# HOW DOES INTERVENTION BY AUTHORITY FIGURES AFFECT POLARIZATION IN THE TURKISH TWITTERSPHERE?

ÖZGÜR TOGAY

BOĞAZİÇİ UNIVERSITY

2023

# HOW DOES INTERVENTION BY AUTHORITY FIGURES AFFECT POLARIZATION IN THE TURKISH TWITTERSPHERE?

Thesis submitted to the

Institute for Graduate Social Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Sociology

by

Özgür Togay

Boğaziçi University

2023

# DECLARATION OF ORIGINALITY

I, Özgür Togay, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature……

Date……

ABSTRACT

How Does Intervention by Authority Figures Affect Polarization in the Turkish Twittersphere?

The thesis study examines the impact of statements made by authority figures on political polarization observed in the Turkish Twittersphere, focusing on the divisive conversations over Turkey's exit from the Istanbul Convention and sexual orientation of the star volleyball player Ebrar Karakurt. It draws from the most recent advances in the field of computational social science to conduct a novel study of political polarization through big data. In addition to proposing a methodological framework on how to study polarization on social media using innovative approaches, the findings of this study provide insights into the degree of polarization within the Turkish Twittersphere and illustrate how Twitter users react to messaging by authority figures.

# ÖZET

## Otorite Figürlerinin Açıklamaları Türkiye Twitter'ındaki Kutuplaşmayı Nasıl Etkiliyor?

Bu tez çalışması, otorite figürleri tarafından yapılan açıklamaların Türkiye Twitter'ında gözlemlenen siyasi kutuplaşmayı nasıl etkilediğini incelemektedir. Türkiyeli Twitter kullanıcılarının karşıt kamplara ayrıldığı tartışmalı iki vakaya odaklanan çalışma, yenilikçi bilişimsel sosyal bilimler yaklaşımları üzerinden Twitter'daki kutuplaşmanın boyutlarını araştırıyor. Çalışma, siyasi kutuplaşmayı büyük veri üzerinden inceleyebilmek için bilişimsel sosyal bilim alanındaki en son gelişmelerden yararlanıyor. Yenilikçi yaklaşımlar kullanarak sosyal medyadaki kutuplaşmanın nasıl inceleneceğine dair metodolojik bir çerçeve önermenin yanı sıra, çalışmanın bulguları Türkiye Twitter küresindeki kutuplaşmanın derecesine dair fikir veriyor ve Twitter kullanıcılarının otorite figürlerinin mesajlarına nasıl tepki verdiğini gösteriyor.

# ACKNOWLEDGEMENTS

I am deeply grateful to my advisors, Assist. Prof. Başak Taraktaş and Assist. Prof. Sinan Erensü, for their unwavering support, guidance, and kindness throughout the thesis process.

I am forever indebted to my family, my mother, sister, and father, who have been my unyielding pillars of strength and motivation. My closest companion, Yumak, whose mischievous antics never fail to bring joy to my life, deserves a special mention.

My sincere appreciation extends to the esteemed members of my thesis committee, Assist. Prof. Mert Arslanalp, Assist. Prof. Uzay Çetin, and Assoc. Prof. Akın Ünver, for their invaluable time and constructive feedback that have significantly enhanced the quality of my research.

Furthermore, I would like to extend my sincere thanks to the Scientific and Technological Research Council of Turkey (TÜBİTAK 2210-A) for their financial support throughout my graduate studies.

Finally, I express my heartfelt gratitude to all those who have supported me during this academic journey, my teammates in the BLM project, my friends and comrades-in-arms Deniz and Emre, and logistical supporter Anıl, I am lucky to have had the privilege of knowing them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

The existence of a healthy public sphere, as a place where individuals come together to discuss the matters of the public and reach a consensus, has long been deemed essential for democracy (Habermas, Lennox, & Lennox, 1974). As the leading institution that informs public opinion, the traditional media was instrumental in achieving this. However, traditional media outlets have been losing influence due to the emergence of new media platforms and the de-legitimization of existing ones (Rajendran & Thesinghraja, 2014). This loss of confidence undermines the media's status as a trustworthy institution and leads its previous audience to increasingly fragmented spaces where personalized truths can be found, thus creating echo chambers (Garimella, Gionis, Morales, & Mathioudakis, 2018). Furthermore, social media platforms provide a medium where anyone can be both the disseminator and the consumer of information or truths, which has brought about unseen levels of fragmentation (Conover, et al., 2011).

Survey data shows that 40 to 60 percent of adults in developed countries get their news on social media (Newman, Fletcher, Kalogeropoulos, & Nielsen, 2017). While earlier on, social media was seen as a global force of democratization, now hopes have turned into fears that social media may contribute to the trend of democratic backsliding (Tucker, Theocharis, Roberts, & Barberá, 2017). It is revealing that the Journal of Democracy featured an article on social media titled "Liberation Technology" in 2010, while less than a decade later, another titled "Can

Democracy Survive the Internet?" appeared on its pages (Persily, 2017). While social media may give a platform to marginalized voices and increase plurality in public debates, campaigns conducted by states and extremist organizations reveal that they also provide grounds for manipulation. As the research on social media grew out of its infancy at the end of the 2010s, the perspective shared by many scholars is that social media itself is not intrinsically democratic or undemocratic; instead, it provides a field where political actors –both democratic and anti-democratic- can compete for power and influence.

In recent years, there has been growing concern that social media platforms may be undermining democracy, with various stakeholders, including the media, academia, civil society, and politicians themselves, expressing concern over this issue. This concern arises from the fact that while social media has been lauded for its potential to promote political engagement, it is also possible that social media experiences may make individuals less interested in politics in an era of hyper-partisanship.

One possible explanation for this phenomenon is the rise of political polarization on social media platforms. Polarization refers to the increasing ideological divide between individuals or groups, resulting in a lack of consensus on political issues and a corresponding increase in hostility and resentment towards those with opposing views. This can lead to a toxic political environment in which individuals are less likely to engage with political issues and may even become disengaged from the political process altogether. In addition, the proliferation of fake news and echo chambers on social media platforms may exacerbate this trend, as individuals are exposed to a narrow range of perspectives and may become entrenched in their own beliefs.

The emergence of computational social science provides novel tools to understand political polarization on social media. Computational social science is an interdisciplinary field that uses computational methods and tools to study social behavior and its underlying processes. This approach is particularly useful for studying complex phenomena, such as political polarization, as it allows researchers to analyze large amounts of data systematically and efficiently. Thus, it can provide valuable insights into the mechanisms that drive political polarization on social media platforms, as well as the factors that contribute to its formation and perpetuation.

Furthermore, computational social science has the potential to shed light on how political polarization on social media may affect offline political behavior and attitudes. For example, studies using computational methods can help identify whether the echo chamber effects commonly observed on social media platforms are also present in the offline world and can provide insight into the extent to which social media may be contributing to the further polarization of political attitudes.

Overall, the use of computational methods in studying political polarization on social media is an important and necessary approach for gaining a better understanding of this complex and important phenomenon. Moreover, by providing a more comprehensive and nuanced view of the factors that contribute to political polarization on social media, computational social science can help develop effective strategies for mitigating this issue and promoting more civil and productive political discourse online.

Finally, Turkey has been widely recognized as a society marked by high levels of political polarization and substantial social media engagement. As a nation grappling with the formidable challenge of democratic backsliding, Turkey offers an

illuminating case study for researchers of democracy. Consequently, Turkish social media presents a compelling avenue for exploring the dynamics of political polarization through such computational methods.

## 1.2 Plan of the thesis

The following thesis is composed of six chapters. Following the introductory chapter, Chapter 2 presents a review of the literature on key topics relevant to the thesis. In Chapter 3, the cases to be analyzed are introduced, along with a brief overview of Twitter as a social media platform. Chapter 4 details the process of cleaning and utilizing data, as well as an overview of the methodological tools and measures used in the study. Chapter 5 presents the analysis of cases and discusses the study's findings in relation to the literature. Finally, Chapter 6 provides a summary of the chapters, identifies the study’s limitations, suggests potential avenues for future research, and considers the implications of the findings.

# CHAPTER 2

# LITERATURE REVIEW

This chapter will provide a comprehensive overview of the existing literature on political polarization, including its study on social media. The literature review will also highlight the key developments and research that contribute to the theoretical framework of political polarization. Additionally, this chapter will examine the literature on political polarization in Turkey, including its manifestation on Turkish social media. The review will identify gaps in the literature and outline how this thesis aims to contribute to the existing knowledge on the subject.

## 2.1 Theoretical framework

Political polarization is a phenomenon in which individuals or groups within a society hold increasingly extreme political beliefs and attitudes and become more entrenched in their positions (Mason, 2018), where various differences within a society become increasingly aligned along a single dimension and individuals begin to view and understand politics and society in terms of "us" versus "them." (McCoy, Rahman, & Somer, 2018) . This can lead to a situation in which political discourse becomes increasingly divisive, and individuals or groups on opposite sides of an issue become increasingly isolated from one another.

Researchers have employed various concepts to study the different forms of political polarization. The earliest approach studied what is currently referred to as ideological polarization, which occurs when individuals or groups become increasingly extreme in their political beliefs and values (McCoy, Rahman, & Somer,

2018). This can manifest as a growing divide between political parties and within political parties, as individuals or groups on opposite sides of an issue become increasingly entrenched in their positions and unwilling to compromise or engage in productive dialogue. Ideological polarization is sometimes studied under or referred to as issue or policy-based polarization.

The past decade has seen the development of another concept for studying political polarization, as its effects were extending far beyond ideological disagreement. Affective political polarization refers to the increasing animosity and negative emotions between individuals on opposite sides of the political spectrum. This can manifest as increased hostility and contempt towards those on the other side and a tendency to view the other side negatively. From a social identity perspective, affective polarization is a consequence of partisan group identity, as individuals who identify with rival parties, e.g., as Republicans or Democrats, are likely to view opposing partisans negatively and co-partisans positively (Iyengar, Lelkes, Levendusky, & Westwood, 2019).

In contrast, ideological polarization focuses on policy issues and the extent to which individuals hold extreme positions. Affective polarization does not necessarily require individuals to hold extreme views but instead refers to the growing divide between individuals with different political beliefs. This can manifest as a situation where individuals on opposite sides of the political spectrum are increasingly entrenched in their positions and less willing to engage in productive dialogue or compromise. While affective and ideological polarization are related, they are distinct phenomena. Affective polarization is more concerned with the emotional and interpersonal aspects of political division, while ideological polarization focuses on

the policy differences between individuals on opposite sides of the political spectrum.

Researchers have identified several factors that contribute to this phenomenon, such as increased social sorting, i.e., the tendency for individuals to identify with a particular political party and to associate with others who share their beliefs. This can lead to a situation in which individuals are socially distant from those on the opposite side of the political spectrum and can make it easier for individuals to make generalized and often inaccurate assumptions about the other side. Tracking the course of affective polarization in the USA through American National Election Studies, Iyengar and Krupetkin (2018) found that while partisan animus began to increase in the 1980s, it has grown significantly over the past two decades. As partisan affect has intensified, it has also become more structured, with ingroup favoritism increasingly associated with outgroup animus. Hostility towards the opposing party has become a more powerful motive for political participation than positive affect towards one's own party. These findings suggest that partisan affect has significant consequences for political attitudes and behavior.

Mason (2015, 2018b) argues that the decline of cross-cutting identities, in which individuals are affiliated with multiple social groups that have different political beliefs, is at the root of affective polarization. Increased sorting, which can be influenced by factors such as race and religion, has made it easier for individuals to reinforce their social identities and make assumptions about those on the other side of the political spectrum.

While reinforcing social identities appears to be a key factor in explaining affective polarization, other research has found that ideological polarization also plays a role. For example, Rogowski and Sutherland (2016), Bougher (2017),

Webster and Abramowitz (2017) have all found that ideological polarization can impact affective polarization, as individuals with extreme beliefs may be more likely to have negative feelings towards those on the other side. Furthermore, Mason (2018) argues that social connections to the groups that hold particular ideological labels is more important than the actual policy preferences, in other words, partisanship is borne out of the belonging felt towards a "liberal" social group rather than adherence to liberal policies.

The proliferation of partisan news outlets, which often reinforce extreme ideological positions, has also been blamed for the current polarized environment. These outlets may activate partisan identities and lead individuals to adopt more extreme positions, increasing affective polarization. However, there are counterarguments to this hypothesis. Some argue that it is unclear whether exposure to partisan news causes affective polarization, as it may be the case that individuals who are already highly polarized are more likely to consume this type of news. Additionally, it is not certain that individuals only consume news that aligns with their ideology, and some research has found evidence to the contrary.

Overall, while there is evidence that both increased sorting and ideological polarization can contribute to affective polarization, the precise mechanisms by which this occurs are still not fully understood. Further research is needed to understand these processes and identify effective strategies for mitigating the harmful effects of affective polarization.

It is important to note that affective and ideological polarization often overlap. For example, Mason's research shows that a person's ideology or party affiliation can become part of their identity, influencing their views on ingroup and

outgroup members. This demonstrates the interconnectedness of affective and ideological polarization.

Political polarization is also studied through which segment of the society it occurs, namely, elite and mass polarization. Elite polarization refers to the polarization of political elites, i.e. politicians and opinion leaders, while mass polarization of the general public.

Barber and Pope (2019) sought to understand whether party identification or ideology is more influential in shaping political views. The study used an experiment in which participants were presented with policy positions that were either liberal or conservative and were attributed to Trump. The researchers measured any changes in the participants' own policy positions in response to these cues. The results of the study showed that low-knowledge respondents, strong Republicans, those who approve of Trump, and self-described ideological conservatives were the most likely to respond to the cues. This suggests that a significant portion of the Republican electorate may be party loyalists rather than being primarily motivated by ideology. These findings indicate that for some people, party identification may be more influential in shaping their political views than their personal beliefs and values.

Another study comparing the United States, the United Kingdom, Spain, and Belgium found that partisanship is a stronger dividing factor than social cleavages such as race, religion, language, or ethnicity (Westwood, et al., 2018). The researchers used surveys to measure polarization and a trust game experiment to determine whether partisanship had a greater influence than other social divides.

The researchers found that partisanship exerts a stronger psychological bond than affiliation with social groups, even when those cleavages are the basis for the parties' ideological positions and electoral appeals. This may be because there are no

social norms that constrain the disapproval of political opponents, as there are for other social divides. Additionally, people are more likely to be held responsible for their party affiliation, which is a choice, than group affiliations based on immutable characteristics. The study also noted that negative campaigning, as is common in the United States, may contribute to the strength of partisan bias.

Some researchers posited that priming a common identity could reduce affective polarization. This is based on the common ingroup identity model which proposes that individuals possess multiple identities that vary in their strength and relevance depending on the context and the most salient of these identities shapes an individual's perception of the political world. Following that, Levendusky (2018) conducted both natural and laboratory experiments to examine the effect of priming American national identity on affective polarization. The results of these experiments showed that priming national identity reduced affective polarization by prompting a psychological recategorization of the opposing party. However, a common ground may not necessarily reduce affective polarization. Jahani et al. (2022) tested the hypothesis that the threat of a common enemy can reduce conflict between members of rival groups through an online experiment. The results showed that Republicans were significantly less likely to learn from Democrats when primed about a common enemy, particularly those with strong conservative views. Jahani et al.'s findings suggest that common enemies may not reduce inter-group conflict in highly polarized societies and support the existence of asymmetric political polarization in the United States.

Polarization is often viewed as a danger to democracy. Arbatli and Rosenberg (2021) argue that polarization creates "enemy camps," and such hostile portrayal makes voters more willing to accept anti-democratic measures against their political

rivals. This tacit approval becomes more evident during elections, where political controversy is prevalent and the stakes are higher. They find that political polarization positively correlates with electoral manipulation and intimidation against the opposition. Political polarization can have several negative consequences for society, including increased social and political instability, decreased trust and cooperation among individuals and groups, and a reduced ability to address and solve critical public policy issues.

Effects of political polarization can also spill over to nonpolitical settings, as shown by McConnell et al. (2017) who combined a field experiment with a survey to examine the influence of partisanship on economic behavior. The results of their study showed that partisanship has a systematic influence on individuals' economic behavior in both the workplace and the market. For example, workers requested lower wages when their employer was a co-partisan, and consumers preferred copartisan dealers. These findings suggest that partisanship can affect behavior in nonpolitical settings and has implications for the economic outcomes of individuals and firms. There is diverse literature focusing on the broader implications of polarization. As partisanship becomes a salient part of identity, vilification of opposition is exacerbated. The lack of civility and mutual respect turns politics into a zero-sum game where elites shy away from compromises on policies. Partisanship is also linked with limited accountability of own-party politicians (Parker & Dull, 2011). Political elites can also benefit from their constituents' increased vulnerability to disinformation to manipulate the public toward their self-serving policy preferences (Flynn, Nyhan, & Reifer, 2017). In their study of the use of trolls on Twitter in Turkey, Bulut and Yörük (2017) found that the ruling party actively

deploys a large army of trolls to exacerbate polarization. This silences positional public figures, creating “a digital culture of lynching and censorship.”

It is, therefore, crucial for researchers and policymakers to understand the different forms of political polarization and to develop strategies for mitigating its adverse effects.

## 2.2 Social media and political polarization

Social media interactions take place in a very different environment. Ho and McLeod's (2008) experiment on face-to-face and computer-mediated interactions found that people are more willing to speak openly and less likely to filter their expressions in the latter. User anonymity allowed by specific platforms, including Twitter, has frequently correlated with incivility (Santana, 2014). Social media is also distinguished by what can be called the social consumption of news. In their experiment on selective exposure, Messing and Westwood (2012) found that social endorsements play a higher role than partisan affiliation in selecting news, revealing how personal networks influence the selection and consumption of news. Barberá et al.'s (2015) study of retweet networks in the U.S. found that while non-political topics (sports, entertainment, etc.) look more like "national conversations," those of political importance regularly take the form of "echo chambers" with high polarization.

Social media provides a platform allowing both elites and masses to disseminate narratives and contribute to polarization. Studies have shown that elite polarization can lead to mass political polarization (Hetherington, 2002), while elites can also be very potent distributors of polarizing content and disinformation (Ünver, 2019). It is commonly thought that false information is amplified in echo bubbles,

where outside voices are filtered through algorithms. Increased partisanship is seen as a byproduct of this process. In an exemplary study of the dissemination of rumors on Facebook, del Vicario et al. (2016) have found that “information related to distinct narratives” create homogenous and polarized communities.

That said, there also have been studies challenging these assumptions. Regardless of echo bubbles, people's exposure to opposing views on social media is found to be high (Bakshy, Messing, & Adamic, 2015) (Duggan & Smith, 2016). A report by Reuters shows that forty percent of social media users across various countries encounter diverse sources (Newman et al., 2017). Access to the internet or social media does not correlate with increased polarization, and the effect of misinformation on people's political knowledge seems to be limited (Alcott & Gentzkow, 2017). On the contrary, the increase in polarization in the USA was most profound among people least likely to use the internet and social media. Boxell et al. (2017) suggest that the predominance of weak ties (acquaintances, distant relatives, etc.) in a user's personal network causes exposure to diverse views on social media, which may explain the lack of increase in polarization. Moreover, there are studies showing that intergroup interactions online may actually increase polarization. An online experiment conducted by Bail et al. (2018) with a large group of Democrat and Republicans revealed that the Republican participants displayed significantly more conservative views after following a liberal account, while the Democrats also became more liberal, although the effect was not statistically significant. This is also mirrored by a survey from the Pew Research Center (2016) which found that most social media users in the U.S. find people with opposing political views online disrespectful and angry, and political interactions with them to be stressful and frustrating. This is doubly important in light of Weeks's (2015) experimental study,

which found that "anger encourages partisan, motivated evaluation of uncorrected misinformation that results in beliefs consistent with the supported political party". This was further evidenced an experiment demonstrating the polarizing effects of online partisan criticism (Suhat, Bello-Pardo, & Maurer, 2017).

While most research focused on the United States, a study by Urman (2020) compared levels of political polarization on Twitter in 16 democratic countries, using a "network-analytic audience duplication approach" based on the followers of political parties' Twitter accounts. In short, this analyzes the intersection between the followers of different parties, with a low percentage of shared followers suggesting higher polarization. The study found that each country's Twittersphere differed in its level of polarization, with some being perfectly integrated, integrated, mixed, polarized, or perfectly polarized. The highest levels of polarization were found in two-party systems with plurality electoral rules, while the lowest levels were found in multi-party systems with proportional voting, suggesting a link between political systems and online political polarization. The author argues that extrapolation from single-case studies is problematic and this may explain contradictory findings of the earlier research. Although this study is valuable for its comparative approach, there are questions regarding its methodology. Specifically, it is questionable whether polarization can be accurately inferred from the distribution of party account followers. Alternative methods, such as measuring homophily and sentiment over interactions, may be more effective in assessing polarization. Additionally, it can be argued that parties as institutions may not be particularly polarizing, but rather that polarization occurs around specific figures or partisan outlets, e.g. Trump is often more polarizing than the Republican Party (TRUMPOLGOP)  Moreover, all parties were given equal weight in multi-party systems, which does not match the political

reality in many countries, as parties have widely differing political power and following.

It should be noted that studies on Twitter are significantly overrepresented within studies of polarization on social media, or as a matter of fact, in any studies on social media, and cross-platform studies are a rarity. A significant exception is a three-platform study conducted by Yarchi et al. (2021) that aimed to improve the conceptual precision of research on polarization on social media. In addition to including underrepresented platforms of Facebook and Whatsapp in their study, the authors have also analyzed both the textual content and networks with these platforms, which are seldom taken together. Focusing on a particularly polarizing case in Israel, the researchers collected approximately 250,000 posts from social media split over four time periods of importance and used computational social science methods for analysis. This allowed the study to address three key aspects of political polarization on social media: interactional polarization, positional polarization, and affective polarization. The results indicated that political polarization cannot be viewed as a unified phenomenon, as there are significant differences between platforms. Specifically, interactions on Twitter tended to be homophilic and exacerbated positional polarization, with pronounced inter-group hostility, while de-polarization occurred over time on WhatsApp. In contrast, Facebook was found to be the least homophilic platform in terms of interactions, positions, and emotions expressed. These findings underline the importance of making conceptual distinctions in the study of political polarization and highlight the need to further investigate the factors that contribute to polarization on social media. This research played a key role in the development of this thesis, particularly owing

to its comprehensive examination of polarization and its clear operationalization of important concepts.

To recap, social media is differentiated by the social consumption of news, which is already filtered and engaged by affiliated network members. Incivility and hostility rampant in social media is found to promote affective polarization, which further widens the partisan divides. On the other hands, the elites can use their influence to flood the media with disinformation as a tactic to avoid responsibility and stoke tensions.

## 2.3 Political polarization in Turkey

The literature on political polarization in Turkey is mostly centered around political parties and their electorate. Erdoğan and Uyan-Semerci (2018) trace the discussions on polarization in Turkey back to Mardin's interpretation of center-periphery relations in the country. They argue that the tension between the center and periphery was replaced with a secular-conservative divide, which carries similar undertones of urban versus rural and Western versus Eastern. According to Kiriş (2011), polarization is a significant factor in explaining the political system in Turkey. The high level of polarization among rival political parties in Turkey leads to a consolidation of the electorate, with individuals becoming more loyal to their own parties and less willing to associate with the opposition. This creates pressure on the political moderates, who face the risk of being pushed out of the political sphere. Furthermore, Bilgiç et al. (2014) argue that the polarization in Turkey carries the potential to threaten social peace and public security and warn of the potential relationship between polarization and inter-group injustice and conflicts based on identity and ethnicity. Sunar and Kaya (2015) disputed the notion of increasing

political and social polarization between conservatives and secularists in Turkey, using data from the World Value Survey and the European Value Surveys from 1990 to 2011 to demonstrate that there were no significant differences between the two groups. However, this analysis has limitations, suffering from surveys' inability to capture affective polarization, and a time frame that does not coincide with the period of autocratization in Turkey, which is posited as a key factor in polarization by Laebans and Öztürk (2021). In their study on partisan identities, Laebens and Öztürk found that the autocratization of Turkey exacerbated polarization and emboldened partisan identities. A vital part of this identity is the perceived threat to their well-being if other parties take power. However, in contrast to democratic countries, extensive patronage networks have weakened the effect of partisanship over political action, making the clientelist relationship a better indicator of political mobilization.

Mete-Dökücü and Just (2021) examined party polarization between 1950 and 2018 through party policies and manifestos. The authors found that party system polarization is not a persistent feature of Turkish politics and that it primarily reflects party differences on social issues rather than economic or European integration issues. Additionally, the results demonstrate that military interventions in 1960 and 1980 reduced party system polarization in subsequent elections, even when controlling for other determinants of polarization. Following a different framework, Ertan et al. (2022) introduced a network-based approach for measuring perceived party polarization in multiparty systems, challenging the traditional two-party or left-right dichotomy often used to measure party polarization. The authors employ a nationally representative survey in Turkey to create cognitive political networks for each respondent by asking about their perceived relationships among major political

parties. These networks are then used to calculate various measures that may represent perceived party polarization, which are compared to the commonly used Left-Right political ideology distance scale. The findings indicate limited overlap between a L-R scale and the multiparty measure developed by the authors. This suggests that the cognitive political network approach may provide a flexible and direct method for measuring party polarization in survey studies within multiparty systems.

Seeking to understand if and how the Turkish political elite perceived polarization, Aydın-Düzgit and Balta (2018) studied their views using four political frames, namely, harmony, continuity/decline, conspiracy and conflict. The authors observe that there is a significant degree of polarization among the Turkish elites in regards to their views on the presence of polarization within Turkey, with this divide overlapping with that between the government and the opposition. An analysis of the justificatory arguments employed by these elites in constructing their frames of reference reveals that while those who deny the existence of polarization attribute its absence to fundamental characteristics of society, reductive comparisons with history, or internal/external enemies, those who acknowledge its presence attribute it to political and institutional factors and processes. The authors' bleak overlook suggests that given the denial of polarization by the pro-government elite and the substantial gap between the two camps' justificatory narratives, the currently reported high rates of polarization in Turkey can be expected to remain as is in the near future. With a wholly different perspective, Karaosmanoğlu's (2020) study investigates how the ruling elite in Turkey uses food as a means of polarization. The author argues that the government's attempts to promote a national culinary culture and the specific choices made in the palace's menu, are both in line with and against to their efforts to

polarize and appeal to the general public. While the government has attempted to polarize the public through pitting non-alcoholic native beverages against rakı, the palace kitchen has also pursued neo-Ottoman glamour through exotic items such as dragon fruit smoothies and chia seeds.

The most comprehensive field research on political polarization in Turkey was conducted by Erdoğan and Uyan-Semerci (2020) through face-to-face interviews and surveys. The research was conducted in three rounds under the Dimensions of Polarization in Turkey project in 2015, 2017, and most recently in 2020 with a sample size of 4006 interviewees. The researchers identified three key components in affective polarization: social distance, moral superiority, and political intolerance. They found that these components were present among supporters of all political parties.

Recent years have seen an upsurge in social media studies in Turkey. For example, Ozduzen et al. (2021) conducted a manually labeled sentiment analysis of 106K tweets collected from refugee-related hashtags, revealing the representations of anti-immigrant rhetoric on social media. The authors argue that the lack of regulation and anonymity provided by social media platforms bolsters the reach of racist commentary. In a similar yet smaller study, Erdogan-Ozturk and Guler (2020) analyzed a sample of tweets collected under the '#ülkemdesuriyeliistemiyorum' hashtag. Their findings also revealed that social media displayed a language of exclusion and hate used towards Syrian refugees not found in mainstream newspapers.

Tunç and Furman (2019) utilized social network analysis to investigate polarization on Twitter during the 2017 Constitutional Referendum, through mention and reply networks, as well as a hashtag network that connected users who posted

under the same hashtags. These networks are considered to be polarized if the nodes can be partitioned into two highly cohesive subgroups, which may represent opposing viewpoints. Modularity scores were also employed as a metric for polarization. Their findings indicated that homophilic interactions were dominant, although some cross-partisan interaction was observed in replies to the status updates of news-affiliated accounts.

Bozdağ (2020) conducted a qualitative study of 29 social media users through interviews, revealing how users experience polarization and use strategies to manage their diverse networks. The results showed that polarization leads to a decrease in network diversity, even among moderate or politically disengaged users. Twitter was perceived as a more political space, with interactions being seen as more partisan and less diverse, compared to Facebook which includes more family networks. Bozdağ and Koçer (2022) found that even self-critical users rely on sources they feel closer to when sharing news.

While there has been a welcome increase in the use of computational methods to study Turkish social media, there is a lack of studies that combine both content and network analysis. Previous content analyses have been limited in scope, often relying on manual labeling, and have not adequately addressed user stances or polarization. Network analyses, on the other hand, lack the detailed insights provided by content analysis. Additionally, there is a lack of research on the role of elites in driving or mitigating polarization on social media as most studies are descriptive in scope. The aim of this thesis is to address the current lack of multi-modal approaches combining content and network analysis in the study of social media.

# CHAPTER 3

# THE MEDIUM AND THE CASES

In this chapter, I will present a brief summary of Twitter as a social media platform, highlighting its core functionalities. This will be followed by an overview of the studied cases, including the reasons for their selection, their timelines, key events, and key parties.

## 3.1 A functional overview of Twitter

Twitter is a popular social media platform that allows users to send and receive messages called tweets, which are limited to 280 characters but can include visual media such as pictures, gifs and videos. Users sign up with an email, choose a username, and design their Twitter homepage, including uploading a profile photo and adding a short biographical description.

A tweet is shown anyone who follows you on Twitter, and tweets by public accounts may be shown to non-followers depending on Twitter's algorithm. Public tweets also show up under trends or search results. Instant updates make Twitter a suitable medium for both distribution and consumption of news. Tweets are used for various other purposes, including status updates, opinion sharing, advertising, and political campaigns and propaganda.

A tweet can include hyperlinks, media, mentions, and hashtags. A hyperlink is a clickable URL within a tweet that takes the user to another web site. Mentions are tweets containing another user's Twitter username preceded by the "@" symbol, used to draw their attention. A hashtag is a keyword or phrase preceded by the "#"

symbol used to describe a topic. When tweeted, hashtags automatically become clickable links that take users to a page featuring a feed of all recent tweets containing that hashtag. Twitter users use hashtags to categorize their tweets, making it easy for others to find and follow tweets about specific topics or themes. Popular hashtags are shown as trending and represent the hot topics of the day. Hashtags are often used for political messaging, which can include demands by online social movements, criticisms, praises or attacks.

### 3.2 Ebrar Karakurt and LGBTQ rights

The issue of LGBT rights has become a new axis of culture war in Turkey, with polarizing views on the subject across society. While Turkey has traditionally been more socially liberal than its neighbors in the Middle East, recent years have seen a rise in conservatism and nationalism that has led to a backlash against LGBT rights.

Turkey has a long history of LGBT activism, with the country's first LGBT rights organization, Lambda Istanbul, established in 1993. In 2014, the country made headlines as the annual pride parade was attended by more than 100,000 people. However, in recent years, the government's stance on LGBT rights has become increasingly hostile, reflecting broader cultural and political trends in the country. The 2015 Pride Parade was dispersed by the police, and the following parades were banned due to security concerns, and subsequent attempts to hold parades have been met with police brutality and violence from far-right groups. In addition, LGBT individuals have faced increasing discrimination and persecution, with reports of arrests, detentions, and harassment by the police.

During the 2021 Women's European Volleyball Championship, Turkish volleyball player Ebrar Karakurt faced criticism and homophobic slurs on social

media after sharing a photo of herself and her girlfriend on her Instagram account on 14 August 2021. The post was published by the conservative newspaper Takvim, which labeled her sexual orientation and relationship as "scandalous." Turkish Imam Ahmet Mahmut Ünlü, known as Cübbeli Ahmet Hoca, condemned Karakurt, expressing concern about the "troubles" she could potentially bring to Turkish people and calling for her removal from the team, a stance shared by other influential religious figures including İhsan Şenocak and Mehmet Boynukalın. However, Karakurt quickly received messages of support on social media, from figures including popular singer Demet Akalın, teammates Eda Erdem Dündar and Naz Aydemir, sports journalist İsmail Şenol, voiced their support for Karakurt. An official statement was made by the Turkish National Volleyball Federation spokesperson Kurtaran Mumcu, who praised Karakurt's skill and ambition, and called people to respect her private life.

The outing of Ebrar Karakurt on social media quickly sparked a debate about the status of LGBTQ individuals in Turkey. The conversation included discussions about the morality of being an LGBTQ person, the ability of an LGBTQ person to hold an official position in the country or represent the country, whether homosexuality can be considered a perversion or disease, or whether LGBTQ was a device employed by malicious actors against the Turkish public. Some participants also reflected on the criticisms by pointing out cases of sexual exploitation within religious wakifs and the support of child marriage within some conservative circles.

LGBTQ rights remain a highly contested and polarizing issue in Turkey, with little space for open conversation. Mumcu's statement, which did not address LGBTQ rights but rather focused on the private life, can be seen as an attempt to mediate this polarizing issue. The present case shares some similarities with a study

conducted by Ozduzen and Korkut, which examined polarization in Turkish social media regarding LGBTQ identity (Ozduzen & Korkut, 2020). The study analyzed the polarization that ensued after the outing of singer Intizar during a custody battle, and focused on how an ostensibly mundane and everyday event transformed into a macro-political polarization. The study revealed that the opposing factions swiftly employed their habitual frames in the discussion, thereby highlighting the major points of division in society, such as religion, family values, and patriotism. The act of pitting the rights and identities of the LGBT community against the welfare of the state and the nation is a rhetorical strategy that is frequently adopted by conservative authoritarian politicians, as observed in countries such as Russia and Hungary (Edenborg, 2021).

The selected case for this study is particularly well-suited for an examination of polarization due to a number of factors. First, choosing a more recent case allows for the inclusion of tweets that may have been removed or accounts that may have been banned or closed over time. Additionally, the size of the data set is an important consideration, as a small case may not be suitable for computational methodologies, while a very large case may be excessively time-consuming and computationally expensive for a single researcher to analyze. The case is also noteworthy in that, unlike other studies of polarization in Turkey, it does not center around political parties, which allows for a broader focus beyond party lines. The case of the Turkish national volleyball team is significant because international competitions are often times when the politically polarized Turkish public comes together to support their national team, potentially priming the shared national identity.

Finally, the presence of an intervention from an official political authority figure provides a clear point of reference to study any changes in polarization in

response to the intervention. It is important to note that for the purposes of this study, the term "authority" refers only to an individual's official political capacity, and does not include other figures who may be considered part of the elite, such as religious leaders or public opinion makers. This narrower definition allows for a clearer identification and operationalization of interventions. Given that polarization is widely recognized as a threat to democracy and social cohesion, this conceptualization also has the potential to make a policy contribution. The findings of this research will be useful for interested political or civil parties who wish to understand what strategies are effective and ineffective in mitigating polarization.

### 3.3 Turkey's exit from the Istanbul Convention

The Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence, usually referred as the Istanbul Convention, is an international treaty that aims to prevent and combat violence against women and domestic violence. The convention was prepared through the contribution of numerous women's rights organizations, and was opened for signature in Istanbul, Turkey, in 2011.

The Istanbul Convention is considered to be the most comprehensive international treaty on preventing and combating violence against women and domestic violence. The convention defines violence against women as "a violation of human rights and a form of discrimination against women and shall mean all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life" (Council of Europe, 2011).

The convention requires signatories to take a range of measures to prevent and combat violence against women and domestic violence, including providing support services for victims, prosecuting perpetrators, and promoting gender equality. The convention also recognizes the importance of addressing the root causes of violence against women, including discrimination, inequality and gender stereotypes.

Turkey was one of the first countries to sign the Istanbul Convention, and the Turkish parliament showed a rare moment of inter-party consensus in its ratification. It was also shown as a badge of honor by the ruling party figures, including then prime-minister Erdoğan. The convention entered into force in 2014, reflecting the country's commitment to women's rights and gender equality. However, in recent years, the government's stance on women's rights and gender equality has become increasingly conservative and nationalist, reflecting the demands of influential religious figures and communities.

The opposition to the Istanbul Convention in Turkey in 2018, with the publication of critical articles in religious newspapers. However, the first official indication of the government's stance on the Convention was voiced by President Recep Tayyip Erdoğan during a closed-doors party camp in June 2019. During this speech, Erdoğan expressed his belief that the Convention was not a sacred law and suggested that it could be reevaluated. In Februray 2020 the government announced its intention to review the convention's provisions and possibly withdraw from it. The government's decision was met with widespread opposition from human rights groups and international organizations, who argued that the convention was a vital tool for protecting women's rights and combating violence against women, also by the opposition parties, who emphasized that any potential withdrawal from the

Convention would require parliamentary approval. There have been reports of opposition to withdrawal within the AKP, mainly by the women members. In addition, KADEM, a conservative women's rights organization in which the President's daughter, Sümeyye Erdoğan, is an active member, has defended the convention (KADEM, 2020). Despite the opposition, the government moved forward with its plans to withdraw from the convention. In March 2021, President Erdogan issued a decree announcing Turkey's official withdrawal from the Istanbul Convention.

The reasons for Turkey's withdrawal from the Istanbul Convention are complex and multifaceted, reflecting broader cultural and political trends in the country. One of the main reasons cited by the government for its withdrawal from the convention is the perceived threat to traditional gender roles and family values. The government argued that the convention promoted "immoral" and "un-Islamic" values, and that it was a threat to the country's conservative and nationalist identity.

Another reason cited by the government was the belief that the convention was being used to promote LGBT+ rights, which the government views as a threat to social order and traditional values. The government argued that the convention's provisions on gender identity and sexual orientation were "incompatible with Turkey's social and family values" (BBC News, 2021).

The arguments voiced by the opponents of the convention mirror those of the conservative right abroad (Kováts & Pető, 2017). In recent years, there has been a growing trend among conservative right-wing movements to fight against the "gender agenda" (Korolczuk & Graff, 2018). This agenda is believed to be a concerted effort to undermine traditional family values and social norms by promoting "gender ideology" and LGBTQ+ rights. In contrast, conservative right-

wing movements tend to view gender as an innate and immutable characteristic that is determined by biological sex. Any attempt to blur the lines between male and female or to promote alternative gender identities is thus perceived as an assault on nature, family, religion and the nation.

# CHAPTER 4

# DATA, DESIGN, AND METHODOLOGY

This chapter presents the methodological framework, data collection, and research design of the study. It begins by introducing computational social science and its relevant methods for the study. The specific approaches and metrics used in this study are then described. Finally, the process of collecting and cleaning up the data is described.

## 4.1 Computational social science

Computational social science is an interdisciplinary field that uses computational methods to study social phenomena. It encompasses a wide range of subfields, including network analysis, agent-based modeling, natural language processing, sentiment analysis, and big data analytics. Computational social science is an increasingly important field, as it allows researchers to analyze and understand complex social phenomena at a large scale. This is particularly relevant in the digital age, where vast amounts of data are generated through social media, digital communication, and other online platforms and interactions. By using computational methods to analyze this data, researchers can gain insights into social dynamics, patterns of behavior, and the mechanisms underlying social systems.

Natural language processing (NLP) is a subfield of computational linguistics that focuses on the interaction between computers and human (natural) languages. It involves the development of algorithms and models that can automatically process and analyze large amounts of text data. NLP is an essential tool for computational

social science, as it allows researchers to automatically analyze and interpret large amounts of textual data systematically and efficiently. NLP is particularly relevant for studying social media data, as it allows researchers to automatically identify and analyze the content and sentiment of social media posts, providing a rich source of information about social dynamics and public opinion.

### 4.1.1 Stance detection

One of the key challenges in quantifying polarization is accurately identifying the positions of individuals on a particular issue, referred to as "stance detection." Manual labeling of documents is a common approach for identifying positions in smaller datasets. However, this method becomes infeasible when dealing with large volumes of data encompassing thousands of documents. This led to the rise of automated stance detection methods. Several studies have utilized traditional feature engineering techniques in stance detection. For example, Sen et al. (citation) introduced a novel set of features and employed both a support vector machine (SVM) model and a feedforward neural network model (Ming, Wong, Tan, Sen, & Chiang, 2016) while Küçük et al. employed unigrams, bigrams, hashtags, external links, emoticons, and named entities as features in an SVM model (Küçük & Can, 2018).

Following the rise of neural networks in natural language processing, numerous recent studies utilized neural models for stance detection. Du et al. employed an attention-based model for classification of stances (Du, Xu, He, & Gui, 2017). Zarrella and Marsh proposed a transfer learning method utilizing features learned through distant supervision on two large unlabeled datasets (Zarrella & Marsh, 2016). Chen and Ku applied a neural network model to classify the stance of

social media posts by considering users' tastes and topics and user comments on posts (Chen & Ku, 2016). Dey et al. utilized a two-phase long short-term memory (LSTM) model (Dey, Shrivastava, & Kaushik, 2018).

A significant advancement in the field occurred with the development of transformer-based models, particularly Google's BERT model in 2019. BERT models trained for stance detection have consistently demonstrated superior performance compared to previous models, as demonstrated by Ghosh et al (Ghosh, Singhania, Singh, Rudra, & Ghosh, 2019).

Despite the impressive progress made in recent years, stance detection is still faced with particular challenges. Many proposed models tend to be domain or language-specific and may not generalize well to other datasets. Additionally, a lack of access to the underlying code can hinder efforts to replicate or re-implement these models. Among the various approaches, Hybrid Content Analysis and Unsupervised Stance Detection show particular promise, as they are claimed to be language and domain agnostic, the code is readily available, and the models were tested on non-evaluation datasets.

### 4.1.2 Topic models

Topic modeling is a statistical approach that utilizes techniques such as Latent Dirichlet Allocation (Blei & Lafferty, 2009) to group expressions into distinct sets of words that discriminate between different groups of documents. Each document contains a minimal set of different topics. This allows for the identification of distinct topics within a collection of documents and the assignment of documents to those topics.

The obtained word groupings are then represented for researchers to interpret (Chang, Boyd-Graeber, Gerrish, Wang, & Blei, 2009). The process is entirely data-driven and controlled by a mathematical model. While the selection of the algorithm enables some limited modeling of theoretically known structures within the text (Roberts, Stewart, & Tingley, 2019), this selection is limited by our limited understanding of the statistical patterns in language use and their impact on the appropriate algorithm and preprocessing decisions.

Topic models have many practical applications in various fields. In natural language processing, topic models can automatically identify the main themes in a large corpus of text, allowing for the efficient organization and analysis of large amounts of unstructured data. In digital humanities, topic models can be used to identify patterns and trends in historical texts, providing insights into the evolution of language and culture over time. In marketing, topic models can be used to analyze consumer behavior and identify trends in customer preferences and feedback. In political science, topic models can be used to analyze the content of political speeches and identify the main themes and issues being discussed. Overall, topic models provide a powerful tool for automatically identifying and analyzing the main themes in large bodies of text.

### 4.1.3 Sentiment analysis

Sentiment analysis is a technique used in natural language processing (NLP) to identify the emotional tone of a piece of text, such as whether it is positive, negative, or neutral. This can be useful for a wide range of applications, such as analyzing customer feedback to identify patterns and trends, or automatically classifying social media posts by sentiment. There are various methods for

performing sentiment analysis, including dictionary-based approaches, which use pre-defined lists of words and their associated sentiments, and machine learning-based approaches, which involve training a model on labeled text data to learn the characteristics of different sentiments. Once trained, the model can then be applied to new text data to predict the sentiment of that text. Sentiment analysis can be useful for gaining insight into the attitudes and opinions expressed in large amounts of text data.

Machine learning models can also be trained to perform sentiment analysis by being fed a large amount of labeled text data. This allows the model to learn the patterns and characteristics associated with different sentiments. Once trained, the model can then be applied to new text data to predict the sentiment of that text.

## 4.2 Overview of methodological tools

### 4.2.1 Content analysis

Content analysis studies and interprets textual data by categorizing and quantifying its meaning using pre-defined categories. This method is based on the assumption that texts, despite their variability in language and meaning, can be equivalent in certain qualities that allow for their classification and aggregation. The content analysis process involves mapping theoretical constructs onto textual forms (Kutter & Kantner, 2012). The meaning of these texts is obtained through interpretation, and operationalizations of these constructs can be applied at two levels. The first level involves providing theoretical criteria and arguments to shape a shared understanding among human coders and guide their interpretation of texts. The second level involves formulating rules on the level of textual indicators and algorithmic procedures that both human coders and computers can apply.

The ability of human coders to comprehend theoretical concepts allows them to recognize valid classifications beyond those mandated by algorithms, but this subjectivity can also introduce inconsistencies and errors in the classification process. In contrast, while content analysts have long recognized the capacity of computers to classify textual contents in a consistent and scalable manner, these computational tools are limited by their inability to comprehend theoretical constructs (Skalski, Neuendorf, & Cajigas, 2017), which restricts their ability to move beyond their reliance on manifest indicators and imperfect algorithms.

In this study, I employ content analysis to identify key words, n-grams, hashtags and narrative frames and study their trends over time. This is done through exploratory data analysis, which includes examining the most popular tweets, the most frequent words, n-grams, and hashtags and topic modeling, which reveals the underlying themes and frames of the discussion over time. While this approach is simplistic, it provides a basic and reliable understanding of the Twitter discussion.

### 4.2.2 Hybrid Content Analysis

Hybrid Content Analysis (HCA) attempts to address the difficult trade-off between scalable computational strategies and insightful small-scale manual labeling (Baden, Kligler-Vilenchik, & Yarchi, 2020). Instead of manually classifying individual texts, HCA uses unsupervised pattern-finding algorithms, in other words, topic models, to identify regularities in the data and then manually classifies these patterns using coding schemes derived from theory-driven categories (see Figure 1). The authors argue that this approach allows researchers to use computational tools to reduce the effort required to analyze social media content massively and focus their manual work on the informed classification of a finite set of recurrent patterns.

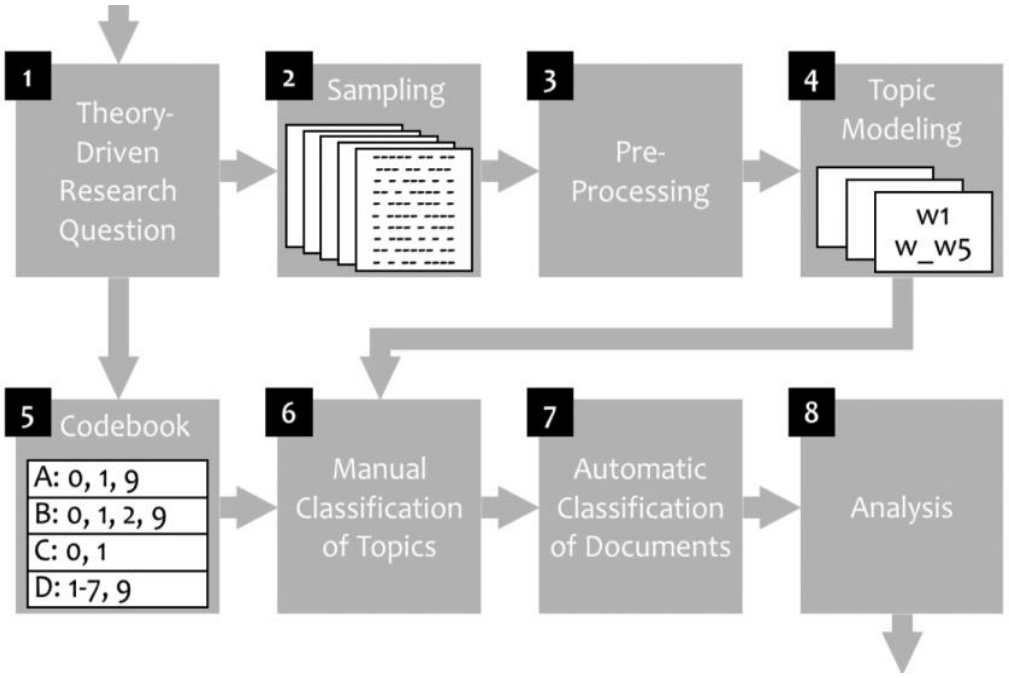


Figure 1. HCA flowchart

This method was showcased by the previously referred study by Baden et al. on political polarization on Israeli social media. The researchers have used the Structural Topic Model and examined the most probable words in each topic, as well as other word-weights such as score, frex, and lift, and representative documents to disambiguate meanings. This allowed them to identify and label topics based on their sentiments, stances, themes, coherence, referred actors, and topicality. These labels were then applied to tweets, with a pre-defined threshold determining whether a tweet would be considered as expressing a particular sentiment or stance. For example, if topics A and B were labeled as expressing negative sentiment and a tweet had a probability of 0.3 of belonging to topic A and a probability of 0.4 of belonging to topic B, and the threshold was set at 0.5, the tweet would be labeled as negative.

The authors used the consistent stances displayed by users in their tweets to label users with a particular stance. This allowed them to classify interactions

(replies, shares, etc.) as homophilic (between users with the same stances) or heterophilic (between users with different stances). They also examined affective polarization by comparing the sentiments expressed in ingroup and outgroup interactions, and used the themes most frequently referred to by a particular stance to measure positional polarization.

### 4.2.3 Unsupervised machine learning

A significant amount of research has focused on using machine learning models to address tasks such as stance detection, text classification, and sentiment analysis. However, there are challenges with replicating the results of these studies due to a lack of transparency in their implementation and the focus on specific datasets, platforms, and languages. One promising alternative is the embeddings-based clustering method proposed by Rashed et al. (Rashed, Kutlu, Darwish, Elsayed, & Bayrak, 2021). This method uses subword-level embeddings to represent users based on the content of their tweets on a particular topic. Subword embeddings also help address the Turkish language's morphological complexities and increase resistance to misspellings common on social media. After averaging embeddings, the authors project the users onto a lower-dimensional space, which results in similar users being brought closer together and dissimilar users being pushed further apart. They are then clustered to detect communities of users with similar stances. The authors found that their method effectively detects fine-grained stances toward different topics. As an example, they used their approach to study user stances towards polarizing issues in Turkish society and found correlations between positions across different topics. Additionally, the resulting clusters can be used to measure polarization on other topics of interest.

## 4.3 Data

### 4.3.1 Ebrar Karakurt dataset

All tweets were collected using search-tweets v2 package for Python (Pigott, Kolb, Montague, Gonzales, & Moffitt, 2021). For tweets on the Ebrar Karakurt case, queried keywords were “ebrar,” “lgbt,” “eşcinsel,” “lezbiyen” and their alternative spellings. The Twitter API lacks substring matching capability, resulting in potential issues with querying for specific terms such as "lezbiyen" which may not include related variations such as "lezbiyenler", “lezbiyenleri” or "lezbiyenlik". This issue is particularly significant for agglutinative languages like Turkish, where words can have an extensive number of suffixes and derivatives. Despite the importance of this matter, I found it has often been neglected or unreported in current literature. To address this issue, I have employed Zargan's 1.3M word dictionary to incorporate the sixty most frequently occurring forms for each spelling of the queried words, achieving comprehensive data collection as a result. Tweets were collected from thirty days, between 08-10-2021 and 09-10-2021, yielding a total of 85,322 original tweets from 48,909 users and 168,172 re-tweets from 116,658 users.

In order to obtain Turkish tweets exclusively, the built-in language filter on Twitter was utilized. However, it was discovered that this filter occasionally misclassified Azeri tweets as Turkish. To address this issue, a total of 222 Azeri tweets were identified and subsequently filtered out by examining the presence of the letter "schwa" (ə), which is not included in the Turkish alphabet and is not used by Turkish social media users in contrast to other out-of-alphabet letters such as x, q, and w. I have also used Twitter’s built-in “nullcast” filter to disallow paid advertisements.

Twitter bot and spam accounts are commonly known for generating low-quality content by combining trending hashtags or words in meaningless posts. To address this issue, I conducted an exploratory analysis of tweets using topic models, word and hashtag frequencies. This analysis allowed me to identify specific, irrelevant words frequently used by bot accounts, such as #tbtxtoken, #crypto, and Barzani, which were subsequently filtered out. Additionally, as some of the queried words were associated with sexuality, this resulted in an abundance of posts from porn accounts that were not relevant to the discussion, and were therefore filtered out. To further refine the filtering process, accounts associated with a coinciding charity campaign for a child suffering from SMA disease, also named Ebrar, were detected and removed by filtering out words associated with the disease and the campaign. In each stage, a manual evaluation of removed tweets was conducted through random sampling in order to ensure the accuracy of the filtering process. A total of 3,221 tweets were removed as a result.

### 4.3.2 Istanbul Convention dataset

In collecting the dataset on the Istanbul Convention, I employed the search queries "istanbulsozlesmesi", "istanbul AND sozlesmesi", as well as their related spellings and derivatives, following the methodology used in previous research. The "and" operator was utilized to identify tweets that contained both the terms "istanbul" and "sozlesmesi," regardless of their order. This approach facilitated the capture of relevant tweets that utilized the terms in various contexts, such as "istanbul ihanet sozlesmesi." However, to ensure the accuracy of the data, I employed a filter that permitted only one word between the terms. It is important to note that the filter was necessary to avoid matches such as "istanbul boğazı montrö sözleşmesi."

Additionally, tweets in Azeri and those identified as trend-spammers were removed from the dataset.

The date of 1st June 2019 has been determined as the pivotal moment when President Erdoğan first expressed reservations regarding the Istanbul Convention. To capture the discourse surrounding this critical event, tweets were collected until the publication of the presidential decree on 20 March 2021. To ensure a comparable timeframe for analysis, tweets dating back two years were also collected, commencing from 1 January 2017.

## 4.4 Measuring Polarization

In order to measure polarization, it is necessary to label user stances accurately. The baseline labels for this task will be obtained using the label propagation method. The effectiveness of these labels will be compared to those obtained using clustering-based and HCA-based methods. Once the stances have been detected, we can measure various dimensions of polarization. I follow Baden et al.'s approach to quantify and measure polarization (Baden, Yarchi, & Kligler-Vilenchik, 2021).

To study affective polarization, which focuses on how people feel about members of the opposing group, I will use sentiment analysis to detect the tone of users' messages when addressing ingroup versus outgroup members. I will evaluate three methods for performing this analysis: a traditional dictionary-based approach, an HCA-based method, and a language-model-based method.

Ideological polarization refers to the differences in the issues people are interested in and how they view them. In order to analyze ideological polarization on social media, I will use topic modeling, the most commonly used tool for this purpose.

Interactional polarization is often studied in terms of homophily and heterophily, which refer to the frequency of ingroup and outgroup interactions. To investigate this aspect of polarization, I will create a map of the reply, retweet, and mention networks of users participating in the discussion and identify the ratio of homophilic and heterophilic interactions within these networks.

Based on previous research, the following hypotheses are proposed:

H1: Affective polarization, as measured by sentiments expressed, will be more prevalent in heterophilic interactions.

H2: Users with opposing stances will use different or contrasting frames, indicating positional polarization.

H3: Homophilic interactions will be more common than heterophilic interactions, demonstrating interactional polarization.

H4: All measures of polarization will decrease following the intervention of the authority figure.

# CHAPTER 5

# ANALYSIS AND DISCUSSION

This chapter presents various analyses and discusses their connections. The previously described methods and their results will be reviewed, including an evaluation of their strengths and weaknesses. The smaller dataset on Ebrar Karakurt will be analyzed first to explain and measure the performance of competing methods, which will be followed by a more succinct analysis of the Istanbul Convention dataset. The final discussion will integrate all of the results and engage with relevant literature in order to address the research question and better understand the impact of interventions on social media discussions.

## 5.1 Case on Ebrar Karakurt

The case study is divided into four phases for analysis. Phase 0 encompasses the time period prior to Karakurt's Instagram post, from August 10th to August 14th at 20:30, serving as the neutral phase. Phase 1 begins with Karakurt's post on August 14th at 20:30 and includes the time period before official intervention, denoting the pre-intervention phase. Phase 2 begins with Mumcu's statement on August 16th at 10:30 and extends until August 19th, when tweet activity related to the topic significantly decreases, acting as the post-intervention phase. Period 3, the immediate aftermath, covers the three days starting on August 19$^{th}$, and is the post-event phase. Therefore, Phases 1 and 2 are of particular significance and will be the focus of our analysis, as Phase 1 represents the initial phase of polarization, while Phase 2 represents the time period following intervention.

### 5.1.1 Stance detection

Accurate labeling of user stances is a crucial aspect of measuring polarization. To address this task, I will utilize three different methods, as recommended by the literature, and compare their accuracy and validity for this case study. Due to the limited size of the dataset pertaining to the Ebrar Karakurt case, it will be utilized first to compare the performance of the selected methodologies.

#### 5.1.1.1 Semi-automated labeling

After conducting an exploratory analysis, I devised a semi-automated labeling scheme for the initial step of stance detection. A crucial aspect of this process is identifying reliable indicators for user stances and maximizing automation in the labeling process as much as possible. Numerous approaches are suggested in the literature for this purpose, including examining usernames and descriptions, using shared URLs as a feature, analyzing retweets, analyzing hashtags, analyzing word choices (e.g., vocabulary, ngrams, phrases), and examining follower/followee networks, among others.

To begin with, I searched for the terms "LGBT," "eşcinsel," "lezbiyen," "sapık," and "sapkın," as well as the "rainbow" and "pride flag" emojis, which are commonly used by the LGBT community, within the username and description sections of all the users in my dataset. This yielded 928 Pro-LGBT labels and 21 Anti-LGBT labels. The asymmetry in these results is expected, as being a part of the LGBT community is often a central aspect of one's identity and therefore tends to be prominently featured in usernames and descriptions, while being Anti-LGBT is less

likely to be emphasized. Next, I examined the tweets and retweets of all labeled users, and this method proved to be highly accurate and time-efficient.

For the second step, I examined all URLs shared in tweets to identify those that could serve as indicators. This aligns with literature suggesting that partisan news exposure and dissemination are correlated. For example, Bozdağ and Koçer found that even self-critical users tend to rely on sources that they feel more closely aligned with when sharing news on social media (Bozdağ & Koçer, 2022). A crucial aspect of this process was identifying sites with clear editorial positions on the issue under investigation. For example, while the oppositional Sözcü newspaper/news site is highly political, it does not have a clear stance on LGBT rights, unlike YeniAkit (Anti-LGBT) or KaosGL (Pro-LGBT). The URL labeling was tested using a stringent scoring parameter: for a user to be labeled based on URLs, they must share more than twice the number of URLs from the opposing group.

For the third step, I manually labeled tweets that were retweeted more than 100 times. This resulted in 240 tweets (122 Pro-LGBT, 118 Anti-LGBT) with clear stances on the issue. While a significant number of the most frequently retweeted tweets expressed a remarkable appreciation for Ebrar Karakurt as a player and the volleyball team, they were not coded as Pro-LGBT because they did not contain any such indication. By examining which users retweeted these tweets, I was able to label 36,585 users, with 24,893 being Pro-LGBT and 11,692 being Anti-LGBT. As with URLs, retweets were scored and had to exceed the threshold to be used for labeling (see Appendix A). Retweets are often considered to be among the most reliable indicators of stance, and my findings supported this (Kutlu, Darwish, & Elsayed, 2018).

Some studies have used vocabulary-based methods for stance detection, but these proved unreliable for this study. There were numerous keywords or phrases that could be associated with a particular stance, but the use of reported speech and irony resulted in conflicting scores, and the additional labels produced by this method were in the low thousands. Given these limitations, I chose to prioritize validity in my analysis.

In many studies on Twitter, hashtags associated with a particular position have been used as a feature for stance detection. However, such use of hashtags can be problematic because they may be hijacked or used by multiple parties, especially when a single dominant hashtag is used to represent an issue, such as the "BlackLivesMatter" hashtag. Although my dataset included popular and seemingly highly opinionated hashtags such as "#SapkınlığaDurDe" and "#EbrarKarakurtYalnızDeğildir," an examination of random samples of tweets posted under these hashtags revealed that the opposing camp used them at a significant frequency. This made hashtags an unreliable feature for stance detection.

Finally, while there have been many promising studies that utilize a follower/followee-based approach for stance detection, this method was not feasible for my study due to several limitations. First, the Twitter USER endpoint is much slower than the TWEETS endpoint and returns only one user per minute. Collecting the follower/followee data would have required two months of continuous data collection. Second, there is a high computational cost to networking 116,000 users, which a single, unfunded researcher cannot meet. Some studies have employed a more restricted approach that builds the follower networks of key parties/politicians and uses it to label lay users. However, the small size of key parties in this case made that method ineffective.

In total, description, URL, and retweet labeling provided 39,298 labels for the dataset. It is worth noting that a consonance/dissonance check was applied when finalizing the labels. This addresses instances where an URL label might conflict with an RT label, and so on. The final check helped me to refine the URLs and scoring parameters until the end result yielded no dissonance between stances.

As the final step, I applied a simple label propagation method, following the approach of Darwesh et al., which assigns labels through retweet networks (Stefanov, Darwish, Atanasov, & Nakov, 2019). For instance, an unlabeled user who retweets a Pro-LGBT user would be given the Pro-LGBT label. The algorithm ran five times until no new labels could be assigned, adding 5,027 labels. In addition to assigning new labels, the algorithm also served as a validity check, as it reported instances of consonance (i.e., a Pro-LGBT user sharing a Pro-LGBT tweet) or dissonance (a user labeled Pro-LGBT sharing a tweet labeled Anti-LGBT). In the end, this method validated 35,000 stances, while only 20 users were assigned conflicting stances. These users were dropped as their stance could not be inferred. Furthermore, a manual inspection of 1,000 random users revealed only three wrong labels, suggesting remarkable accuracy.

#### 5.1.1.2 Hybrid Content Analysis

While Hybrid Content Analysis allows for efficient categorization of large datasets, it may require extensive pre-processing depending on the chosen topic model. For this study, I used the Structural Topic Model (STM) and BERTopic, both commonly utilized in academic research. STM, in particular, is a popular choice due to its ease of use and implementation in the R programming language (Roberts, Stewart, & Tingley, 2019). In addition, it provides researchers with various metrics to evaluate topic quality, allows for different weightings of topic words, and includes

automated functions for determining an appropriate starting number of topics. As a critical modification, the authors of HCA employed an inversely weighted document feature matrix as their model's input to increase the importance of words unique to a particular topic while diminishing the significance of common words. I also found this performed better than the standard document feature matrix in my tests.

That said, STM required significant pre-processing to provide interpretable results. This is exacerbated by the lack of NLP resources for Turkish. As an agglutinative language, Turkish words can have multiple (almost arbitrary) suffixes, making it hard to lemmatize. Even the usually simple task of identifying stopwords turns challenging as they come in many forms and require an extensive vocabulary for detection. Moreover, Turkish grammar heavily relies on adverbs, which also connate negation. The existing Turkish NLP tools expect a normalized and grammatically correct input, which is most often not the case with tweets due to the informal nature of the conversation, the abundance of misspellings, abbreviations, and asciifying of Turkish characters.

I have iterated many times over the dataset to get the best results from the topic models. This required me to check the most frequent words and ngrams each time to identify stopwords and concatenate common ngrams and compound verbs. Additionally, I employed a selective lemmatization technique for the identified keywords to trim them down to the smallest meaningful part. Compared to standard lemmatization, this allows maintaining the negative words (e.g. “ilgilendirir” vs. “ilgilendirmez”) and conceptually different derivates (e.g. “ahlak” vs. “ahlaksız”) which are essential signifiers.

I also evaluated an alternative model using a lemmatized corpus. Since a reliable Turkish lemmatizer was unavailable, I modified the Zargan dataset as a

lookup source, adding asciified entries and negative forms to the dictionary, which increased its size from 1.6 million words to 2.2 million. That said, this did not yield any noticeable improvement in the quality of generated topics compared to selective lemmatization.

It should also be noted that while the preceding enhancements significantly improved topic model results, I still had to employ a more restrained approach to labeling than the original study to ensure validity. Most importantly, I have given a lot more prominence to the most representative documents assigned to topics. This was necessitated by the fact that even when a topic had multiple words signifying a label (e.g., “Pro-LGBT”), the top representative documents might convey an opposite meaning. This was a necessary trade-off between more labels and more validity, where I favored the latter.

BERTopic is a recent development in topic modeling and has gained increasing use in academic research. Unlike classical methods like STM, BERTopic generates document embeddings using pre-trained language models, projects them onto a lower-dimensional space, clusters them, and finally generates topic representations using a class-based term frequency-inverse document frequency (TF-IDF) procedure (Grootendorst, 2022).

One of the main strengths of BERTopic, its reliance on embedding models, can also be viewed as a limitation, as its performance is highly dependent on the availability of well-trained models. While BERTurk provides state-of-the-art performance as a transformer-based model for Turkish, BERTopic works best with sentence models, which are not currently available for Turkish at the time of writing. As an alternative, BERTopic can utilize multilingual models or average word embeddings provided by a BERT model. In this study, I tested BERTopic using

Google's Multilingual Universal Sentence Encoder, a 512-dimensional sentence model trained on 16 languages including Turkish (Cer, et al., 2018), and BERTurk (Schweter, 2020). I also "pre-trained" or "updated" a BERTurk model on my corpus through a process also referred to as "continual training" or "post-training," which involves adjusting the BERTurk embeddings based on the new corpus. This was done for two main reasons: first, similar to other base BERT models, BERTurk was trained on a corpus consisting primarily of formal texts, which can lead to reduced performance compared to models trained explicitly on Twitter data, as demonstrated by the comparison between BERT-English and BERTweet (Nguyen, Vu, & Tuan Nguyen, 2020). Additionally, evidence suggests improved performance on downstream tasks following pre-training. For the BERTopic corpus, I performed minimal pre-processing. This included cleaning up mentions, links, and punctuation (excluding hashtags), replacing numbers with a token, translating emojis into text, and correcting common misspellings.

Embedding-based topic models often struggle with grouping excessively similar words, resulting in topics comprising various forms of the same word (e.g., lezbiyen, lezbiyenler, lezbiyene, lezbiyenmiş), which is particularly prevalent in agglutinative languages. Fortunately, a recent update to BERTopic allows for considering word diversity when constructing topics and setting the diversity parameter to 0.5 resulted in significant improvements. Overall, while BERTopic produced informative and coherent topics, relying solely on them was still not possible to label user stances as proposed by Hybrid Content Analysis. An evaluation of candidate topics for stance detection revealed conflicting messages, similar to the results obtained with STM.

In conclusion, while topic models can provide valuable insights into a corpus and help track historical trends, using them for stance detection as suggested by Hybrid Content Analysis was not feasible in this context. Therefore, researchers should be cautious when using topic models and consider more than just the top-topic words. Examining the documents assigned to a particular topic is essential. In cases with clearly partisan topics, this may assist in the semi-automated labeling process.

#### 5.1.1.3 Unsupervised Stance Detection

Unsupervised Stance Detection was evaluated on the dataset through multiple parameters. In addition to using MUSE as in the original study, I also employed averaged BERTurk embeddings and a domain-trained BERTurk model. While the algorithm generated notable, distinct clusters as shown in Figure 2, a detailed examination based on semi-automated labels revealed that the clusters did not align with the stance labels, as observed in Figure 3. This discrepancy may be due to the relatively smaller size of my corpus.

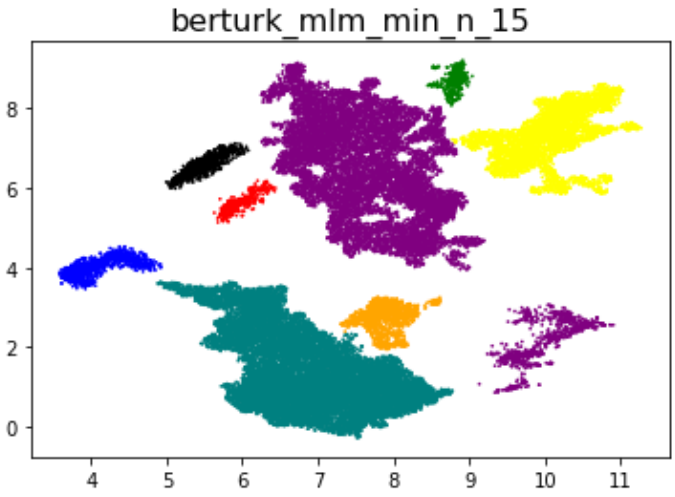


Figure 2. Embeddings colored by clusters

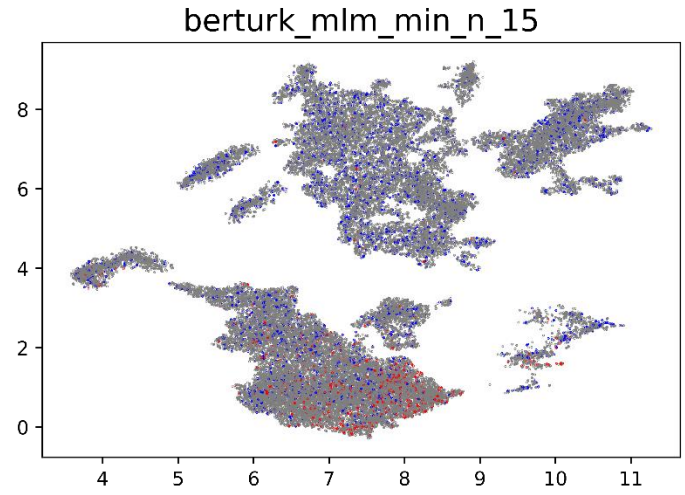

Figure 3. Embeddings colored by stance labels

In recap, stance detection proved to be a formidable task. While label propagation provided valid results, it left a sizable part of the dataset unlabeled. Hybrid Content Analysis provided less than acceptable accuracy. While unsupervised stance detection could cluster most of the users, low accuracy made it unusable.

### 5.1.2 Content analysis

In order to understand the predominant themes within the discussion, I identified the most frequent hashtags, words, and ngrams. While this does not necessarily reveal the users' stance, it provides insight into the conversation's principal axes and trends over time. For starters, a rough analysis of the number of tweets per date shows peaks following the statements and coincides with volleyball matches, as shown in Figure 4.

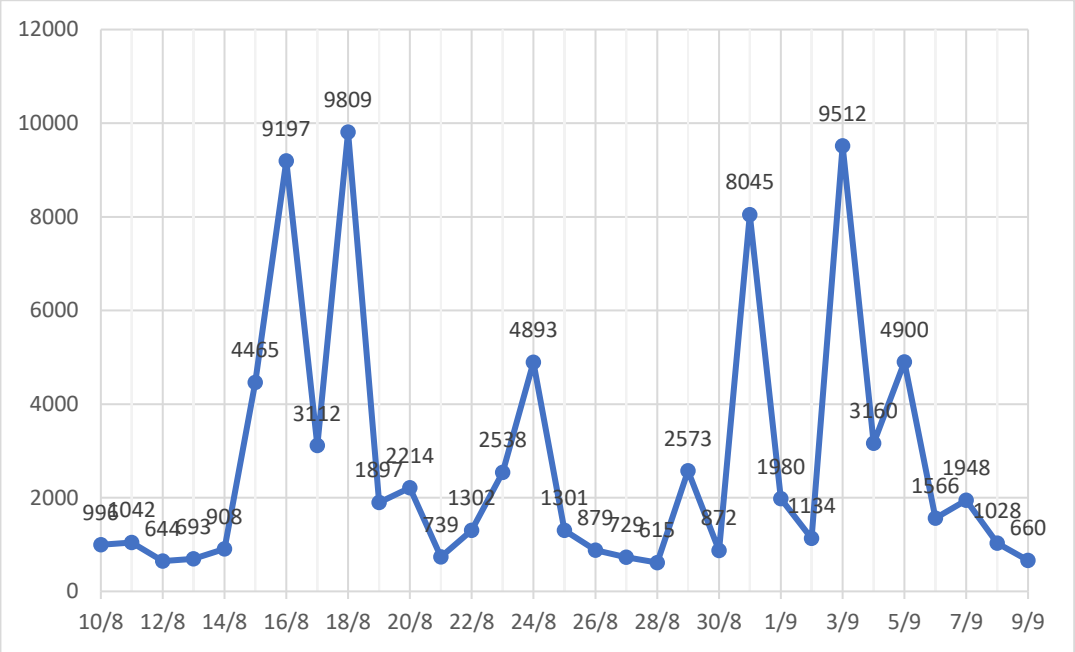


Figure 4. Tweet volume by date, Ebrar Karakurt

Table 1 presents an analysis of the hashtags, showing that the volleyball team was a significant focus of discussion. The official statement of the spokesperson did not have a significant impact, but the Pro-Ebrar tags saw significant growth following Cübbeli Ahmet's statement, as visible in Figure 5.

Table 1. Changes in Hashtag Frequencies from Phase 1 to Phase 2, Ebrar Karakurt

| Hashtag | Phase 1 | Phase 2 | % Change |
|---|---|---|---|
| #ebrarkarakurt | 1435 | 322 | -77.5 |
| #ebrarkarakurtyalnizdegildir | 33 | 414 | 1154.4 |
| #ebrarkarakurtyalnizdegil | 92 | 79 | -13.9 |
| #lgbtihaklariinsanhaklaridir | 7 | 79 | 1032.2 |
| #loveislove | 58 | 8 | -85.4 |
| #fileninsultanlari | 11 | 9 | -16.1 |
| #lgbtsapikliktir | 1 | 18 | 1669.7 |
| #ebrarkarakurttakimdanatilsin | 1 | 9 | 823.3 |

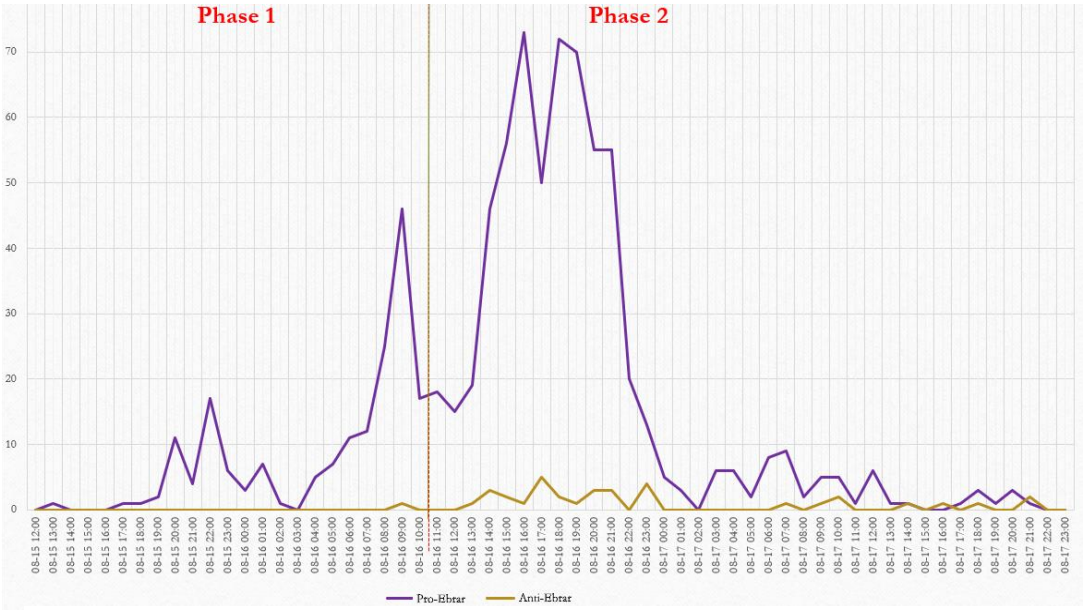

Figure 5. Hashtag trends in Phase 1 and Phase 2, red line marks Mumcu's statement

I have organized the key terms into categories of LGBT/sexuality, immorality, social norms, religion, national, rights-based, anti-conservative, and volleyball. Some of these categories contained neutral terms that simply described the topic being discussed (e.g., LGBT, national), while others contained emotionally and politically charged terms that indicated a more heated exchange (e.g., immorality, religion, anti-conservative). These themes demonstrate striking parallels to those observed in Ozduzen & Korkut's study.

Figure 6 reveals a strong inverse relationship between mentions of volleyball and mentions of LGBT-related themes, while others followed a much flatter trend. This suggests that the primary individuals discussing these topics often switch between discussing one or the other, with more discussion of volleyball occurring in match days and more discussion of LGBT themes occurring at other times. This finding suggests that the primary participants in discussing these themes may be the same people.

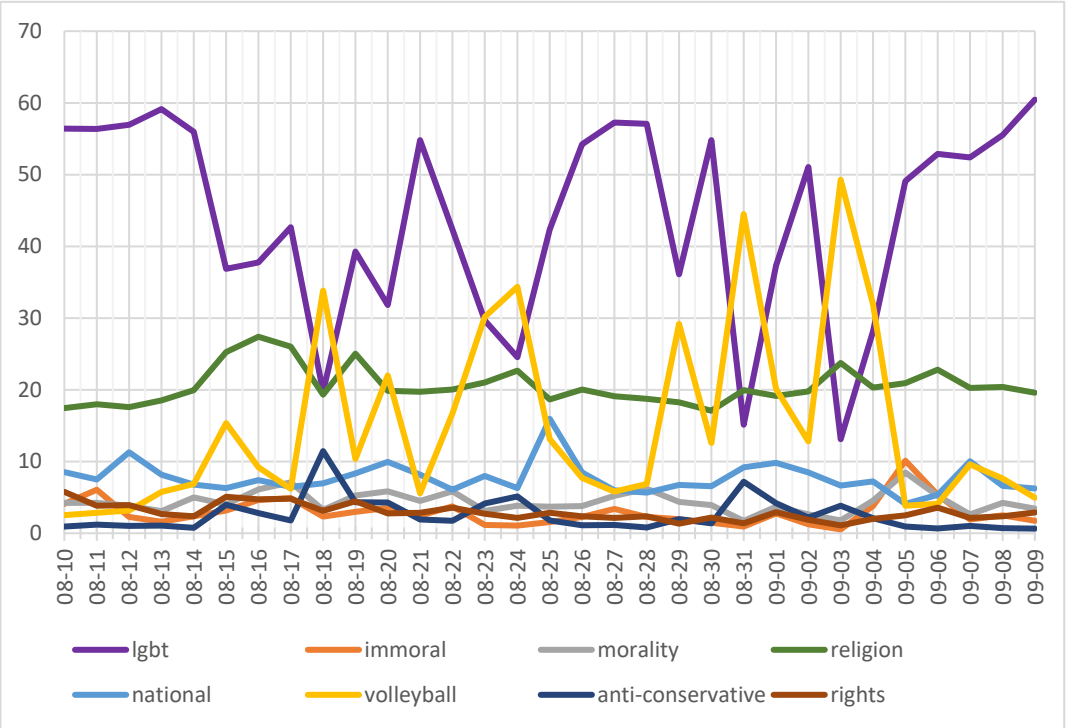


Figure 6. Keyword shares by date, Ebrar Karakurt

BERTopic with BERTurk embeddings was utilized to identify the topics present in a given dataset through topic modeling. As a starting point, the number of topics was set by the auto-reduction algorithm provided by BERTopic, which yielded 15 topics. However, it was observed that this process did not capture all of the key frames identified through exploratory analysis, leading to the decision to set the initial number of topics at the higher value of 100. While this may seem excessive, the similar and redundant topics yielded by such a higher number can be merged, while keeping the benefit of more granular topics. Upon examination of the key words and representative tweets for each identified topic, two additional frames were identified: praise for Ebrar Karakurt and praise for the national volleyball team. Ultimately, the topics were merged based on shared frames for the purpose of easier interpretation, resulting in a final number of 24 topics.

Topic modeling showed that messages with anti-LGBT themes frequently included political attacks against the opposition. These messages exhibited uniformly negative sentiments, as expected. Nationalist themes were present in both pro-LGBT and anti-LGBT messages, with terms such as "ataturkunkizlari" appearing in pro-LGBT messages and "sporcu, ahlakli" in anti-LGBT messages. Both sides appealed to morality, with anti-LGBT messages invoking family values and pro-LGBT messages addressing child abuse in religious communities. Among the coherent topics identified, those expressing praise for Ebrar or the team were most common. Pro-LGBT messages were divided in terms of sentiment, with positive sentiment appearing in topics related to "ask, sevgi" and negative sentiment in those centered around the term "yobaz."

Analyzing topics reveals well-defined frames that highlight the points of polarization, as shown in Table 2. For example, the anti-LGBT camp utilizes a variety of frames in their messaging, including a religious frame that condemns homosexuality, an "abnormality" frame that portrays homosexuality as a mental disorder, a social frame that emphasizes the perceived threat of homosexuality to family values and children, and a political frame linking homosexuality, opposition parties, and foreign actors. Additionally, there is a notable appeal to national themes, including references to Atatürk.

Table 2. Topics by Phase and Stance, Ebrar Karakurt

| | LGBT | Volleyball | Rights | Anti-Cons | Religious | Praise-Team | Praise-Ebrar | Immorality |
|---|---|---|---|---|---|---|---|---|
| | Pro / Anti | Pro / Anti | Pro / Anti | Pro / Anti | Pro / Anti | Pro / Anti | Pro / Anti | Pro / Anti |
| Phase 0 | 43 / 33 | 19 / 0 | 11 / 2 | 4 / 0 | 4 / 7 | 7 / 0 | 47 / 1 | 4 / 11 |
| Phase 1 | 86 / 25 | 6 / 0 | 83 / 1 | 5 / 0 | 11 / 14 | 8 / 0 | 190 / 3 | 2 9 |
| Phase 2 | 136 / 67 | 53 / 1 | 99 / 8 | 68 / 0 | 24 / 56 | 34 / 0 | 735 / 10 | 7 / 40 |
| Phase 3 | 53 / 19 | 27 / 0 | 4 / 0 | 10 / 0 | 6 / 2 | 15 / 0 | 148 / 0 | 4 / 8 |

The pro-LGBT side utilizes frames that emphasize privacy and personal rights, as well as secularism and the sanctity of love. They also appeal to social issues such as corruption, child abuse, and violence against women, accusing conservatives of remaining silent on these issues. Finally, there is a strong "othering" frame that groups anti-LGBT individuals with the opposing theme, framing victories over other national teams as victories over the anti-LGBT camp.

### 5.1.3 Sentiment and network analyses

For sentiment analysis, I have utilized the sentiment analysis model provided by Köksal and Özgür, which is based on BERTurk (Köksal & Özgür, 2021). This model, which was fine-tuned specifically on a Turkish twitter corpus, was found to be the best performing solution available at the time of writing. Lexicon-based approaches such as TurkishSentiNet were not deemed suitable for the analysis of noisy textual data, and transformer-based models were found to be more effective in general, even when analyzing formal text datasets (Dehkharghani, Saygin, Yanikoglu, & Oflazer, 2015).

Firstly, I will analyze the sentiments expressed in tweets to understand the attitudes and opinions of different groups of Twitter users during different periods in the timeline. Particular attention will be paid to tweets posted by authors who have been labeled according to their stance on the issue. The results presented on Table 2 show that the overall sentiments expressed in all phases and by all groups were negative. However, there were periods of marked change, with the positive-to-negative ratio increasing from 0.38 to 0.64 for tweets that were from Pro-LGBT users in Phase 1, likely reflecting a positive reaction to Ebrar Karakurt’s Instagram post. This trend continued in Phase 2, after Mumcu’s statement, although to a lesser

extent. Tweets posted by users who were labeled as Anti-LGBT were overwhelmingly negative, which is unsurprising given that opponents of an issue are unlikely to express positive sentiments on it.

Table 3. Sentiments by Phase and Stance, Ebrar Karakurt

| | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Pro-LGBT | 98 | 259 | 0.38 | 490 | 760 | 0.64 | 1030 | 1433 | 0.719 | 219 | 379 | 0.58 |
| Anti-LGBT | 10 | 198 | 0.05 | 11 | 196 | 0.06 | 36 | 783 | 0.046 | 12 | 156 | 0.08 |
| Total | 471 | 2560 | 0.18 | 2210 | 4987 | 0.44 | 5283 | 12176 | 0.434 | 1011 | 2672 | 0.38 |

In the next part of the study, the sentiments expressed in user interactions, starting with replies, will be analyzed. The results showed that negative sentiments were prevalent even in interactions between individuals who share similar views on the subject, highlighting its polarizing nature. As before, interactions between users who expressed pro-LGBT sentiments were an outlier with a higher presence of positive sentiments. This trend persisted from Period 0 to Period 1, similar to what was observed in the analysis of tweets. The majority of reply interactions remained homophilic, with users interacting with others who share similar views. However, there was a notable increase in heterophilic interactions in Period 2, with the homophily-to-heterophily ratio dropping from a typical value of approximately 10 to 1.6. While this may suggest a decrease in interactional polarization, it should be noted that the sentiments expressed in these interactions remained predominantly negative, meaning that we cannot conclude that they represented a step towards depolarization.

Table 4. Sentiments in Replies, Ebrar Karakurt

| REPLIES | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 8 | 38 | 0.2 | 32 | 57 | 0.6 | 41 | 82 | 0.5 | 15 | 27 | 0.6 |
| | Anti-Anti | 0 | 19 | 0 | 1 | 9 | 0.1 | 2 | 46 | 0.04 | 1 | 10 | 0.1 |
| Heterophilic | Pro-Anti | 0 | 4 | 0 | 1 | 3 | 0.3 | 3 | 73 | 0.04 | 0 | 3 | 0 |
| | Anti-Pro | 0 | 2 | 0 | 0 | 7 | 0 | 0 | 33 | 0 | 0 | 6 | 0 |

Quotes allow a user to share another user's tweet along with their own comment. This is distinct from a retweet, which represents a simple share of the original tweet. Despite comprising the smallest dataset among interactions, quotes revealed the most dissimilar results in two respects, visible in Table 4. Firstly, quotes exchanged between users who both expressed pro-LGBT sentiments had a positive-to-negative ratio of 0.94. Secondly, there was a higher prevalence of heterophilic interactions, or interactions between users with opposing views, in Periods 2 and 4. Given the small sample size, it was possible to conduct a more detailed analysis, which revealed that the heterophilic interactions between pro- and anti-LGBT users in Period 2 were in response to anti-LGBT tweets posted by religious figures.

Table 5. Sentiments in Quotes, Ebrar Karakurt

| QUOTES | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 6 | 6 | 1 | 24 | 12 | 2 | 18 | 29 | 0.62 | 2 | 9 | 0.22 |
| | Anti-Anti | 0 | 8 | 0 | 0 | 2 | 0 | 0 | 16 | 0.00 | 0 | 3 | 0 |
| Heterophilic | Pro-Anti | 0 | 0 | NA | 0 | 0 | NA | 4 | 47 | 0.09 | 0 | 0 | NA |
| | Anti-Pro | 0 | 0 | NA | 0 | 1 | 0 | 1 | 6 | 0.17 | 0 | 2 | 0 |

Finally, the analysis of mention interactions revealed a similar overall pattern to what was observed in other forms of interaction, with a higher-than-usual positive-to-negative ratio in mentions between users who both expressed pro-LGBT

sentiments. Also, there was an increase in heterophily albeit with a predominantly negative sentiment, as observed in Table 5.

Table 6. Sentiments in Mentions, Ebrar Karakurt

| MENTIONS | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 6 | 9 | 0.67 | 35 | 68 | 0.51 | 56 | 56 | 1 | 12 | 37 | 0.32 |
| | Anti-Anti | 2 | 26 | 0.08 | 0 | 10 | 0 | 2 | 71 | 0.03 | 1 | 7 | 0.14 |
| Heterophilic | Pro-Anti | 0 | 4 | 0 | 1 | 1 | 1 | 4 | 88 | 0.05 | 0 | 2 | 0 |
| | Anti-Pro | 0 | 1 | 0 | 0 | 11 | 0 | 0 | 40 | 0 | 0 | 7 | 0 |
| Mixed | Pro-Mixed | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 4 | 0.25 | 1 | 2 | 0.5 |
| | Anti-Mixed | 0 | 0 | NA | 0 | 0 | NA | 0 | 10 | 0 | 0 | 0 | NA |

### 5.1.4 Discussion

In the case analyzed, the authority figure had no notable impact on decreasing polarization on social media, polarization persisted or even intensified after the intervention. The reach and appeal was contained to one side. The low political power of the TVF spokesperson may have contributed to the lack of impact on depolarization.

The analysis of user interactions shows a dominance of homophily, i.e., the tendency for individuals to interact with others with similar views. While the large size of the unlabeled majority prevents firm conclusions, it is clear that homophilic interactions are more common than heterophilic interactions. It was also observed that the number of labeled stances increased significantly during Period 2, which could indicate that users with strong opinions joined the conversation or those already participating became more open about their position. Notably, influential anti-LGBT actors joined during Phase 3, which could explain the significant increase in anti-LGBT interactions.

While the literature suggests homophilic interactions to have more positive sentiment, our findings do not fully validate this. The most positive sentiment was observed within Pro-LGBT homophilic interactions, yet even in the best case, it was barely equal to the that of negative sentiments. On the other hand, Anti-LGBT homophilic interactions were predominantly negative, suggesting that the nature of the subject plays a significant role in determining the sentiments expressed. Previous research has indicated that homophilic interactions tend to have more positive sentiment. However, our findings do not fully support this conclusion. Pro-LGBT homophilic interactions demonstrated the most positive sentiment, yet even in this case, it was only slightly higher than negative sentiment. In contrast, Anti-LGBT homophilic interactions were predominantly negative, indicating that the subject matter may play a significant role in determining the sentiments expressed.

To summarize, the first hypothesis predicted that heterophilic interactions would have negative sentiments. While our findings do not contradict this, it is important to note that Anti-LGBT homophilic interactions had negative sentiments comparable to those of heterophilic interactions. Therefore, a more nuanced approach is necessary when examining homophilic interactions.

The second hypothesis proposed that opposing stances would employ different frames. Analysis of the tweets by users with opposing stances showed that this was the case. In both keyword groups and topics, users with opposing stances used different frames in their arguments. However, there were also areas or frames that were contested, such as national imagery, which was used by both sides.

The third hypothesis on interactional polarization simply posited that homophilic interactions would be more common than heterophilic interactions, and our findings support this conclusion.

The fourth and final hypothesis addressed the core research question of whether authority figure statements do have an effect on polarization. This hypothesis was tested through various metrics. The intervention did not appear to have an impact on affective polarization, as negative sentiments continued to dominate the discussion. Similarly, the politically and emotionally charged frames and keywords were still used by users, indicating no change in positional polarization. There was an increase in heterophilic interactions in Phase 2, however, these interactions exclusively expressed negative sentiments and therefore may not indicate depolarization. In conclusion, the intervention by the authority figure had no significant effect on polarization.

## 5.2 Case on the Istanbul Convention

The second case study is again subdivided into four distinct phases, albeit with a slight deviation from the previous one. Phase 0 represents the neutral pre-event phase, prior to the formal announcement of the withdrawal. The subsequent phases are demarcated by the president's remarks on the withdrawal process. Unlike the prior case, there is no discernible point of intervention in this instance.

### 5.2.1 Stance detection

Drawing upon the insights garnered from the previous case, I have decided to only employ the Semi-Automated Labeling technique for the Istanbul Convention dataset. An exploratory analysis of the influential tweets and hashtags provided suitable identifiers for the contrasting stances. I have explored the identifiers "istanbul sozlesmesi yasatir", "feminist", "kadin haklari", "vazgecmiyoruz", "istanbul sozlesmesini uygula", "6284u uygula", and the female symbol "♀" as indicators of a

pro-Convention stance. Correspondingly, I have searched for "aileyi koru", "aileme dokunma", "once aile", "aile meclisi", "iftira yasasi", "anti-feminist", "feminist teror", "genc evli", "suresiz nafaka", "cocuk haczi", and "6284magduru" to identify users espousing an anti-Convention stance. This process yielded a cohort of 520 pro-Convention and 157 anti-Convention users.

Subsequently, I proceeded to label users based on their shared URLs. Specifically, I analyzed the most frequently shared URLs and domains, and assigned labels based on the evident editorial stance. Through this process, I identified 44 anti-Convention and 62 pro-Convention sites, allowing the labeling of 5722 pro-Convention and 2316 anti-Convention users who shared them (see Appendix B).

As the third step in the labeling process, tweets that were retweeted over 100 times were examined and assigned labels. This resulted in the identification of 109 pro-Convention and 77 anti-Convention tweets, yielding 8952 pro-Convention and 4537 anti-Convention users who retweeted them.

Finally, by propagating the assigned labels through retweets, the final tally of pro-Convention and anti-Convention users was determined to be 89817 and 57956, respectively.

### 5.2.2 Content analysis

Through exploratory analysis of keyword, hashtag, ngram frequencies, and topic modeling, several keywords were identified. The key themes identified were "gender-lgbt," "morality," "religion," "family," "nation," "women's rights," and "anti-convention." The identified themes bear similarity to those of the previous case, which is expected as there is some overlap between the subject of LGBT rights and women’s rights.

When examining the use of keywords within the discussions, it is evident that religious themes dominate, followed by themes relating to family and the nation. However, the latter themes gradually lose their prevalence in subsequent phases.

An analysis of hashtags reveals that anti-convention hashtags can be categorized into three groups, consisting of calls for the protection of family, expressions of perceived injustice against men, and demands for withdrawal. In contrast, pro-convention hashtags can be categorized into two groups, consisting of denouncement of violence against women and direct references of the convention's significance, as demonstrated through the widespread use of the hashtag #istanbulsozlesmesiyasatir (see Table 7). Despite the popularity of this hashtag within the total dataset, it is worth noting that it only gained traction during the final phase of the conversation, whereas the anti-convention stance was also consistent in the earlier phases.

Table 7. Hashtag Frequencies, Istanbul Convention

| Hashtag Group | Phase 0 | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|---|
| Pro-Convention | 0 | 175 | 28 | 22220 |
| Against Violence | 4 | 63 | 31 | 1836 |
| Anti-Convention | 1 | 233 | 592 | 1269 |
| Family | 14 | 265 | 101 | 141 |
| Grievances | 40 | 347 | 391 | 846 |

Furthermore, the themes revealed through topic modeling, as seen in Table 8, indicate that the initial phase focused heavily on highlighting injustices against men, specifically in relation to "perpetual alimony," "slander," and "child foreclosure." The framing of family and gender also emerged as important themes throughout the conversation, both of which are areas of contention between the two opposing sides. Religious themes, as before, were utilized heavily by the anti-convention users.

Table 8. Topics by Stance and Phase, Istanbul Convention

| | Gender | Grievances | Violence | Religious | Family | Immorality |
|---|---|---|---|---|---|---|
| | Pro-Con / Anti-Con | Pro-Con / Anti-Con | Pro-Con / Anti-Con | Pro-Con / Anti-Con | Pro-Con / Anti-Con | Pro-Con / Anti-Con |
| Phase 0 | 2 / 54 | 0 / 24 | 2 / 0 | 0 / 7 | 9 / 53 | 4 / 11 |
| Phase 1 | 42 / 92 | 2 / 168 | 24 / 1 | 3 / 28 | 24 / 283 | 2 / 9 |
| Phase 2 | 48 / 41 | 2 / 442 | 29 / 8 | 11 / 92 | 11 / 144 | 7 / 40 |
| Phase 3 | 234 / 184 | 31 / 310 | 306 / 23 | 18 / 122 | 122 / 286 | 4 / 8 |

### 5.2.3 Sentiment and network analyses

The network interactions and the corresponding sentiments expressed therein are consistent with those observed in the preceding case, as heterophilic interactions constitute a clear majority, with homophilic interactions having a more adversary tone. Nevertheless, it is noteworthy that the overall tenor of these interactions leans towards the negative end of the spectrum. There is a marked in the final phase for Pro-Con tweets and homophilic interactions, which is caused by the emergence of "#istanbulsozlesmesiyasatir" hashtag, which conveys a positive tone.

Table 9 shows the overall sentiments expressed in tweets, while Table 10, Table 11 and Table 12 present the sentiments expressed in reply, quote and mention interactions, respectively.

Table 9. Tweet Sentiments by Phase, Istanbul Convention

| | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Pro-Con | 4 | 85 | 0.05 | 77 | 741 | 0.1 | 39 | 414 | 0.09 | 1916 | 6957 | 0.28 |
| Anti-Con | 31 | 680 | 0.05 | 236 | 6553 | 0.04 | 252 | 7261 | 0.03 | 662 | 14728 | 0.04 |

Table 10. Sentiments in Replies, Istanbul Convention

| | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 0 | 5 | 0 | 7 | 82 | 0.09 | 4 | 33 | 0.12 | 93 | 624 | 0.15 |
| | Anti-Anti | 10 | 135 | 0.07 | 56 | 1428 | 0.04 | 50 | 1422 | 0.04 | 144 | 2893 | 0.05 |
| Heterophilic | Pro-Anti | 1 | 0 | NA | 1 | 31 | 0.03 | 2 | 7 | 0.29 | 5 | 123 | 0.04 |
| | Anti-Pro | 1 | 13 | 0.08 | 2 | 150 | 0.01 | 2 | 79 | 0.03 | 17 | 475 | 0.04 |

Table 11. Sentiments in Quotes, Istanbul Convention

| | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 0 | 8 | 0 | 2 | 29 | 0.07 | 4 | 23 | 0.17 | 260 | 503 | 0.52 |
| | Anti-Anti | 0 | 31 | 0 | 15 | 350 | 0.04 | 9 | 307 | 0.03 | 35 | 933 | 0.04 |
| Heterophilic | Pro-Anti | 1 | 1 | 1 | 2 | 17 | 0.12 | 2 | 7 | 0.29 | 6 | 74 | 0.08 |
| | Anti-Pro | 0 | 2 | 0 | 1 | 22 | 0.05 | 2 | 10 | 0.20 | 1 | 83 | 0.01 |

Table 12. Sentiments in Mentions, Istanbul Convention

| | | Phase 0 | | | Phase 1 | | | Phase 2 | | | Phase 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N | Pos | Neg | P/N |
| Homophilic | Pro-Pro | 0 | 12 | 0 | 8 | 153 | 0.05 | 5 | 50 | 0.1 | 348 | 982 | 0.35 |
| | Anti-Anti | 28 | 556 | 0.05 | 222 | 4555 | 0.05 | 183 | 5886 | 0.03 | 702 | 9543 | 0.07 |
| Heterophilic | Pro-Anti | 2 | 1 | 2 | 0 | 104 | 0 | 2 | 22 | 0.09 | 19 | 296 | 0.06 |
| | Anti-Pro | 1 | 39 | 0.03 | 3 | 294 | 0.01 | 1 | 190 | 0.01 | 27 | 1067 | 0.03 |
| Mixed | Pro-Mixed | 0 | 0 | NA | 0 | 6 | 0 | 0 | 3 | 0 | 1 | 34 | 0.03 |
| | Anti-Mixed | 0 | 116 | 0 | 9 | 205 | 0.04 | 7 | 237 | 0.03 | 304 | 421 | 0.7 |

### 5.2.4 Discussion

The present case poses a challenge in analyzing the impact of authority intervention, as the conversation gains momentum only subsequent to the president's statement. Nevertheless, this can be construed as a form of influence, illustrating the capacity of an influential figure to swiftly steer the conversation towards a more polarizing subject matter. As with the preceding case, however, the efficacy of the

authority figure's influence is not evenly distributed among both sides of the argument. While the opposition towards the Istanbul Convention within the anti-AKP camp diminishes in the aftermath of the statement, the views held by those in support of the Convention remain unaltered and impervious to any such attempts at mediation. The users organized around the grievances theme were influential in directing the course of discussion, even when much of these grievances did not have anything to do with the convention itself.

With respect to the hypotheses posited in this study, the results obtained from the present case closely resemble those of the preceding one. The heterophilic interactions are predominantly characterized by negative sentiment, which serves as evidence of affective polarization (H1). The majority of the interactions take place between users who share similar opinions, thereby reflecting interactional polarization (H3). Moreover, the divergent themes utilized by the opposing sides instead of engaging in a discourse centered around shared themes indicate positional polarization (H2). Furthermore, the measures of polarization do not indicate any significant decline following the pronouncements made by authoritative figures, thereby again rejecting our fourth hypothesis.

# CHAPTER 6

# CONCLUSION

In the final chapter of this thesis, I will discuss key points to conclude the study in four sections. First, I will provide a summary of the thesis. An examination of the limitations of the study will follow this. Lastly, I will propose potential avenues for future research and consider the potential implications of the thesis.

## 6.1 Overview of the thesis

This thesis began by providing a review of the existing literature on political polarization. In Chapter 2, the concepts of ideological and affective polarization were explained, and key studies in the field were summarized. The role of social media in political polarization was then examined, including the concepts of echo bubbles, selective exposure, and partisan sorting, informing the reader of the current state of the literature. Finally, the state of research on political polarization in Turkey was examined, identifying the gap in the literature for large-scale studies on social media that utilize the latest methodological developments.

In Chapter 3, I introduced the cases under investigation. The first case was described as a fusion of two polarizing issues (secularism and LGBT rights), and how it transformed a previously unifying subject, national sports, into a divisive one. The second case showed how a previously unifying subject, prevention of violence against women, was quickly polarized. A brief overview of the platform of focus, Twitter, was then provided.

Chapter 4 began with a brief overview of tools frequently used in computational social science, followed by a detailed examination of the specific methodological challenges of the study and the tools chosen to address them. The critical issue addressed was stance detection, i.e., identifying the position a user takes on a particular issue. From the various methods proposed in the literature, semi-automated labeling, hybrid content analysis, and unsupervised stance detection were selected due to their scalability and adaptability to different languages and datasets. The subsequent tasks of sentiment and topical analysis of the tweets were also addressed, including standard dictionary-based methods and recent language-model-based approaches for sentiment analysis and a brief description of topic models for topical analysis.

Chapter 4 continued with a description of the data collection and cleaning processes for each case. The preprocessing approach was outlined, which included minimal pre-processing-up and spam detection. The chapter concluded with an overview of the metrics to detect polarization, including sentiment scores for measuring affective polarization, topical analysis for ideological polarization, and network analysis for interactional polarization.

In Chapter 5, the data and tools described in the previous chapter were applied. While unsupervised stance detection and hybrid content analysis failed to provide accurate stances, semi-automated stance detection allowed the labeling of many users. The content analysis for each case then began with a basic examination of hashtags and word frequencies, followed by a dynamic analysis of the topics featured in the discussion. This analysis revealed the contrasting frames employed by the opposing sides, some of which were claimed by both and were contested. Finally,

the interactions between users through retweet networks and their evolution over time were analyzed, demonstrating both homophily and increasing polarization.

## 6.2 Limitations

The field of computational social science is characterized by a wealth of innovative methods and perspectives, but this state of flux also means that there are few established and reliable guidelines. As is often the case in emerging fields of research, replication is needed to establish the reliability and validity of these methods. This was a significant issue encountered during the course of this study, as many promising approaches failed to deliver the expected results. Therefore, it is essential for researchers in this field to be mindful of the need for replication and to carefully evaluate the robustness of new methods before adopting them in their own work.

As a study that approaches the interdisciplinary field of computational social science from a social science perspective, this thesis is also limited by the current state of the computational field. This is particularly evident in the case of Turkish NLP tools, which are scarce and not readily available to researchers. While many promising methodological papers suggest innovative approaches, the lack of ready-to-use packages means that these methods are often difficult to implement without the necessary engineering expertise and resources, creating a significant barrier for researchers without this background.

Noisy text on social media presents a challenge for natural language processing in all languages, but it is especially challenging in Turkish due to the asciified nature of the text. Normalization is often used to address this issue, but available Turkish normalizers have been unsuccessful in improving the text and have

even caused further problems. This is due to the fact that normalizers expect non-asciified text and often incorrectly transform out-of-vocabulary words. To overcome this challenge, either a powerful normalizer or a language model trained on an informal corpus is needed. Given the current focus on language models and the success of BERTweet, the latter may be a more robust and future-proof solution.

As is often the case, some computational methods require significant computing power, which can restrict the feasible research approaches. This was the primary reason the current study could not create an entire network of all users within the dataset.

The unique and exceptionally brief nature of the first case made it difficult to draw firm conclusions. This is because such a limited and specific sample may not represent broader trends or phenomena and may not provide sufficient evidence to support broader claims or generalizations. The second case, on the other hand, lacked a specific intervention against polarization. For the sake of a clear research design and operationalization, both cases forego non-official authorities, such as opinion leaders, religious figures and celebrities, who may have comparable or even larger effect than an official figure. It is essential to consider the limitations of individual cases and be cautious when attempting to draw conclusions that extend beyond the study's specific context. In order to arrive at more robust and reliable conclusions, it is often necessary to examine a more extensive and diverse sample over a longer period of time. It should be noted that different mediums may create different patterns of interaction and polarization.

## 6.3 Implications and venues for further research

This study presents a framework for researchers to examine political polarization and its manifestation on social media. It does so by focusing on the impact of a statement made by an authority figure on a divisive issue on the tone of the conversation, labeling of stances of parties involved in the discussion and measuring their patterns and nature of interaction. Further research that employs a similar perspective on other cases, such as those involving more influential authority figures, less divisive social events, or those taking place on different platforms, could enhance our understanding of political polarization and, importantly, ways to address it.

## 6.4 Conclusion

Political polarization remains a significant issue for democracies around the world, both developed and developing. This takes place during a time when society is transitioning towards digitalization, with a pace so fast that the revolutionary platforms of yesterday, such as Facebook and Twitter are having a hard time keeping up. Researchers and members of civil society are also struggling to understand the different modes of interaction taking place on these platforms. Each platform has its own unique challenges and manifestations of polarization, and may require different strategies to address it. This thesis aims to thoroughly examine polarization through social media and provide a methodology for social scientists to continue studying this topic. The findings of this study indicate that addressing polarization is a complex task and simple messaging by political elites is insufficient. While it is unclear what will replace the current media landscape, it is likely that new platforms that build upon current ones with novel modes of interaction will emerge. It is

therefore essential to understand the platforms of the day in order to prepare for the next phase of digital communication.

Overall, it is likely that addressing polarization on social media will require a multifaceted approach that combines several different strategies. Further research is needed to identify the most effective methods for combating polarization in the digital age.

# APPENDIX A

## URL LABELS, EBRAR KARAKURT

| | | | |
|---|---|---|---|
| aylikdergisi.com | Anti-LGBT | cumhuriyet.com.tr | Pro-LGBT |
| haberkurulu.com | Anti-LGBT | esitlikadaletkadin.org | Pro-LGBT |
| habervakti.com | Anti-LGBT | marksist.org | Pro-LGBT |
| yenidenrefahpartisi.org.tr | Anti-LGBT | bianet.org | Pro-LGBT |
| bncmedyahaber.com | Anti-LGBT | evrn.sl | Pro-LGBT |
| akittv.com.tr | Anti-LGBT | gmag.com.tr | Pro-LGBT |
| dogruhaber.com.tr | Anti-LGBT | ilerihaber.org | Pro-LGBT |
| farklibakis.net | Anti-LGBT | kaosgl.org | Pro-LGBT |
| habermotto.com | Anti-LGBT | kayiprihtim.com | Pro-LGBT |
| hasbahcegazetesi.com | Anti-LGBT | sendika.org | Pro-LGBT |
| ilkkursun.site | Anti-LGBT | siyasihaber6.org | Pro-LGBT |
| kadem.org.tr | Anti-LGBT | sol.org.tr | Pro-LGBT |
| diyanet.gov.tr | Anti-LGBT | t24.com.tr | Pro-LGBT |
| haber7.com | Anti-LGBT | tele1.com.tr | Pro-LGBT |
| barandergisi.net | Anti-LGBT | birgun.net | Pro-LGBT |
| cubbeliahmethoca.com | Anti-LGBT | demokrathaber.org | Pro-LGBT |
| gunisigigazetesi.net | Anti-LGBT | diken.com.tr | Pro-LGBT |
| haksozhaber.net | Anti-LGBT | dokuz8haber.net | Pro-LGBT |
| ihvanlar.net | Anti-LGBT | evrensel.net | Pro-LGBT |
| milatgazetesi.com | Anti-LGBT | gazeteduvar.com.tr | Pro-LGBT |
| milligazete.com.tr | Anti-LGBT | gazeteemek.net | Pro-LGBT |
| yenisafak.com | Anti-LGBT | gercekgundem.com | Pro-LGBT |
| ysafak.com | Anti-LGBT | queercyprus.org | Pro-LGBT |

# APPENDIX B

## URL LABELS, ISTANBUL CONVENTION

| | | | |
|---|---|---|---|
| star.com.tr | Anti-Con | artigercek.com | Pro-Con |
| cocukaile.net | Anti-Con | gazeteduvar.com.tr | Pro-Con |
| m.star.com.tr | Anti-Con | amnesty.org.tr | Pro-Con |
| gozcuhaber.com | Anti-Con | journo.com.tr | Pro-Con |
| milligazete.com.tr | Anti-Con | sendika.org | Pro-Con |
| yenisafak.com | Anti-Con | ekmekvegul.net | Pro-Con |
| aileakademisi.org | Anti-Con | teyit.org | Pro-Con |
| furkangurbetoglu.wordpress.com | Anti-Con | evrimagaci.org | Pro-Con |
| m.haber7.com | Anti-Con | sol.org.tr | Pro-Con |
| risalehaber.com | Anti-Con | dokuz8haber.net | Pro-Con |
| islamianaliz.com | Anti-Con | anitsayac.com | Pro-Con |
| ilkha.com | Anti-Con | diken.com.tr | Pro-Con |
| hudapar.org | Anti-Con | kesk.org.tr | Pro-Con |
| m.seslimakale.com.tr | Anti-Con | birgun.net | Pro-Con |
| dogruhaber.com.tr | Anti-Con | esikplatform.net | Pro-Con |
| haber7.com | Anti-Con | evrensel.net | Pro-Con |
| memursen.org.tr | Anti-Con | ilerihaber.org | Pro-Con |
| turkgun.com | Anti-Con | yeniyasamgazetesi2.com | Pro-Con |
| ilkkursun.site | Anti-Con | dogrulukpayi.com | Pro-Con |
| hakses.net | Anti-Con | esitlikicinkadinplatformu.net | Pro-Con |
| haberpars.com | Anti-Con | haber.sol.org.tr | Pro-Con |
| akittv.com.tr | Anti-Con | kadincinayetlerinidurduracagiz.net | Pro-Con |
| ebs.org.tr | Anti-Con | kadinsavunmasi.org | Pro-Con |
| haksozhaber.net | Anti-Con | yeniyasamgazetesi3.com | Pro-Con |
| aydinlik.com.tr | Anti-Con | istanbulbarosu.org.tr | Pro-Con |
| nevzattarhan.com | Anti-Con | kisadalga.net | Pro-Con |
| habervakti.com | Anti-Con | 5harfliler.com | Pro-Con |
| kuran-ikerim.org | Anti-Con | bianet.org | Pro-Con |
| turkiyeailemeclisi.net | Anti-Con | egitimsen.org.tr | Pro-Con |
| anadolugenclik.com.tr | Anti-Con | politikyol.com | Pro-Con |
| | | birikimdergisi.com | Pro-Con |
| | | gazetekarinca.com | Pro-Con |
| | | kaosgl.org | Pro-Con |

# APPENDIX C

## TRANSLATION TABLE FOR EMOJIS

```
 ":)": ":gülen_surat:",
 ":1st_place_medal:": ":altın_madalya:",
 ":3rd_place_medal:": ":bronz_madalya:",
 ":Czechia:": ":çekya:",
 ":Italy:": ":italya:",
 ":Netherlands:": ":hollanda:",
 ":OK_hand:": ":iyi_işareti:",
 ":Poland:(": ":polonya:",
 ":Romania:": ":romanya:",
 ":Sweden:": ":isveç:",
 ":Turkey:": ":türkiye:",
 ":Ukraine:": ":ukrayna:",
 ":VS_button:": "versus",
 ":alarm_clock:": ":çalar_saat:",
 ":anger_symbol:": ":sinir_işareti:",
 ":angry_face:": ":sinirli_surat:",
 ":angry_face_with_horns:": ":sinirli_surat:",
 ":anxious_face_with_sweat:": ":tedirgin_surat:",
 ":astonished_face:": ":şaşırmış_surat:",
 ":baby_bottle:": ":biberon:",
 ":balloon:": ":balon:",
 ":bar_chart:": ":bar_grafik:",
":beaming_face_with_smiling_eyes:":
":mutlu_surat:",
 ":birthday_cake:": ":pasta:",
 ":blossom:": ":çiçek:",
 ":boar:": ":yaban_domuzu:",
 ":bomb:": ":bomba:",
 ":bottle_with_popping_cork:": ":şampanya:",
 ":bouquet:": ":buket:",
 ":broken_heart:": ":kırık_kalp:",
 ":butterfly:": ":kelebek:",
 ":call_me_hand:": ":ara_beni:",
 ":camera_with_flash:": ":fotoğraf_makinesi:",
 ":cat_with_tears_of_joy:": ":mutlu_surat:",
 ":check_mark:": ":onay_işareti:",
 ":check_mark_button:": ":onay_işareti:",
 ":cherry_blossom:": ":çiçek:",
 ":cigarette:": ":sigara:",
 ":clapping_hands:": ":alkış:",
 ":clinking_beer_mugs:": ":şerefe:",
 ":clinking_glasses:": ":şerefe:",
 ":clown_face:": ":palyaço:",
 ":cold_face:": ":hasta_surat:",
 ":collision:": ":çarpışma:",
 ":comet:": ":yıldız:",
 ":confetti_ball:": ":parti:",
 ":confused_face:": ":kafası_karışık_surat:",
 ":cowboy_hat_face:": ":kovboy_surat:",
 ":cross_mark:": ":çarpı_işareti:",
 ":cross_mark_button:": ":çarpı_işareti:",
 ":crossed_fingers:": ":bol_şans:",
 ":crossed_swords:": ":kılıçlar:",
 ":crown:": ":taç:",
 ":crying:": ":ağlayan_surat:",
 ":crying_cat:": ":ağlayan_surat:",
 ":crying_face:": ":ağlayan_surat:",
 ":crystal_ball:": ":kristal_küre:",
 ":cucumber:": ":salatalık:",
 ":disappointed_face:": ":hayal_kırıklığı_surat:",
 ":disguised_face:": ":takma_burun:",
 ":dizzy:": ":baş_dönmesi:",
 ":dolphin:": ":yunus_balığı:",
 ":double_exclamation_mark:": "!",
 ":dove:": ":kumru:",
 ":drooling_face:": ":hevesli_surat:",
 ":eagle:": ":kartal:",
 ":eight-pointed_star:": ":yıldız:",
 ":enraged_face:": ":sinirli_surat:",
 ":envelope:": ":zarf:",
 ":exclamation_question_mark:": "!?",
 ":exploding_head:": ":patlayan_kafa:",
 ":expressionless_face:": ":düz_surat:",
 ":eye:": ":göz:",
 ":eyes:": ":gözler:",
 ":face_blowing_a_kiss:": ":öpücük_surat:",
 ":face_savoring_food:": ":muzip_surat:",
 ":face_screaming_in_fear:": ":çığlık_surat:",
 ":face_vomiting:": ":kusan_surat:",
 ":face_with_crossed-out_eyes:":
":şaşıran_surat:",
 ":face_with_hand_over_mouth:":
":gülen_surat:",
 ":face_with_head-bandage:": ":yaralı_surat:",
 ":face_with_monocle:": ":elit_surat:",
 ":face_with_raised_eyebrow:":
":sorgulayan_surat:",
 ":face_with_rolling_eyes:":
":gözdeviren_surat:",
 ":face_with_steam_from_nose:":
":sinirli_surat:",
 ":face_with_symbols_on_mouth:":
":sinirli_surat:",
 ":face_with_tears_of_joy:": ":mutlu_surat:",
 ":face_with_thermometer:": ":hasta_surat:",
 ":face_with_tongue:": ":muzip_surat:",
 ":fairy:": ":peri:",
 ":female_sign:": ":kadın_işareti:",
 ":fire:": ":alev:",
 ":flamingo:": ":flamingo:",
 ":flexed_biceps:": ":güçlü:",
 ":flushed_face:": ":kızaran_yanak:",
 ":folded_hands:": ":şükran:",
 ":four_leaf_clover:": ":yonca:",
 ":front-facing_baby_chick:": ":civciv:",
 ":frown_sad_andry_pouting:": ":asık_surat:",
 ":frowning_face:": ":asık_surat:",
":frowning_face_with_open_mouth:":
":asık_surat:",
 ":full_moon_face:": ":mutlu_surat:",
 ":gem_stone:": ":elmas:",
 ":glowing_star:": ":yıldız:",
```

```
  ":goat:": ":keçi:",
  ":grimacing_face:": ":ekşi_surat:",
  ":grinning_face:": ":sırıtan_surat:",
  ":hand_with_fingers_splayed:": ":açık_el:",
  ":handshake:": ":el_sıkışmak:",
  ":happy_face_smiley:": ":mutlu_surat:",
  ":hear-no-evil_monkey:":
":duymayan_maymun:",
  ":heart:": ":kalp:",
  ":hibiscus:": ":çiçek:",
  ":high_voltage:": ":elektrik:",
  ":honey_pot:": ":bal:",
  ":honeybee:": ":arı:",
  ":hot_face:": ":sıcaklamış_surat:",
  ":hot_springs:": ":duman:",
  ":hundred_points:": ":yüz_puan:",
  ":kiss_mark:": ":öpücük:",
  ":kiss_woman_woman:":
":kadın_kadına_öpücük:",
  ":kissing_cat:": ":öpücük:",
  ":kissing_face_with_closed_eyes:": ":öpücük:",
  ":kissing_face_with_smiling_eyes:": ":öpücük:",
  ":kite:": ":uçurtma:",
  ":left-facing_fist:": ":yumruk:",
  ":lollipop:": ":lolipop:",
  ":loudly_crying_face:": ":ağlayan_surat:",
  ":loudspeaker:": ":megafon:",
  ":love-you_gesture:": ":sevgi_işareti:",
  ":love_letter:": ":aşk_mektubu:",
  ":magnifying_glass_tilted_right:": ":büyüteç:",
  ":man_dancing:": ":dans:",
  ":man_facepalming:": ":utanç_verici:",
  ":man_fairy:": ":peri:",
  ":mechanical_arm:": ":robot_kol:",
  ":middle_finger:": ":hareket_çekmek:",
  ":mouth:": ":öpücük:",
  ":movie_camera:": ":kamera:",
  ":musical_notes:": ":müzik:",
  ":nail_polish:": ":keyif:",
  ":nauseated_face:": ":mide_bulantısı:",
  ":nazar_amulet:": ":nazar_değmesin:",
  ":nerd_face:": ":muzip_surat:",
  ":neutral_face:": ":nötr_surat:",
  ":new_moon_face:": ":dolunay_surat:",
  ":newspaper:": ":gazete:",
  ":oncoming_fist:": ":yumruk:",
  ":open_book:": ":kitap:",
  ":palms_up_together:": ":dua:",
  ":party_popper:": ":parti:",
  ":partying_face:": ":parti:",
  ":peach:": ":şeftali:",
  ":person_facepalming:": ":utanç_verici",
  ":person_raising_hand:": ":yumruk:",
  ":person_shrugging:": ":omuz_silkmek:",
  ":person_with_veil:": ":gelin:",
  ":pile_of_poo:": ":bok:",
  ":ping_pong:": ":raket:",
  ":pleading_face:": ":yalvaran_surat:",
  ":prince:": ":prens:",
  ":princess:": ":prenses:",
  ":prohibited:": ":yasak:",
  ":puzzle_piece:": ":yapboz:",
  ":rainbow:": ":gökkuşağı:",
  ":rainbow_flag:": ":lgbt_bayrağı:",
  ":raised_fist:": ":yumruk:",
  ":raised_hand:": ":el:",
  ":raising_hands:": ":eller_yukarı:",
  ":red_apple:": ":elma:",
  ":red_exclamation_mark:": "!",
  ":red_question_mark:": "?",
  ":relieved_face:": ":rahatlamış:",
  ":revolving_hearts:": ":kalp:",
  ":right-facing_fist:": ":yumruk:",
  ":ring:": ":yüzük:",
  ":rocket:": ":roket:",
  ":rolling_on_the_floor_laughing:":
":gülen_surat:",
  ":rose:": ":çiçek:",
  ":rosette:": ":rozet:",
  ":round_pushpin:": "",
  ":sad_but_relieved_face:": "",
  ":see-no-evil_monkey:":
":gözü_kapalı_maymun:",
  ":shield:": ":kalkan:",
  ":shushing_face:": ":sus:",
  ":sign_of_the_horns:": ":metal:",
":skeptical_annoyed_undecided_uneasy_hesitant:
": ":şüpheci:",
  ":sleepy_face:": ":uykulu:",
  ":slightly_frowning_face:": ":asık_surat:",
  ":slightly_smiling_face:": ":gülen_surat:",
  ":smiling_face:": ":gülen_surat:",
  ":smirking_face:": ":sırıtan_surat:",
  ":sneezing_face:": ":hapşırık:",
  ":sparkles:": ":yıldız:",
  ":speak-no-evil_monkey:":
":ağzı_kapalı_maymun:",
  ":star-struck:": ":aşık:",
  ":star:": ":yıldız:",
  ":strawberry:": ":çilek:",
  ":studio_microphone:": ":mikrofon:",
  ":sun:": ":güneş:",
  ":tears_of_happiness:": ":mutlu:",
  ":television:": ":televizyon:",
  ":thinking_face:": ":düşünceli:",
  ":thumbs_down:": ":reddediyorum:",
  ":thumbs_up:": ":onaylıyorum:",
  ":tired_face:": ":yorgun_surat:",
  ":tongue:": ":dil_çıkarmak:",
  ":top_hat:": ":şapka:",
  ":tornado:": ":hortum:",
  ":transgender_flag:": ":trans_bayrağı:",
  ":trophy:": ":zafer_kupası:",
  ":tulip:": ":çiçek:",
  ":two_hearts:": ":kalp:",
  ":unamused_face:": ":gözdeviren_surat:",
  ":unicorn:": ":pegasus:",
  ":upside-down_face:": ":gülen_surat:",
  ":victory_hand:": ":zafer_işareti:",
  ":volleyball:": ":voleybol:",
```

":vulcan_salute:": ":selam:",
":warning:": ":uyarı:",
":waving_hand:": ":selam:",
":weary_face:": ":yorgun_surat:",
":wink_smirk:": ":göz_kırpan_surat:",
":winking_face:": ":göz_kırpan_surat:",
":winking_face_with_tongue:":
":göz_kırpan_surat:",
":wolf:": ":kurt:",
":woman_dancing:": ":dans:",
":woman_facepalming:": ":utanç_verici",
":woman_fairy:": ":peri:",
":woman_in_lotus_position:": "",
":woman_playing_handball:": ":voleybolcu:",
":woman_running:": ":koşucu:",
":woman_shrugging:": ":omuz_silkmek:",
":woman_tipping_hand:": ":ironi:",
":woman_walking:": ":yürüyüş:",
":women_with_bunny_ears:": "",
":woozy_face:": ":garip_surat:",
":writing_hand:": ":yazmak:",
":yawning_face:": ":sıkıcı:",
":zany_face:": ":garip_surat:",
":zipper-mouth_face:": ":sır_vermez:"

# REFERENCES


Alcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 31*(2), 211-236.

Arbatli, E., & Rosenberg, D. (2021). United we stand, divided we rule: how political polarization erodes democracy. *Democratization, 28*(2), 285-307.

Aydın-Düzgit, S., & Balta, E. (2018). When elites polarize over polarization: Framing the polarization debate in Turkey. *New Perspectives on Turkey, 60*, 1-24.

Baden, C., Kligler-Vilenchik, N., & Yarchi, M. (2020). Hybrid Content Analysis: Toward a Strategy for the Theory-driven, Computer-assisted Classification of Large Text Corpora. *Communication Methods and Measures, 14*(3), 165-183.

Baden, C., Yarchi, M., & Kligler-Vilenchik, N. (2021). Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication, 38*(1-2), 98-139.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B., . . . Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216-9221.

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132.

Barbera, P., Jost, J., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science, 28*(10), 1531–1542.

BBC News. (2021, March 20). *Domestic violence: Turkey pulls out of Istanbul convention*. Retrieved from BBC News: https://www.bbc.com/news/world-europe-56467689

Bilgiç, M. S., Koydemir, F. S., & Akyürek, S. (2014). Türkiye'de Kimlikler Arası Kutuplaşmanın Sosyal Mesafe Üzerinden Ölçümü ve Toplumsal Güvenliğe Etkisi. *Bilge Strateji, 6*(11), 163-205.

Bougher, L. (2017). The Correlates of Discord: Identity, Issue Alignment, and Political Hostility in Polarized America. *Political Behavior, 39*, 731-762.

Boxell, L., Gentzkow, M., & Shapiro, J. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences, 114*(40), 10612-10617.

Bozdağ, Ç. (2020). Managing Diverse Online Networks in the Context of Polarization: Understanding How We Grow Apart on and through Social Media. *Social Media + Society*, 6(4).

Bozdağ, Ç., & Koçer, S. (2022). Skeptical Inertia in the Face of Polarization: News Consumption and Misinformation in Turkey. *Media and Communication, 10*(2), 169-179.

British Psychological Society. (2017, - -). *Ethics Guidelines for internet-mediated research*. Retrieved from British Psychological Society: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf

Bulut, E., & Yörük, E. (2017). Digital Populism: Trolls and Political Polarization of Twitter in Turkey. *International Journal of Communication, 11*(25), 4093–4117.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R., . . . Kurzweil, R. (2018). Universal Sentence Encoder. *arXiv digital preprint*, arxiv.org/abs/1803.11175.

Chang, J., Boyd-Graeber, J., Gerrish, S., Wang, C., & Blei, D. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, (pp. 288-296). Vancouver, British Columbia.

Chen, W., & Ku, L. (2016). UTCNN: a deep learning model of stance classification on social media text. *arXiv digital preprint*, arXiv:1611.03599 [cs.CL].

Conover, M., Ratkiewicz, J., Francisco, M., ạlves, B. G., Flammini, A., & Menczer, F. (2011). Political Polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media, 5*(1), 89-96.

Council of Europe. (2011). *Convention on Prventing and Combating Violence Against Women and Domestic Violence.* -: -.

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2015). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation, 50*(3), 667–685 .

Del Vicario, M., Bessi, A., Zollo, F., Fabio, P., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). The spreading of misinformation online. *PNAS, 113*(3), 554-559.

Dey, K., Shrivastava, R., & Kaushik, S. (2018). Topical Stance Detection for Twitter: A Two-Phase LSTM Model Using Attention. *Advances in Information Retrieval, 5*(1), 529–536.

Du, J., Xu, R., He, Y., & Gui, L. (2017). Stance Classification with Target-specific Neural Attention. *Proceedings of the Twenty-Sixth International Joint Conference on*, (pp. 3988-3994).

Duggan, M., & Smith, A. (2016, October 25). *The Political Environment on Social Media*. Retrieved from Pew Research Center: https://www.pewresearch.org/internet/2016/10/25/politics-and-social-media-methodology/

Duggan, M., & Smith, A. (2016, October 25). *The tone of social media discussions around politics*. Retrieved from Pew Research Center: https://www.pewresearch.org/internet/2016/10/25/the-tone-of-social-media-discussions-around-politics/

Edenborg, E. (2021). 'Traditional values' and the narrative of gay rights as modernity: Sexual politics beyond polarization. *Sexualities, 26*(1-2), 37-53.

Erdoğan, E., & Uyan-Semerci, P. (2018). *Fanusta Diyaloglar: Türkiye'de Kutuplaşmanın Boyutları.* İstanbul: İstanbul Bilgi Üniversitesi Yayınları.

Erdoğan, E., & Uyan-Semerci, P. (2020). *Dimensions of Political Polarization in Turkey 2020.* Istanbul Bilgi University Center for Migration Research.

Erdogan-Ozturk, Y., & Isik-Guler, H. (2020). Discourses of exclusion on Twitter in the Turkish Context: #ülkemdesuriyeliistemiyorum (#idontwantsyriansinmycountry). *Discourse, Context & Media, 36*(100), 400.

Ertan, G., Çarkoğlu, A., & Aytaç, E. (2022). Cognitive political networks: A structural approach to measure political polarization in multiparty systems. *Social Networks, 68*, 118-126.

Flynn, D., Nyhan, B., & Reifer, J. (2017). The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics. *Political Psychology, 38*(51), 127-150.

Furman, I., & Tunç, A. (2019). The End of the Habermassian Ideal? Political Communication on Twitter During the 2017 Turkish Constitutional Referendum. *Policy & Internet, 12*(3), 311-331.

Garimella, K., Gionis, A., Morales, G. D., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. *Proceedings of the 2018 World Wide Web Conference*, (pp. 913-922).

Ghosh, S., Singhania, P., Singh, S., Rudra, K., & Ghosh, S. (2019). Stance Detection in Web and Social Media: A Comparative Study. *International Conference of the Cross-Language Evaluation Forum for European Languages*, 75-87.

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

Habermas, J., Lennox, S., & Lennox, F. (1974). The Public Sphere: An Encyclopedia Article. *New German Critique, 3*, 49-55.

Ho, S., & McLeod, D. (2008). Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication. *Communication Research, 35*(2), 190–207.

Iyengar, S., & Krupenkin, M. (2018). The Strengthening of Partisan Affect. *Advances in Political Psychology, 39*, 201-218.

Iyengar, S., Lelkes, Y., Levendusky, M., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science, 22*(1), 129-146.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly, 76*(3), 405-431.

Jahani, E., Gallagher, N., Merhout, F., Cavalli, N., Guilbeault, D., Leng, Y., & Bail, C. A. (2022). An Online experiment during the 2020 US–Iran crisis shows that exposure to common enemies can increase political polarization. *Scientific Reports, 12*, 19304.

Johnson, D. R., Jasper, D. M., Griffin, S., & Huffman, B. L. (2013). Reading Narrative Fiction Reduces Arab-Muslim Prejudice and Offers a Safe Haven From Intergroup Anxiety. *Social Cognition, 31*(5), 578-598.

KADEM. (2020, 08 01). *İstanbul Sözleşmesi Hakkında*. Retrieved from KADEM: https://kadem.org.tr/istanbul-sozlesmesi-hakkinda/

Karaosmanoğlu, D. (2020). From Ayran to Dragon Fruit Smoothie: Populism, Polarization, and Social Engineering in Turkey. *International Journal of Communication, 14*, 1253-1274.

Kaya, Y., & Sunar, L. (2015). The Culture Wars Redux? The Polarization of Social and Political Attitudes in Turkey. *Social Currents, 2*(4), 393-412.

Kiriş, H. (2011). Parti Sisteminde Kutuplaşma ve Türk Parti Sistemi Örneği. *Amme İdaresi Dergisi, 44*(4), 33-67.

Köksal, A., & Özgür, A. (2021). Twitter Dataset and Evaluation of Transformers for Turkish Sentiment Analysis. *2021 29th Signal Processing and Communications Applications Conference (SIU).*

Korolczuk, E., & Graff, A. (2018). Gender as "Ebola from Brussels": The Anticolonial Frame and the Rise of Illiberal Populism. *Journal of Women in Culture and Society, 43*(4), 797-821.

Kováts, E., & Pető, A. (2017). Anti-Gender Discourse in Hungary: A Discourse without a Movement? In R. Kuhar, & D. Paternotte, *Anti-Gender Campaigns in Europe: Mobilizing against Equality* (pp. 117-131). London: Rowman & Littlefield.

Küçük, D., & Can, F. (2018). Stance Detection on Tweets: An SVM-based Approach. *arXiv digital preprint*, arXiv:1803.08910 [cs.CL].

Kutlu, M., Darwish, K., & Elsayed, T. (2018). Devam vs. Tamam: 2018 Turkish Elections. *arXiv digital preprint*, arXiv:1807.06655 [cs.SI].

Kutter, A., & Kantner, C. (2012). Corpus-based Content Analysis: A Method for Investigating News Coverage on War and Intervention. *International Relations Online Working Paper Series*.

Laebens, M., & Öztürk, A. (2021). Partisanship and Autocratization: Polarization, Power Asymmetry, and Partisan Social Identities in Turkey. *Comparative Political Studies, 54*(2), 245-279.

Levendusky, M. (2018). Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization? *The Journal of Politics, 80*(1), 59-70.

Mason, L. (2018). Ideologues Without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly, 82*(S1), 866-887.

McConnell, C., Margalit, Y., Malhotra, N., & Levendusky, M. (2017). The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science, 62*(1), 5-18.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the Global Crisis of Democracy: common Patterns, Dyanmics and Pernicious Consequeunces for Democratic Polities. *American Behavioral Scientist, 62*(1), 16-42.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist, 62*(1), 16–42.

Messing, S., & Westwood, S. (2012). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research, 41*(8), 1042–1063.

Mete-Dökücü, H., & Just, A. (2021). Party system polarization in developing democracies: the case of Turkey, 1950–2018. *Turkish Studies, 23*(3), 331-353.

Ming, F., Wong, F., Tan, C. W., Sen, S., & Chiang, M. (2016). Quantifying Political Leaning from Tweets, Retweets, and Retweeters. *IEEE Transactions on Knowledge and Data Engineering, 28*(8), 2158–2172.

Newman, N., Fletcher, R., Kalogeropoulos, A. L., & Nielsen, R. (2017). *Reuters Institute Digital News Report 2017.* Oxford: Reuters Institute for the Study of Journalism.

Nguyen, D. Q., Vu, T., & Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English Tweets. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 9-14). Online: Association for Computational Linguistics.

Ozduzen, O., & Korkut, U. (2020). Enmeshing the mundane and the political: Twitter, LGBTI+ outing and macro-political polarisation in Turkey. *Contemporary Politics, 26*(5), 493-511.

Ozduzen, O., Korkut, U., & Ozduzen, C. (2021). 'Refugees are not welcome': Digital racism, online place making and the evolving categorization of Syrians in Turkey. *New Media & Society, 23*(11), 3349-3369.

Parker, D., & Dull, M. (2011). Divided We Quarrel: The Politics of Congressional Investigations, 1947–2004. *Legislative Studies Quarterly, 34*(3), 319-345.

Persily, N. (2017). Can Democracy Survive the Internet? *Journal of Democracy, 28*(2), 63-76.

Pew Research Center. (2018). *News Use Across Social Media Platforms 2018.* -: Pew Research Center.

Pew Research Center. (2018). *Partisans Remain Sharply Divided in Their Attitudes About the News Media.* -: Pew Research Center.

Pew Research Center. (2019, June 27). *Republicans far more likely than Democrats to say fact-checkers tend to favor one side*. Retrieved from Pew Research Center: https://www.pewresearch.org/fact-tank/2019/06/27/republicans-far-more-likely-than-democrats-to-say-fact-checkers-tend-to-favor-one-side/

Pigott, F., Kolb, J., Montague, J., Gonzales, A., & Moffitt, J. (2021). *searchtweets-v2.* Retrieved from pypi.org: https://pypi.org/project/searchtweets-v2/

Qiang, J. a. (2019). Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *arXiv digital preprint* , arXiv:1904.07695.

Rajendran, L., & Thesinghraja, P. (2014). The Impact of New Media on Traditional Media. *Middle-East Journal of Scientific Research, 22*(4), 609-616.

Rashed, A., Kutlu, M., Darwish, K., Elsayed, T., & Bayrak, C. (2021). Embeddings-Based Clustering for Target Specific Stances: The Case of a Polarized Turkey. *Proceedings of the International AAAI Conference on Web and Social Media*, 537-548.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software, 91*(2), 1-40.

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior, 38*, 485-508.

Santana, A. (2014). Virtuous or Vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice, 8*(1), 18-33.

Schaffner, B., Macwilliams, M., & Nteta, T. (2018). Understanding White Polarization in the 2016 Vote for President: The Sobering Role of Racism and Sexism. *Political Science Quarterly, 133*(1), 9-34.

Schweter, S. (2020, April). *BERTurk - BERT models for Turkish.* Retrieved from https://doi.org/10.5281/zenodo.3770924

Skalski, P. D., Neuendorf, K. A., & Cajigas, J. A. (2017). Content Analysis in the Interactive Media Age. In P. D. Skalski, K. A. Neuendorf, & J. A. Cajigas, *The Content Analysis Guidebook* (pp. 201-242). London: SAGE.

Stefanov, P., Darwish, K., Atanasov, A., & Nakov, P. (2019). Predicting the Topical Stance of Media and Popular Twitter Users. *arXiv digital preprint*, arXiv:1907.01260 [cs.SI].

Suhat, E., Bello-Pardo, E., & Maurer, B. (2017). The Polarizing Effects of Online Partisan Criticism: Evidence from Two Experiments. *The International Journal of Press/Politics, 23*(1), 95–115.

Tucker, J., Theocharis, Y., Roberts, M., & Barberá, P. (2017). From Liberation to Turmoil: Social Media And Democracy. *Journal of Democracy, 28*(4), 48-59.

Ünver, H. A. (2019, March). *Russian Digital Media andInformation Ecosystem in Turkey.* Retrieved from EDAM: https://edam.org.tr/wp-content/uploads/2019/04/Ak%C4%B1n-%C3%9Cnver-Russian-Digital-Media.pdf

Urman, A. (2020). Context matters: political polarization on Twitter from a comparative perspective. *Media, Culture & Society, 42*(6), 857-879.

we are social & Hootsuite. (2019). *Digital in 2019: Turkey.* : We are social ; Hootsuite.

Webster, S., & Abramowitz, A. (2017). The Ideological Foundations of Affective Polarization in the U.S. Electorate. *American Politics Research, 45*(4), 621–647.

Weeks, B. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on susceptibility to Political Misinformation. *Journal of Communication, 65*(4), 699-719.

Westwood, S., Iyengar, S., Walgrave, S., Leonisio, R., Miller, L., & Strijbis, O. (2018). The tie that divides: Cross-national evidence of the primacy of partyism. *European Journal of Political Research, 57*(2), 333-354.

Wong, J. C. (2017, 12 12). *Former Facebook executive: social media is ripping society apart*. Retrieved from The Guardian: https://www.theguardian.com/technology/2017/dec/11/facebook-former-executive-ripping-society-apart

Zarrella, G., & Marsh, A. (2016). MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, (pp. 458–463).