency and WD, we implement a sampling procedure. Specifically, for each word, we draw 20 samples without replacement of 75 contextual representations from each corpus. For each sample, we calculate the WD, and take the mean across the 20 WD estimates. We implement WMD using the POT Python package.

Joint contextual measures

Intuitively, for words used similarly, the corpus-tokens should carry no information, whereas, for words used differently, the difference in the embeddings of the corpus-tokens should help distinguish between the different uses of the word. Hence, similar in spirit to the approach taken in d_w^{PCA} and d_w^{SVD}, we then create a measure based on the difference between the corpus-specific tokens in the following way. First, we obtain the contextual representations for each corpus-specific token, $\{v_i^{C_1}\}_{i=1}^{N_{C_1}}$ and $\{v_j^{C_2}\}_{j=1}^{N_{C_2}}$. Then we calculate the mean embedding for each set of corpus-tokens, $\bar{v}^{C_1} = \frac{1}{N_{C_1}} \sum_{i=1}^{N_{C_1}} v_i^{C_1}$ and likewise for C_2. For each word in the vocabulary, we then calculate the two centroids of its contextual representations in the two corpora, $\bar{v}_{C_1,w}$ and $\bar{v}_{C_2,w}$, and project these centroids onto the offset between the mean corpus-token embeddings, $\bar{v}^{C_1} - \bar{v}^{C_2}$, to obtain the normalised scalar projections equal to the cosine similarity (CS) between word centroid and corpus-token-offset

$$
\text{CS}\left(\bar{v}_{C_1,w}, \bar{v}^{C_1} - \bar{v}^{C_2}\right) = \frac{\langle \bar{v}_{C_1,w}, \bar{v}^{C_1} - \bar{v}^{C_2}\rangle_2}{\| \bar{v}_{C_1,w} \|_2 \| \bar{v}^{C_1} - \bar{v}^{C_2} \|_2}
\tag{6.21}
$$

Finally, as our measure of difference, we then calculate the absolute value of the difference in the projections of the centroids (DPC),

$$
d_w^{DPC} = |\text{CS}\left(\bar{v}_{C_1,w}, \bar{v}^{C_1} - \bar{v}^{C_2}\right) - \text{CS}\left(\bar{v}_{C_2,w}, \bar{v}^{C_1} - \bar{v}^{C_2}\right)|
\tag{6.22}
$$

6.3.6 Validation results

Table A.43 reports the correlations based on all words. In general, cosine distances and our suggested co-occurrence separation measure display the highest Spearman correlation with the swap degrees. Cosine distance based on the centroids from contextual embeddings does not correlate better than their static counterparts, but they do correlate much better when using the SVM distance and offset PCA measures, in which cases the static embeddings are useless. The nearest-neighbour measure works for static and dynamic centroids alike but is always correlating less than the corresponding cosine distance measure.[12]

The non-statics measures also correlate subpar compared to the static cosine measure. Cosine WMD correlate less than both cosine distance, offset PCA and 30-NN, while the projection difference from the joint embedding is more promising, despite still correlating less than the cosine distance.

As the swap degree is zero for ∼89% of words in the vocabulary, the difference in correlation might to a high degree be driven by the sorting of non-swapped words. While correlation based on all words most closely mimics the explorative situation, we also perform the correlations for the swapped words only, to assess how well the measures work for words in which we expect a difference. Table A.44 reports the correlation for swapped words only. Generally, correlations are higher and cosine distances achieve the highest correlation, with static embeddings this time correlating higher that the centroids from contextual embeddings. The sense separation measure does, however, not perform as well in this setting, suggesting that its strength is less in deciding the degree of swap, but rather whether words have been swapped or not.

In addition to assessing how well the models and measures recover the induced differences, we can also ask how well the static embeddings identify which words a given word has been swapped with (if any). For words that are not swapped, as well as for words that are swapped less than 50%, we expect the word to translate consistently to itself (i.e. be its own nearest neighbour). For words that are swapped more than 50%, we expect words to translate to their paired word, while for words that are swapped exactly 50%, we expect the translation to be indecisive, that is in some cases translating to the word itself, while in others translating to its paired word.

Table A.45 shows the result of this translation task. Here, there is a clear difference between the static embeddings and contextual centroids - while all embeddings correctly translate words that have not been swapped, as well as (most) words swapped less than 50%, only the static embeddings correctly translate (most) words swapped more than 50%. This is the case because the contextual centroids almost always translate consistently, regardless of the post-training approach.

[12] We did experiment with varying the number of nearest neighbours. There was not a single number that worked best for all embeddings, but even when choosing the optimal number of neighbours specific to each embedding, this measure still does not achieve higher correlation than cosine distance.

Embedding	Cosine distance	SVM distance	Offset PCA	30 NN difference	Sense separation	Cosine WD	Projection difference
W2V, PA	52.0	8.0	0.0	43.0	54.0	–	–
GloVe, PA	54.0	3.0	0.0	50.0	53.0	–	–
W2V, CCA	52.0	2.0	0.0	39.0	52.0	–	–
GloVe, CCA	50.0	3.0	0.0	37.0	46.0	–	–
RoBERTa, pretrained, 4 last layer	53.4	15.0	46.1	36.2	23.8	29.9	–
RoBERTa, 5 epoch, 4 last layer	52.1	29.6	45.6	40.6	23.0	30.2	–
RoBERTa, 20 epoch, 4 last layer	52.7	1.0	41.2	43.2	33.7	33.6	–
RoBERTa, pretrained, last layer	53.3	17.1	45.8	37.1	24.1	32.1	–
RoBERTa, 20 epoch, last layer	49.5	9.1	7.4	41.1	19.5	32.4	–
Joint RoBERTa, 5 epoch, 4 last layer	–	–	–	–	–	–	43.8
Joint RoBERTa, 20 epoch, 4 last layer	–	–	–	–	–	–	44.7
Joint RoBERTa, 20 epoch, last layer	–	–	–	–	–	–	44.8

Table A.43: **All words:** Spearman correlation between the degree of swap and various measures (columns) for different embedding approaches (rows) and all words. Not all measures are applicable to all embeddings (marked with a dash).

Embedding	Cosine distance	SVM distance	Offset PCA	30 NN difference	Sense separation	Cosine WMD	Projection difference
W2V, PA	92.0	6.0	2.0	67.0	56.0	–	–
GloVe, PA	91.0	1.0	0.0	83.0	47.0	–	–
W2V, CCA	92.0	2.0	4.0	66.0	49.0	–	–
GloVe, CCA	91.0	2.0	0.0	66.0	58.0	–	–
RoBERTa, pretrained, 4 last layer	77.7	53.0	46.8	46.4	24.2	51.1	–
RoBERTa, 5 epoch, 4 last layer	82.1	53.3	52.3	65.1	32.5	52.7	–
RoBERTa, 20 epoch, 4 last layer	82.5	40.0	33.2	72.4	38.8	60.2	–
RoBERTa, pretrained, last layer	75.3	52.6	44.2	51.2	21.2	51.8	–
RoBERTa, 20 epoch, last layer	82.0	6.8	32.8	66.4	34.7	58.4	–
Joint RoBERTa, 5 epoch, 4 last layer	–	–	–	–	–	–	46.7
Joint RoBERTa, 20 epoch, 4 last layer	–	–	–	–	–	–	35.2
Joint RoBERTa, 20 epoch, last layer	–	–	–	–	–	–	36.5

Table A.44: **Swapped words only:** Spearman correlation between the degree of swap and various measures (columns) for different embedding approaches (rows) and swapped words only. Not all measures are applicable to all embeddings (marked with a dash). Embeddings are still aligned based on all words, and the nearest neighbours are identified among all words.

	Unswapped words	Less than 50% swap	More than 50% swap
W2V, PA	100.0	94.6	94.3
GloVe, PA	100.0	92.1	93.7
W2V, CCA	100.0	95.0	93.0
GloVe, CCA	99.6	82.5	80.0
RoBERTa, pretrained, 4 last layer	100.0	100.0	0.0
RoBERTa, 5 epoch, 4 last layer	100.0	100.0	0.0
RoBERTa, 20 epoch, 4 last layer	100.0	99.2	0.7
RoBERTa, pretrained, last layer	100.0	100.0	0.0
RoBERTa, 20 epoch, last layer	100.0	100.0	0.0

Table A.45: **Translation accuracy:** Different embedding approaches (rows) by the degree of swap (columns).

6.3.7 Top 100 words used most differently

The following tables show the top 100 words identified by each of the three measures of difference used here. The corresponding dialectograms for these words below can be found at the project GitHub repository.

Highest aligned cosine distance

0-20	20-40	40-60	60-80	80-100
voluntary	🤡	homelessness	goldwater	massacre
juvenile	🇺🇸	gaslighte	ps	aide
com	url	org	additionally	divorce
🤪	lastly	helicopter	ffs	graph
gingrich	forbes_url	disclose	disrupt	cloud
sheet	😡	viral	facilitate	lewis
👏	me	object	gym	pee
settlement	broadband	pipeline	mccarthy	djt
keystone	matt	conservatives	bolster	davis
newt	larry	likewise	alert	blah
spur	light_skin_tone	spill	greenhouse	compliance
❤️	initiative	competitor	charles	jon
karl	elephant	atlanta	monitor	revoke
airline	dissolve	carlson	businessinsider_url	dominion
captain	tldr	initiate	furthermore	tobacco
mission	drilling	upgrade	evidently	jfk
port	andlt	steel	taylor	hurricane
rescue	joy	hunter	offshore	aids
rove	mar	reddit_url	wire	bell
cough	workplace	abusive	supposedly	ma

Table A.46: **100 words with highest aligned cosine distance**

Highest-frequent mistranslations

0-20	20-40	40-60	60-80	80-100
republican	🤣	michigan	permit	warming
democrats	mitch	johnson	nyc	o
republicans	ted	wisconsin	lefty	um
liberal	rand	ha	stupidity	eh
cruz	ah	nancy	suppression	vaccination
leftist	yea	mitt	pete	ga
unfortunately	lmao	korea	audit	mission
subreddit	hundred	youtube_url	kkk	depression
sander	ron	btw	charity	carson
warren	brother	beto	graham	ben
😂	everybody	centrist	woke	bc
kasich	kamala	pa	imgur_url	cuomo
yep	al	anyways	thomas	supposedly
bs	reddit_url	chamber	govt	rnc
harris	ohio	toward	pennsylvania	washingtonpost_url
desantis	communism	rush	garland	hrc
orange	got	found	tucker	san
rubio	maga	hunter	pipeline	sexist
approve	haha	coward	bigot	rioter
son	yup	alien	la	alabama

Table A.47: **100 highest-frequent mistranslating words**

Highest sense separation

0-20	20-40	40-60	60-80	80-100
r	cruz	gun	evangelical	goldwater
biden	obese	dem	we	ron
comment	fox	rand	desantis	stack
democrats	broadband	wedding	democratic	racist
republican	newt	climate	aoc	fascist
liberal	her	oxygen	absentee	suppression
gop	gingrich	paul	dems	devos
left	pete	abrams	install	president
trump	spur	white	fraud	incumbent
she	joe	civility	mayor	homosexual
he	supporter	receipt	party	overcome
conservative	conventional	right	student	neutrality
hillary	greenhouse	orange	sander	mature
clinton	theocracy	illegal	kamala	rubio
republicans	castro	salon	warren	min
democrat	medium	weaponize	competitive	federal
race	obama	newsmax	vote	wi
sanders	his	libertarian	establishment	end
pbs	bernie	hunter	hiv	majority
progressive	leftist	senate	transgender	ballot

Table A.48: **100 words with highest sense separation**

Bibliography

Artetxe, Mikel, Gorka Labaka, and Eneko Agirre (2016). „Learning principled bilingual mappings of word embeddings while preserving monolingual invariance“. en. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, pp. 2289–2294.

Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M Blei (2009). „Reading Tea Leaves: How Humans Interpret Topic Models“. en. In: *Proceedings of the 22nd International Conference on Neural Information Processing Systems*. NIPS'09, pp. 288–296.

Dubossarsky, Haim, Daphna Weinshall, and Eitan Grossman (Sept. 2017). „Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models.“ In: *Conference on Empirical Methods in Natural Language Processing*. MAG ID: 2759556264 S2ID: ab370db3fe0ee0cb272f13ebbb32894de90bec8d, pp. 1136–1145.

Dufter, Philipp and Hinrich Schütze (Nov. 2019). „Analytical Methods for Interpretable Ultradense Word Embeddings“. en. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. arXiv: 1904.08654. Association for Computational Linguistics, Hong Kong, China.

Gladkova, Anna, Aleksandr Drozd, and Satoshi Matsuoka (2016). „Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't.“ en. In: *Proceedings of the NAACL Student Research Workshop*. San Diego, California: Association for Computational Linguistics, pp. 8–15.

Jason Micheal Baumgartner (2022). *Pushshift*.

KhudaBukhsh, Ashiqur R, Rupak Sarkar, Mark S Kamlet, and Tom Mitchell (2021). „We Don't Speak the Same Language: Interpreting Polarization through Machine Translation“. en. In: *Association for the Advancement of Artificial Intelligence* AAAI-21, p. 9.

Kozlowski, Austin C., Matt Taddy, and James A. Evans (Oct. 2019). „The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings“. en. In: *American Sociological Review* 84.5, pp. 905–949.

Kusner, Matt J, Yu Sun, Nicholas I Kolkin, and Kilian Q Weinberger (July 2015). „From Word Embeddings To Document Distances“. en. In: *Proceedings of the 32nd International Conference on Machine Learning*. ICML'15 37, p. 10.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (July 2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. en. arXiv:1907.11692 [cs].

Manjavacas, Enrique and Lauren Fonteyn (June 2022). „Adapting vs. Pre-training Language Models for Historical Languages“. en. In: *Journal of Data Mining & Digital Humanities* NLP4DH.Digital humanities in... P. 9152.

Matthew Podolak (2022). *PMAW*.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (Sept. 2013). „Efficient Estimation of Word Representations in Vector Space“. en. In: *arXiv:1301.3781 [cs]*. ZSCC: 0020188 arXiv: 1301.3781.

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever (Sept. 2013). „Exploiting Similarities among Languages for Machine Translation“. en. In: *arXiv:1309.4168 [cs]*. arXiv: 1309.4168.

Milbauer, Jeremiah, Adarsh Mathew, and James Evans (2021). „Aligning Multidimensional Worldviews and Discovering Ideological Differences“. en. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. ZSCC: 0000000, pp. 4832–4845.

Murphy, Brian, Talukdar, Partha Pratim, and Mitchell, Tom (2012). „Learning Effective and Interpretable Semantic Models using Non-Negative Sparse Embedding“. In: *Proceedings of COLING 2012*. COLING 2012, pp. 1933–1950.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (2014). „Glove: Global Vectors for Word Representation“. en. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ZSCC: 0022883. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543.

r/Democrats (2022). *Pushshift*.

r/Republican (2022). *Pushshift*.

Rodriguez, Pedro L. and Arthur Spirling (Jan. 2022). „Word Embeddings: What Works, What Doesnt, and How to Tell the Difference for Applied Research“. en. In: *The Journal of Politics* 84.1, pp. 101–115.

Rothe, Sascha, Sebastian Ebert, and Hinrich Schütze (2016). „Ultradense Word Embeddings by Orthogonal Transformation“. en. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, pp. 767–777.

Rothe, Sascha and Hinrich Schütze (2016). „Word Embedding Calculus in Meaningful Ultradense Subspaces“. en. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Berlin, Germany: Association for Computational Linguistics, pp. 512–517.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf (Feb. 2020). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. en. arXiv:1910.01108 [cs].

Schönemann, Peter H. (Mar. 1966). „A generalized solution of the orthogonal procrustes problem“. en. In: *Psychometrika* 31.1, pp. 1–10.

Stoltz, Dustin S. and Marshall A. Taylor (July 2019). „Concept Movers Distance: measuring concept engagement via word embeddings in texts“. en. In: *Journal of Computational Social Science* 2.2, pp. 293–313.

Törnberg, Petter (Oct. 2022). „How digital media drive affective polarization through partisan sorting“. en. In: *Proceedings of the National Academy of Sciences* 119.42, e2207159119.

William, Press (May 2011). *Canonical Correlation Clarified by Singular Value Decomposition*.

Bibliography

7

Adler-Nissen, Rebecca, Katrine A. Eggeling, and Pia Wangen (2021). „Machine Anthropology: A View from International Relations". In: *Big Data & Society* 8.2.

Albris, Kristian, E. I. Otto, S. L. Astrupgaard, E. M. Gregersen, L. S. Jørgensen, O. Jørgensen, C. R. Sandbye, and S. Schønning (2021). „A View from Anthropology: Should Anthropologists Fear the Data Machines?" In: *Big Data & Society* 8.2, p. 20539517211043656.

Albris, Kristoffer (2011). *Feltnoter 3: Noter & Spørgsmål*. Private Fieldnotes. Image of Fieldnote was reproduced with the permission of Kristoffer Albris.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow (2020). „The welfare effects of social media". In: *American Economic Review* 110, pp. 3629–3676.

Allcott, Hunt and Matthew Gentzkow (2017). „Social Media and Fake News in the 2016 Election". In: *Journal of Economic Perspectives* 31.2, pp. 211–236.

Allen, Jennifer *et al.* (2020). „Evaluating the fake news problem at the scale of the information ecosystem". In: *Science Advances* 6, eaay3539.

Arseniev-Koehler, Alina and Jacob G. Foster (2020). *Machine learning as a model for cultural learning: Teaching an algorithm what it means to be fat*. arXiv: 2003.12133 [cs.CY].

Azarbonyad, Hosein, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps (2017). *Words are Malleable: Computing Semantic Shifts in Political and Media Discourse*. arXiv: 1711.05603 [cs.CL].

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, Megan B. F. Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky (2018). „Exposure to opposing views on social media can increase political polarization". In: *Proceedings of the National Academy of Sciences of the United States of America* 115, pp. 9216–9221.

Barassi, Veronica (2015). *Activism on the Web: Everyday Struggles against Digital Capitalism*. 1st. Vol. 4. Routledge.

Barbera, Pablo (2015). „Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". eng. In: *Political analysis* 23.1, pp. 76–91.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker (2019). „Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data". In: *American Political Science Review* 113.4, pp. 883–901.

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau (2015). „Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" In: *Psychological Science* 26, pp. 1531–1542.

Bennett, W. Lance and Alexandra Segerberg (2012). „The Logic of Connective Action: The Personalization of Contentious Politics“. In: *Information, Communication & Society* 15.5, pp. 739–776.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. Springer New York.

Bjerre-Nielsen, Andreas and Kristoffer Laigaard Glavind (2022). „Ethnographic data in the age of big data: How to compare and combine“. In: *Big Data & Society* 9.1.

Blei, David M. (2012). „Probabilistic Topic Models“. In: *Communications of the ACM* 55.4, pp. 77–84.

Blok, Anders and Morten Axel Pedersen (2014). „Complementary social science? Quali-quantitative experiments in a big data world“. In: *Big Data & Society* 1.2.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (Oct. 2008). „Fast unfolding of communities in large networks“. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Bode, Leticia (2017). „Closing the gap: gender parity in political engagement on social media“. In: *Information, Communication & Society* 20.4, pp. 587–603.

Boellstorff, Tom, Bill. Maurer, and Tom. Boellstorff (2015). *Data, Now Bigger and Better!* eng. Paradigm ; 46. Chicago: Prickly Paradigm.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai (2016). *Quantifying and Reducing Stereotypes in Word Embeddings*. arXiv: 1606.06121 [cs.CL].

Bond, Robert and Solomon Messing (2015). „Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook“. In: *American Political Science Review* 109.1, pp. 62–78.

Bornakke, Troels and Brian Lystgaard Due (2018). „BigThick Blending: A method for mixing analytical insights from big and thick data sources“. In: *Big Data & Society* 5.1.

Bourdieu, Pierre (1973). „Cultural Reproduction and Social Reproduction“. In: *Knowledge, Education, and Cultural Change*. Ed. by Roger Brown. London: Tavistock Publications, pp. 71–84.

Bourdieu, Pierre (1986a). *Distinction: A Social Critique of the Judgement of Taste*. Routledge & Kegan Paul.

Bourdieu, Pierre (1986b). „The Forms of Capital“. In: *Handbook of Theory and Research for the Sociology of Education*. Ed. by J. G. Richardson. New York: Greenwood Press, pp. 241–258.

Breslin, Samantha Dawn, Anders Blok, Thyge Ryom Enggaard, Tobias Gårdhus, and Morten Axel Pedersen (2022). „"Affective Publics": Performing Trust on Danish Twitter during the COVID-19 Lockdown“. English. In: *Current Anthropology* 63.2, pp. 211–218.

Bullinaria, J. A. and J. P. Levy (2007). „Extracting Semantic Representations from Word Co-occurrence Statistics: A Computational Study“. In: *Behavior Research Methods* 39, pp. 510–526.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova (Dec. 2019). *Social Media and Xenophobia: Evidence from Russia*. Working Paper 26567. National Bureau of Economic Research.

Caliskan, Aylin *et al.* (2017). „Semantics derived automatically from language corpora contain human-like biases“. In: *Science* 356, pp. 183–186.

Carlsen, Hjalmar Bang and Snorre Ralund (Jan. 2022). „Computational grounded theory revisited: From computer-led to computer-assisted text analysis“. en. In: *Big Data & Society* 9.1, p. 205395172210801.

Charles, Victoria and Tudor Gherman (2019). „Big Data Analytics and Ethnography: Together for the Greater Good“. In: *Big Data for the Greater Good*. Ed. by Ali Emrouznejad and Victoria Charles. Vol. 42. Studies in Big Data. Cham: Springer.

Clegg, Nick and Chaya Nayak (Aug. 2020). *New Facebook and Instagram Research Initiative to Look at US 2020 Presidential Election*. https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election/.

Coscia, Michele and Frank Neffke (2017). *Network Backboning with Noisy Data*. arXiv: 1701.07336 [physics.soc-ph].

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs.CL].

DiMaggio, Paul (1982). „Cultural Capital and School Success: The Impact of Status Culture Participation on the Grades of U.S. High School Students“. In: *American Sociological Review* 47.2, pp. 189–201.

Duncan MacRae, Jr. (1958). *Dimensions of Congressional Voting: A Statistical Study of the House of Representatives in the Eighty-First Congress*. Berkeley: University of California Press.

Eady, Gregory, Richard Bonneau, Joshua A. Tucker, and Jonathan Nagler (2023). *News Sharing on Social Media: Mapping the Ideology of News Media, Politicians, and the Mass Public*. Preprint.

Earl, Jennifer (2016). „Protest Online: Theorizing the Consequences of Online Engagement“. In: *The Consequences of Social Movements*. Ed. by Lorenzo Bozi, Marco Giugni, and Katrine Uba. Cambridge: Cambridge University Press, pp. 363–400.

Eltantawy, Nahed and Julie B Wiest (2011). „The Arab spring| Social media in the Egyptian revolution: reconsidering resource mobilization theory“. In: *International journal of communication* 5, p. 18.

Enikolopov, Ruben *et al.* (2020). „Social Media and Protest Participation: Evidence from Russia“. In: *Econometrica* 88.4. Accessed 21 Apr. 2024, pp. 1479–1514.

Ennser-Jedenastik, Laurenz, Christina Gahn, Anita Bodlos, and Martin Haselmayer (2022). „Does social media enhance party responsiveness? How user engagement shapes parties issue attention on Facebook“. In: *Party Politics* 28.3, pp. 468–481. eprint: https://doi.org/10.1177/1354068820985334.

Ethnote (2024). https://ethnote.org/. Accessed: 2024-05-03.

Evans, James A. and Pedro Aceves (2016). „Machine Translation: Mining Text for Social Theory“. In: *Annual Review of Sociology* 42.Volume 42, 2016, pp. 21–50.

Feng, Shiyu and Chin Yee Tan (2024). „Toward conceptual clarity for digital cultural and social capital in student learning: Insights from a systematic literature review“. In: *Humanities and Social Sciences Communications* 11.1, p. 68.

Folkemødet (2022). *About Folkemødet (in English)*. URL: https://folkemoedet.dk/om-folkemodet/about-folkemodet-in-english/.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (2018a). „Word embeddings quantify 100 years of gender and ethnic stereotypes“. In: *Proceedings of the National Academy of Sciences* 115.16, E3635–E3644. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1720347115.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (Apr. 2018b). „Word embeddings quantify 100 years of gender and ethnic stereotypes“. en. In: *Proceedings of the National Academy of Sciences* 115.16, E3635–E3644.

Glaser, Barney and Anselm Strauss (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.

Golovchenko, Yevgeniy (2020). „Measuring the scope of pro-Kremlin disinformation on Twitter“. In: *Humanities and Social Sciences Communications* 7.176, pp. 1–11.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, *et al.* (2011). „The Dynamics of Protest Recruitment through an Online Network“. In: *Scientific Reports* 1, p. 197.

Grimmer, Justin and Brandon M. Stewart (2013a). „Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts“. en. In: *Political Analysis* 21.3. ZSCC: 0002076, pp. 267–297.

Grimmer, Justin and Brandon M. Stewart (2013b). „Text as data: The promise and pitfalls of automatic content analysis methods for political texts“. In: *Political Analysis* 21.3, pp. 267–297.

Grinberg, Nir *et al.* (2019). „Fake news on Twitter during the 2016 U.S. presidential election“. In: *Science* 363, pp. 374–378.

Guess, Andrew, Brendan Nyhan, and Jason Reifler (2018). *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 U.S. Presidential Campaign*. Working Paper. European Research Council, Brussels.

Guess, Andrew M. *et al.* (2023a). „How do social media feed algorithms affect attitudes and behavior in an election campaign?“ In: *Science* 381, pp. 398–404.

Guess, Andrew M. *et al.* (2023b). „Reshares on social media amplify political news but do not detectably affect beliefs or opinions“. In: *Science* 381, pp. 404–408.

Halupka, Max (2014). „Clicktivism: A Systematic Heuristic“. In: *Policy & Internet* 6, pp. 115–132.

Halupka, Max (2018). „The legitimisation of clicktivism“. In: *Australian Journal of Political Science* 53.1, pp. 130–141.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky (2018). *Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change*. arXiv: 1605.09096 [cs.CL].

Hargittai, Eszter (2020). „Potential biases in big data: Omitted voices on social media“. In: *Social science computer review* 38.1, pp. 10–24.

Hoffman, Mark Anthony, Jean-Philippe Cointet, Philipp Brandt, Newton Key, and Peter Bearman (2017). „The (Protestant) Bible, the (Printed) Sermon, and the Word(s): The Semantic Structure of the Conformist and Dissenting Bible, 1660–1780“. In: *Poetics* 68, pp. 89–103.

Hotelling, H. (1933). „Analysis of a complex of statistical variables into principal components“. In: *Journal of Educational Psychology* 24, pp. 417–441, 498–520.

Hsu, Wendy F (2014). „Digital ethnography toward augmented empiricism: A new methodological framework“. In: *Journal of Digital Humanities* 3.1, pp. 3–1.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes (2012). „Affect, Not Ideology: A Social Identity Perspective on Polarization“. In: *Public Opinion Quarterly* 76.3, pp. 405–431.

Jones, Jason, Mohammad Amin, Jessica Kim, and Steven Skiena (2020). „Stereotypical Gender Associations in Language Have Decreased Over Time“. en. In: *Sociological Science* 7. ZSCC: 0000008, pp. 1–35.

katerinaptrv (July 2023). *GPT4 - All Details Leaked*. Medium. Available online: https://medium.com/@daniellefranca96/gpt4-all-details-leaked-48fa20f9a4a.

Khondker, Habibul H. (2011). „Role of the New Media in the Arab Spring“. In: *Globalizations* 8.5, pp. 675–679.

KhudaBukhsh, Ashiqur R., Rupak Sarkar, Mark S. Kamlet, and Tom M. Mitchell (2020). *We Don't Speak the Same Language: Interpreting Polarization through Machine Translation*. arXiv: 2010.02339 [cs.CL].

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019). „The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings“. In: *American Sociological Review* 84.5, pp. 905–949.

Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena (2014). *Statistically Significant Detection of Linguistic Change*. arXiv: 1411.3315 [cs.CL].

Larson, Jennifer M. *et al.* (2019). „Social Networks and Protest Participation: Evidence from 130 Million Twitter Users“. In: *American Journal of Political Science* 63.3. Accessed 21 Apr. 2024, pp. 690–705.

Le, Quoc V. and Tomas Mikolov (2014). *Distributed Representations of Sentences and Documents*. arXiv: 1405.4053 [cs.CL].

Lee, M. and J. Martin (2015). „Coding, Counting and Cultural Cartography“. In: *American Journal of Cultural Sociology* 3, pp. 1–33.

Lévi-Strauss, Claude (1963). *Structural Anthropology*. New York: Basic Books.

Levy, Omer and Yoav Goldberg (June 2014). „Linguistic Regularities in Sparse and Explicit Word Representations“. In: *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*. Ed. by Roser Morante and Scott Wen-tau Yih. Ann Arbor, Michigan: Association for Computational Linguistics, pp. 171–180.

Levy, Ro'ee (2021). „Social Media, News Consumption, and Polarization: Evidence from a Field Experiment“. In: *American Economic Review* 111.3, pp. 831–870.

Liu, Yinhan, 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: 1907.11692 [cs.CL].

Lonkila, Markku and Pertti Jokivuori (2023). „Sharing and liking as youth nano-level participation. Finnish students civic and political engagement in social media“. In: *Journal of Youth Studies* 26.6, pp. 803–820. eprint: https://doi.org/10.1080/13676261.2022.2049731.

Ma, Xiaofei, Zhiguo Wang, Patrick Ng, Ramesh Nallapati, and Bing Xiang (2019). *Universal Text Representation from BERT: An Empirical Study*. arXiv: 1910.07973 [cs.CL].

Maaten, Laurens van der and Geoffrey Hinton (2008). „Visualizing Data Using t-SNE“. In: *Journal of Machine Learning Research* 9.Nov, pp. 2579–2605.

Margetts, Helen, Peter John, Scott Hale, and Taha Yasseri (2015). *Political Turbulence: How Social Media Shape Collective Action*. Princeton: Princeton University Press.

McInnes, Leland, John Healy, and James Melville (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: 1802.03426 [stat.ML].

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781 [cs.CL].

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever (2013). *Exploiting Similarities among Languages for Machine Translation*. arXiv: 1309.4168 [cs.CL].

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Distributed Representations of Words and Phrases and their Compositionality*. arXiv: 1310.4546 [cs.CL].

Milbauer, Jeremiah, Adarsh Mathew, and James Evans (Nov. 2021). „Aligning Multidimensional Worldviews and Discovering Ideological Differences“. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Ed. by Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 4832–4845.

Miller, Melissa L. and Cristian Vaccari (2020). „Digital Threats to Democracy: Comparative Lessons and Possible Remedies“. In: *The International Journal of Press/Politics* 25.3, pp. 333–356.

Mohr, John W., Robin Wagner-Pacifici, and Ronald L. Breiger (2015). „Toward a computational hermeneutics“. In: *Big Data & Society* 2.2.

Morozov, Evgeny (2009). „The Brave New World of Slacktivism“. In: *Foreign Policy*.

Morselli, Davide, Maite Beramendi, Zakaria Bendali, and Olivier Fillieule (2023). „A Longitudinal Approach to Online Collective Identity Work: The Case of the Gilets Jaunes in the Var Department“. In: *Mobilization: An International Quarterly* 28.3, pp. 301–322.

Müller, Karsten and Carlo Schwarz (Aug. 2021). „Fanning the Flames of Hate: Social Media and Hate Crime“. In: *Journal of the European Economic Association* 19.4, pp. 2131–2167.

Müller, Karsten and Carlo Schwarz (2023). „From Hashtag to Hate Crime: Twitter and Antiminority Sentiment“. In: *American Economic Journal: Applied Economics* 15.3, pp. 270–312.

Munk, Anders K. and Brit Ross Winthereik (2022). „Computational Ethnography: A Case of COVID-19s Methodological Consequences“. In: *The Palgrave Handbook of the Anthropology of Technology*. Singapore: Palgrave Macmillan.

Munk, Anders Kristian (2019). „Four Styles of Quali-Quantitative Analysis: Making sense of the new Nordic food movement on the web“. In: *Nordicom Review* 40.s1, pp. 159–176.

Nardi, Bonnie (2010). *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. University of Michigan Press.

Nelson, Laura K. (Feb. 2020). „Computational Grounded Theory: A Methodological Framework“. en. In: *Sociological Methods & Research* 49.1, pp. 3–42.

Nissenbaum, Helen (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books.

Niwa, Yoshiki and Yoshihiko Nitta (1995). *Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries*. arXiv: cmp-lg/9503025 [cmp-lg].

Nyhan, Brendan, Jaime Settle, Emily Thorson, *et al.* (2023). „Like-minded sources on Facebook are prevalent but not polarizing“. In: *Nature* 620, pp. 137–144.

OpenAI, Josh Achiam, Steven Adler, *et al.* (2024). *GPT-4 Technical Report*. arXiv: 2303.08774 [cs.CL].

Paff, Sarah (2022). „Anthropology by Data Science“. In: *Annals of Anthropological Practice* 46, pp. 7–18.

Pariser, Eli (2011). *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*. Penguin.

Pedersen, Morten Axel (2021). „Going Native in Data Science: An (Auto) Ethnography of Interdisciplinary Collaboration“. In: *The Moral Work of Anthropology: Ethnographic Studies of Anthropologists at Work*. Ed. by Hanne Overgaard Mogensen and Birgitte Gorm Hansen. 1st. Vol. 2. Berghahn Books, pp. 133–168.

Pedersen, Morten Axel (2023). „Editorial introduction: Towards a machinic anthropology“. In: *Big Data & Society* 10.1.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (Oct. 2014). „GloVe: Global Vectors for Word Representation“. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Ed. by Alessandro Moschitti, Bo Pang, and Walter Daelemans. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543.

Peponakis, Manolis, Sarantos Kapidakis, Martin Doerr, and Evangelia Tountasaki (2024). „From Calculations to Reasoning: History, Trends and the Potential of Computational Ethnography and Computational Social Anthropology“. In: *Social Science Computer Review* 42.1, pp. 84–102.

Poole, K. T. and H. Rosenthal (1985). „A Spatial Model for Legislative Roll Call Analysis“. In: *American Journal of Political Science* 29.2, pp. 357–384.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever (2018). *Improving language understanding with unsupervised learning*. Technical Report. OpenAI.

Rheault, Ludovic and Christopher Cochrane (2020). „Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora“. In: *Political Analysis* 28.1, pp. 112–133.

Ricoeur, Paul (1981). *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Cambridge, UK: Cambridge University Press.

Salganik, Matthew J. (2018). *Bit by bit: social research in the digital age*. OCLC: on1012406622. Princeton: Princeton University Press.

Sanjek, R., ed. (1990). *Fieldnotes: The Makings of Anthropology*. http://www.jstor.org/stable/10.7591/j.ctvv4124m. Cornell University Press.

Santucci, Jean-François, Albert Doja, and Luca Capocchi (2020). „A Discrete-Event Simulation of Claude Lévi-Strauss Structural Analysis of Myths Based on Symmetry and Double Twist Transformations“. In: *Symmetry* 12, p. 1706.

Schöll, Nikolas, Aina Gallego, and Gaël Le Mens (2023). „How Politicians Learn from Citizens Feedback: The Case of Gender on Twitter“. In: *American Journal of Political Science* n/a.n/a, pp. 557–574. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12772.

Seaver, Nick (2017). „Algorithms as culture: Some tactics for the ethnography of algorithmic systems“. In: *Big Data & Society* 4.2.

Shahin, S., J. Nakahara, and M. Sánchez (2024). „Black Lives Matter goes global: Connective action meets cultural hybridity in Brazil, India, and Japan“. In: *New Media & Society* 26.1, pp. 216–235.

Shmargad, Yotam and Samara Klar (2020). „Sorting the news: How ranking by popularity polarizes our politics“. In: *Political Communication* 37, pp. 423–446.

Sieben, Sabine and Clemens M. Lechner (2019). „Measuring cultural capital through the number of books in the household“. In: *Measurement Instruments for the Social Sciences* 1.1, p. 1.

Siegel, Alexandra A. and Vicky Badaan (2020). „#No2Sectarianism: Experimental Approaches to Reducing Sectarian Hate Speech Online“. In: *American Political Science Review* 114.3, pp. 837–855.

Slapin, Jonathan B. and Sven-Oliver Proksch (2008). „A Scaling Model for Estimating Time-Series Party Positions from Texts“. In: *American Journal of Political Science* 52.3, pp. 705–722.

Stopczynski, Arkadiusz, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, M. Mønster Madsen, and Jakob Eg Larsen (2014). „Measuring Large-Scale Social Networks with High Resolution“. In: *PLoS ONE* 9.4, e95978.

Suk, J., A. Abhishek, Y. Zhang, S. Y. Ahn, T. Correa, C. Garlough, and D. V. Shah (2021). „#MeToo, Networked Acknowledgment, and Connective Action: How Empowerment Through Empathy Launched a Social Movement“. In: *Social Science Computer Review* 39.2, pp. 276–294.

Temporão, M., C. Vande Kerckhove, C. van der Linden, Y. Dufresne, and J. M. Hendrickx (2018). „Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach“. In: *Political Analysis* 26.4, pp. 457–473.

Theocharis, Yannis (2015). „The Conceptualization of Digitally Networked Participation“. In: *Social Media + Society* 1.2.

Tucker, Joshua *et al.* (Oct. 2017). „From Liberation to Turmoil: Social Media and Democracy“. In: *Journal of Democracy* 28.4, pp. 46–59.

Turney, P. D. and P. Pantel (Feb. 2010). „From Frequency to Meaning: Vector Space Models of Semantics“. In: *Journal of Artificial Intelligence Research* 37, pp. 141–188.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2023). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].

Vryonides, Marios (2007). „Social and cultural capital in educational research: issues of operationalisation and measurement“. In: *British Educational Research Journal* 33, pp. 867–885.

Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman (2019). „GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding“. In: In the Proceedings of ICLR.

Wojcieszak, Magdalena, Andreu Casas, Xiao Yu, Jonathan Nagler, and Joshua A. Tucker (Sept. 2022). „Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity“. In: *Science Advances* 8.39, eabn9418.

Wu, Patrick Y., Jonathan Nagler, Joshua A. Tucker, and Solomon Messing (2023). *Large Language Models Can Be Used to Estimate the Latent Positions of Politicians*. arXiv: 2303.12057 [cs.CY].

Young, Lori and Stuart Soroka (2012). „Affective News: The Automated Coding of Sentiment in Political Texts“. In: *Political Communication* 29.2, pp. 205–231. eprint: https://doi.org/10.1080/10584609.2012.671234.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). „Political Effects of the Internet and Social Media“. In: *Annual Review of Economics* 12.Volume 12, 2020, pp. 415–438.