RICE UNIVERSITY

Social Network Censorship: Topics, Techniques, and Impacts

by

Rima S. Tanash

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Dan Wallach, Adviser
Professor of Computer Science

Devika Subramanian
Professor of Computer Science

Melissa Marshall
Professor of Political Science

Chris Bronk
Professor of Computer and Information Systems

Houston, Texas

May, 2017
ABSTRACT

Social Network Censorship: Topics, Techniques, and Impacts

by

Rima S. Tanash

Previous research in digital censorship focused by and large on studying censorship of applications and networks that are heavily controlled by oppressive governments such as China. My research goal is to broaden these studies beyond censorship in autocracies and include western social media, with an explicit focus on Turkish censorship of Twitter, while showing evidence of how aggregating users’ data from public APIs can lead to privacy leaks of users’ political affiliations.

In this research, we made fundamental contributions in the areas of censorship and privacy. We conducted large scale measurements of Twitter in Turkey, and introduced an approach to systematically label censored Twitter posts while showing that it is possible to find more censored tweets than those published by Twitter. We show a simple way to bypass Twitter censorship inside of the censoring country without using a proxy/VPN, which can be used by ordinary users. Our framework can be applied to any geographical boundaries or targeted users’ groups. We propose a novel set of rules to construct a data flow graph of censored tweets that unmasks influential users and their community association that results in revealing their political and social background. This aggregate
analysis of publicly reachable tweets is particularly critical to users’ privacy if unmasked by oppressive governments or malicious persons who can thereafter leverage these users to spread malicious content. Using standard machine learning and NLP algorithms for topic clustering, we show that the dynamics of censorship in democratic countries, such as Turkey, are different than those in dictatorial regimes that are targeting collective action discussions. Our results show that the censors in Turkey target topics that could negatively impact the outcome of an election and the ruling Justice and Development Party (AKP) political interest. Expanding this work to examine the impact of the failed Turkish coup 2016 on censorship, we identified a new censored topic that is also deemed adversarial to AKP. The overwhelming majority of the censored tweets pre-coup, from 2014 to 2015, are on government corruption and Kurdish/terrorism issues, with some pro-government censored tweets, because current Turkish law enables individuals to pursue due process against defamatory posts. On the other hand, post-coup results show that in additional to the previously identified censored Kurdish topics, Gülen-movement related topics are also heavily censored, a movement the government claims to be the mastermind behind the coup. More notably, this empirical comparison study reveals evidence of 72% decline in publicly identifiable government-censored tweets. We attribute this, in part, to an estimated 43% decline in overall Twitter usage in Turkey and in part to users’ self-censorship. Supporting this theory, we detected that 41% of all users in our pre-coup dataset voluntarily removed 18% of their old tweets by either switching their accounts to protected mode, deleting their accounts, or deleting some tweets. Unlike activists who regularly tweet political content, and are more likely to be censored by the government, we found that self-censoring users appear to be more typical users who normally post neutral tweets, and only 6% political tweets on average.
Acknowledgments

This dissertation work is a collection of several years of research in Internet censorship at the Rice University, Department of Computer Science, under the supervision of my adviser, Dr. Dan Wallach. I would like to thank Dr. Wallach for his continuous support and guidance through the past years, and for his constructive feedback. I am grateful for the opportunity he has given me to work with him and his first-class lab students, whom I also want to acknowledge for their never-ending encouragement, support, and friendship.

I want to thank my long-term mentor and friend, Dr. Sam Small, for encouraging me to apply and enroll in the PhD program at Rice, and for always believing in me when I needed it the most.

I would like to especially thank my committee members; Dr. Devika Subramanian for sharing her knowledge in machine learning which incentivized me to widen my research from various perspectives, and for her useful discussions throughout the past years, Dr. Melissa Marschall for her continuous support, and for the collaboration opportunity you made possible, through which I was introduced to new ways to reason about politically challenging research questions, Dr. Chris Bronk, for all his support and feedback, and for making sure my work always received highest exposure. I thank all of you for that, and for the opportunity to work with you.

I also want to acknowledge my collaborators Zhouhan Chen, Abdullah Aydogan, and Tanmay Thakur for their helpful discussions and their contribution toward this research, and 3 other “anonymous” persons for their insight and their translation efforts.

More importantly, I want to thank my parents Said and Viki, for always showing faith
in me and giving me liberty to choose what I desired, and for their unconditional love and support, especially my mom who always helped me pick-up the slack at times of high pressure.

There is one particular individual that I owe a very big thank-you to; my husband, Amr. He is the one who has empowered me to pursue my goals, and encouraged me to reach my highest potential. When at times, the thought of continuing was overwhelming, his outlook, his words of wisdom, and his continued love and support through this journey made it easier to stick to it and stay focused on the end result. I am grateful to my amazing children, Sarah and Adam, for being proud of me and fueling me with love and joy, and the patience they showed during my thesis work and writing, and for giving me the ultimate reason to always work hard. Words could never express how grateful I am to have such a loving and caring family.
## Contents

- Abstract ii
- Acknowledgments iv
- List of Illustrations x
- List of Tables xii

1 About this Dissertation 1
  1.1 Introduction .......................................................... 1
  1.2 Dissertation Overview .................................................. 3

2 Measuring Twitter Censorship in Turkey 7
  2.1 Introduction .......................................................... 7
  2.2 Related Censorship Work ............................................... 9
  2.3 Country-Withheld Content Mechanisms .............................. 10
    2.3.1 Transparency Reporting ........................................... 11
    2.3.2 The “Country_Withheld” Process ................................. 14
    2.3.3 How Twitter Appears from Different Countries ............... 15
    2.3.4 Twitter APIs ......................................................... 16
    2.3.5 Withheld Accounts ................................................ 18
  2.4 Case Study: Turkey ................................................... 18
    2.4.1 Disclosed Censorship ............................................. 20
    2.4.2 Collecting Censored Tweets ....................................... 20
2.4.3 Collecting Censored Accounts .................................................. 22
2.5 Broader Turkish Data Collection .................................................. 22
2.5.1 Data Sources ....................................................................... 23
2.5.2 Data Collection .................................................................... 24
2.5.3 Analysis ............................................................................... 26
2.6 Censorship Topics with tf-idf and NMF ....................................... 31
2.6.1 Hot Topic Clustering ............................................................... 32
2.6.2 Hot Topics from Withheld Accounts ....................................... 34
2.7 Bypassing Censorship ................................................................. 35
2.7.1 Censorship Escalation ............................................................. 37
2.7.2 How Many Users Use Tor in Turkey ......................................... 38
2.8 Summary ................................................................................. 40

3 Modeling Influential Users and Users’ Communities ................. 42
3.1 Influential Users ........................................................................ 42
3.2 Related Work ............................................................................ 42
3.3 Tweet’s JSON Fields .................................................................. 44
3.3.1 Constructing Users’ Influence Data-Flow Graph ..................... 46
3.3.2 Community detection ............................................................. 47

4 The Dynamics of Social Media Censorship in Transitioning Democracies ............................................................................ 49
4.1 Introduction ................................................................................ 49
4.2 Explaining the Impetus and Dynamics of Censorship Beyond Dictatorial Regimes ......................................................... 52
4.3 Censorship Mechanism in Turkey ................................................. 54
5 The Decline of Social Media Censorship and the Rise of Self-Censorship after the 2016 Failed Turkish Coup

5.1 Introduction

5.2 Related work: Censorship in Turkey

5.3 Data

5.4 Hypotheses

5.4.1 Censorship Size

5.4.2 Censored Topics

5.4.3 Self-Censorship of Gülen Topics

5.4.4 Self-Censoring Users

5.5 Twitter Transparency Report

5.6 Summary
Illustrations

2.1 Sample withholding request from Russia, retrieved from Chilling Effects in 2013 ........................................... 13
2.2 Sample of withholding notice from Turkey - Chilling Effects extracted in 2015 .................................................. 14
2.3 Withheld tweet as it appeared from inside the censoring country ................................................................. 15
2.4 Withheld account as it appeared from inside the censoring country ............................................................ 15
2.5 Comparison view of a withheld Russian tweet from Russia (above) and USA (below) ...................................... 16
2.6 Distribution of number of tweets per user ........................................................................................................ 28
2.7 Distribution of number of tweets per user (log-log scale) ............................................................................. 28
2.8 Same user with two profiles withheld in Turkey .......................................................................................... 29
2.9 Tor Metrics: users connecting from Turkey .................................................................................................... 39
2.10 Distribution of withheld tweets by month (log scale) .................................................................................. 39

4.1 Daily comparison of withheld vs. non-withheld tweets .................................................................................. 63
4.2 Top Users’ Communities of Censored Tweets, and Influential Users ......................................................... 72
4.3 Porn community .............................................................................................................................................. 76
4.4 Users without edges (singleton) .................................................................................................................. 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Volume of streamed Tweets post-coup 2015 vs. pre-coup 2016, 24 days bins.</td>
<td>84</td>
</tr>
<tr>
<td>5.2 Daily Tor connection in Turkey - July 2016.</td>
<td>85</td>
</tr>
<tr>
<td>5.3 Daily Tor connection in Turkey - June 2015.</td>
<td>86</td>
</tr>
<tr>
<td>5.4 Users communities and influential users of censored tweets post-coup</td>
<td>90</td>
</tr>
<tr>
<td>5.5 Labeling of Gülen related Tweets from public vs. unreachable tweets</td>
<td>93</td>
</tr>
<tr>
<td>5.6 Percent of political tweets of the top 200 users, from the highest percent to the lowest</td>
<td>95</td>
</tr>
<tr>
<td>A.1 Daily comparison of withheld vs. non-withheld tweets (Oct 21 to Dec 5)</td>
<td>106</td>
</tr>
<tr>
<td>A.2 Daily comparison of withheld vs. non-withheld tweets (Dec 9 to Jan 10)</td>
<td>107</td>
</tr>
</tbody>
</table>
Tables

2.1 Aggregate Transparency Reports disclosed by Twitter for 18 months (Jan. 2012 - Jun. 2013) from all countries. .......................................................... 12
2.2 Distribution of withheld tweets in Turkey, as reported by Twitter. ............. 21
2.3 Hot topics from withheld tweets from non-withheld accounts ..................... 33
2.4 Topics discussed in withheld accounts. .................................................... 34

3.1 Data flow rules for adding directed edges ................................................. 47

4.1 Withheld vs. non-withheld Tweets on street protests. .............................. 62
4.2 Most salient topics from non-withheld tweets in the first dataset ................. 65
4.3 Most salient topics from withheld tweets in the first dataset ..................... 66
4.4 Most salient topics from non-withheld tweets in the second dataset ........... 68
4.5 Most salient topics from withheld tweets in the second dataset ................. 69

5.1 Post-coup user-accounts labeling results ................................................. 88
5.2 Gülen-related vs. all tweets, collected post-coup, counting how many remain public vs. unreachable. ......................................................... 92

A.1 Protests and the Associated Keywords ............................................... 108
A.2 Protests from October 15, 2014 to January 15, 2015 ............................... 109
A.3  Protests from June 11, 2015 to June 23, 2015 ..................................... 110
A.4  Examples of transforming tweets in preprocessing phase .................. 112
Chapter 1

About this Dissertation

1.1 Introduction

Censorship is one form of regulation that is of particular interest since it has such profound implications for democratic institutions and processes, social movements, and democratization. Studies of social media censorship overwhelmingly focus on dictatorships like China and Syria. Scholars across a number of disciplines have been fascinated with understanding not simply how people use this medium, but also how governments regulate users, content, and usage (Zhu et al. [1]; Gunitsky [2]; Rød and Weidmann [3]; Fu et al. [4]; Greitens [5]; Gohdes [6]; Lorentzen [7]; Bamman [8]; Cairns and Carlson [9]; Howard et al. [10]). However, global censorship reports demonstrate that transitioning democracies such as Turkey and Brazil also censor social media. In 2012, Twitter, widely used around the world, introduced a standard interface for government agencies to request that individual tweets or even whole accounts be censored [11], which in this work, we found that this mechanism is relatively easy to circumvent. Twitter, in turn, discloses specific incidents of country-by-country censorship to the lumen database[4], previously known as the Chilling Effects, and statistics about this censorship in its transparency reports [12].

[https://lumendatabase.org/]
on bi-annual bases, which we found to be under-reported in the case of Turkish censorship by one-order-of-magnitude, raising the possibility that similar trends hold for censored tweets from other countries as well.

In this work, we propose a framework for systematically collecting and detecting censored tweets, and extracting influential users and their communities from the censored data, as well as community-based topics using NLP standard algorithms and graph metrics. The totality of this work considered studying over 34 million Turkish tweets from late-2014 to late-2016 spanning the Turkish election corruption scandal in 2014 and the Turkish general election and the Kurdish bombing event in 2016, as well as tweets collected during the 2016 failed Turkish coup.

Despite the prevalence of social media censorship in transitioning democracies [13], we know very little about the nature or dynamics of censorship in these states. The reason is that most existing studies have focused almost exclusively on authoritarian regimes. The work in this dissertation aims to study the dynamics of social media censorship beyond autocracies where censors primarily target posts referencing protests, with an explicit focus on Twitter censorship in Turkey, as Twitter identifies Turkey by far the leader in Twitter censorship in recent years. Focusing on Turkey and by building on theories of issue responsiveness and electoral studies, we argue that the issues prioritized by the public primarily shape the censorship strategies of transitioning democracies. We applied NLP and standard machine learning/clustering techniques, and found that the vast quantity of censored tweets contained political content, often critical of the Turkish government primarily targeting posts referencing corruption and terrorism/Kurdish issues, prior to the 2016 failed Turkish coup, both of which were ranked as citizens’ top priorities during the
studied period. Our analysis shows that a perceived threat of collective action does not explain censorship in Turkey. Instead, we find support for our hypothesis that social media censorship in transitioning democracies is targeted at sensitive public policy issue topics, those perceived by the public as important, but on which the ruling party was unwilling or unable to address.

Although self-censorship of the press is not a novel practice following past military coups in Turkey [14], this is the first work to examine and quantify social media self-censorship, and empirically analyze the coup’s impact on government-censored topics, and the volume of government censorship, by comparing posts from after the coup to posts we collected before the coup during the Turkish general election of 2015, where we show a drastic decline in publicly identifiable government-censored tweets following the coup, attributed, in part, to the decline in overall Twitter usage in Turkey and in part to users’ self-censorship. We identified a new focus of the Turkish government censorship on the Gülen movement after the coup. Our analysis show pro-Gülen tweets being widely self-censored. Additionally, we detected more publicly accessible anti-Gülen tweets. We also studied self-censored users’ profiles, and found that unlike government-censored users, self-censored users are politically neutral.

In the following section, we explain how we organize this dissertation work and its related publications and conferences.

1.2 Dissertation Overview

This dissertation is organized as follows.
In Chapter 2, we present a framework for detecting and collecting public censored tweets using different Twitter public APIs. Censored tweets detected using this framework are mainly tweets ordered to be removed or redacted by governments’ officials or court orders. Using this framework, we conduct a measurement study in Turkey and present findings of the size of censorship in Turkey and the topics being censored during the collection period. We also present a novel way to bypass Twitter censorship by introducing a simple change in the application setting. This work was presented in 2015 at the ACM-CCS-Workshop on Privacy in the Electronic Society (WPES). The suggested citation for this work is as follows:


In Chapter 3, we present methods for identifying influential users as a function of nodes’ out-degree from a set of censored tweets using graph metrics. We first discuss how we generate the data-flow graph using a specific set of rules to represent a snapshot of the users’ communication in our dataset. We next show how we use modularity, which help us achieve better precision in topic extraction using machine learning document classification algorithms tf-idf and NMF, which we discuss in more details in Chapter 4. This work was presented in the 2016 POLMETH conference. The suggested citation for this work is:
In Chapter 4, we study the dynamics of social media censorship beyond autocracies. We build on the line of literatures by moving the focus of research beyond dictatorial regimes to “semi” or transitioning democracies with specific focus on Turkey, and analyze what is censored and who is censored. The work in this chapter is collaboration with Melissa Marschall from the Political Science Department at Rice University and Abdullah Aydogan from the Baker Institute of Public Policy. This work was presented at the American Political Science Association (APSA) 2016 conference. A complete work is scheduled to be submitted for publication consideration in 2017 at the American Political Science Association (APSA), peer-reviewed political science journals.

In Chapter 5, we examine the effect of the 2016 Turkish coup on social media censorship, both by the government ordering Twitter to conduct censorship and as well as by people removing their own tweets. We compare tweets collected from Turkey pre-coup to tweets collected post-coup, and examine and quantify social media self-censorship, and empirically compare its effect relative to government-implemented censorship of social media. The work in this chapter was submitted for publication consideration to the 7th USENIX Workshop on Free and Open Communications on the Internet (FOCI), with the
title “The Decline of Social Media Censorship and the Rise of Self-Censorship after the 2016 Failed Turkish Coup”. Authors who assisted in this work, including the review, are; Zhouhan Chen, Dr. Dan Wallach, and Dr. Melissa Marschall.

In Chapter 6 we examine the history of the Turkish democracy and freedom of speech in light of press censorship, and particularly correlate these historical events relevant to the censored data we identified throughout our research.

Finally, we present the various directions for future work related to these studies, and our conclusion in Chapter 7.
Chapter 2

Measuring Twitter Censorship in Turkey

2.1 Introduction

Twitter, the popular microblogging platform, plays an essential role for communicating news and current events among professional reporters, human-rights activists, and oppressed citizens across the globe. More notably, Twitter was widely used to disseminate news and opinion during a series of uprisings in the Middle East, also known as the Arab Spring, resulting in the overthrow of dictatorships in Egypt, Tunisia, Libya [15][16], and the Taksim Gezi Park protests in Turkey on May 2013 [17].

On January 27, 2012, Twitter announced a new censorship policy known as “Country-Withheld Content” by which Twitter enabled governments and their representatives to formally request that Twitter withhold tweets and/or whole accounts within the boundaries of a specific country [11]. Once Twitter receives such a request, it somehow determines whether the request is lawful and then withholds the tweet in question in that specific country, while still allowing it to be visible elsewhere. Twitter claimed that this policy was a business decision that allowed Twitter to exist in parts of the world that have different ideas of freedom of expression, and that this policy will prevent locally offensive content [18]. The policy received criticism that it was nothing more than a form of government censorship and a threat to freedom of speech [19]. Unsurprisingly, trending hashtags like
#TwitterCensored and #twitterblackout followed this announcement, where many users expressed their outrage. Notably, Twitter also announced a partnership with Chilling Effects to publish withheld content “unless they are legally prohibited from doing so” [11].

Twitter also notes, in their “Withholding Transparency Reports” that the reported data is neither 100% comprehensive nor complete. For example, the Politwoops website, dedicated to collect deleted tweets from politicians—often a source of raw and embarrassing statements—announced that Twitter disabled their API feed on June 4, 2015. Twitter claimed that the company violated their developer agreement [20]. Politicians deleting their own tweets is obviously not the same thing as a country censoring its citizens, but all the same, Twitter has demonstrated that it is opposed to external organizations displaying content that is not visible on Twitter itself.

Given this, the research question is easy to pose: how much censorship is really happening on Twitter? How much is not being reported to Chilling Effects? How much is not being reported on Twitter’s withholding transparency disclosures? And can we determine anything about the machinery behind the censorship? Are tweets being withheld one by one, based on individual requests to Twitter by foreign governments, or are they being withheld in large groups, perhaps based on hashtags or other keywords? We suspect that there is undisclosed censorship on Twitter. We just do not know the depth of the unknowns.
2.2 Related Censorship Work

Chen et al. described that in recent years, social media has risen in prominence in many countries. In China, social media such as Weibo and Renren plays an important role as a platform for breaking news and political commentary outside of the confines of state-controlled news media. However, like all websites in China, Chinese social media is subject to censorship. The magnitude of censorship varies dramatically across topics, with 82% of posts in some topics being censored. The paper also finds that censorship of a topic correlates with high user engagement, suggesting that censorship does not stifle discussion of sensitive topics. Furthermore, the authors find that users create variants of words (known as morphs) to avoid keyword censorship [21].

Florio et al. confirmed that in 2014 the Turkish government hijacked DNS traffic to censor users traffic. Users of traditional computers were able to circumvent censorship using TOR and VPNs. Unlike traditional users, mobile users used a specialized Android application designed to allow DNS configurations of mobile devices to circumvent G/4G service-provider censorship. The application was installed by 130,000 devices by August 2014. The researchers collected data that showed that some censorship activities began after the lift of the Twitter ban on Turkey in March of 2014. However, they did not examine the type of data being censored [22].

Aase et al. argued that motivation, resources, and time are three major elements in the application and potential uses of Internet censorship. The challenge for the online censorship research community is to develop tools for measuring these three elements explicitly when conducting measurement studies [23].
Winter et al. proposed techniques to measure and circumvent Internet censorship that they deployed in three countries. The techniques successfully bypassed the Great Firewall of China [24].

Morrison examined the feasibility of automating the detection of censorship in microblogs without using sensitive keywords but using social network graphs properties and communication flow, and found his automated detection methods feasible when studying Sina Weibo [25].

In recent years, a large number of papers have focused on censorship resistance schemes (CRSs). Khattak et al. proposed an attack model to comprehensively explore censorship capabilities and and developed an evaluation framework to test each CRS’s flexibility [26].

No previous work has been attempted to quantify Twitter’s internal censorship mechanisms.

2.3 Country-Withheld Content Mechanisms

To better understand the steps Twitter uses to implement its “Country-Withheld Content” policy, it is important to examine both its Transparency Reports and the Chilling Effects website that Twitter uses to publish government withholding notices. The Chilling Effects is an independent third party archiving service that publishes cease and desist requests and other related legal demands [27]. Many companies such as Google and Facebook use Chilling Effects as a method for being more transparent about how they handle censorship-related requests (and, perhaps, as a way of disincentivizing those who might wish to issue them various forms of legal demands).
2.3.1 Transparency Reporting

When Twitter announced its censorship program, they noted: “...we have expanded our partnership with Chilling Effects to publish not only DMCA notifications but also requests to withhold content—unless, similar to our practice of notifying users, we are legally prohibited from doing so.” Twitter started publishing Transparency Reports in January 2012, with data grouped into 6-month bins on a per-country basis. We summarize 18 months of this data, from January 2012 to June 2013 in Table 2.1. The bulk of the censorship appears to occur in three countries: Brazil, France, and Russia. A request may specify several user accounts and/or tweets to be withheld, so the number of actual tweets withheld does not necessarily map one-to-one with the number of requests or the number of accounts specified. For example, in the case of Brazil, Twitter withheld 39 tweets and one account, but the report does not specify if the tweets were generated by the withheld account.

From our data collection, it appears that a request to withhold an account causes all tweets from that account to be withheld. In such cases, Twitter appears not to include these tweets from withheld accounts in the overall count of withheld tweets for a given country in its transparency reporting. We verified this manually for Brazil and for Germany, but have not systematically looked at each and every country with withheld accounts.

Our initial effort was to collect as many Twitter-related reports from the Chilling Effects database as we could find. This would enable us, for example, to learn Twitter handles, hashtags, and other sensitive keywords around which we might later build automated searches. Upon initial examination, we quickly discovered how incomplete the Chilling
<table>
<thead>
<tr>
<th>Country</th>
<th>Number of Requests</th>
<th>Number of Accounts Specified</th>
<th>Number of Accounts Withheld</th>
<th>Number of Tweets Withheld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Brazil</td>
<td>26</td>
<td>33</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>Canada</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ecuador</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>France</td>
<td>5</td>
<td>56</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>Germany</td>
<td>6</td>
<td>11</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Greece</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>India</td>
<td>4</td>
<td>18</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Japan</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Korea</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Russia</td>
<td>17</td>
<td>17</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Turkey</td>
<td>14</td>
<td>46</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UK</td>
<td>9</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>US</td>
<td>6</td>
<td>23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>108</strong></td>
<td><strong>264</strong></td>
<td><strong>9</strong></td>
<td><strong>117</strong></td>
</tr>
</tbody>
</table>

Table 2.1: Aggregate Transparency Reports disclosed by Twitter for 18 months (Jan. 2012 - Jun. 2013) from all countries.
Effects database appears to be. We found a grand total of 33 notices posted to Chilling Effects, across all countries, which is far fewer than just the 108 account withholding requests disclosed by Twitter’s own transparency reports. Clearly, the Chilling Effects database is nowhere near a comprehensive disclosure of Twitter’s stream of withholding requests.

Despite these shortcomings, we found that the Chilling Effects postings disclosed a fair bit of information, including the stated reason for the request as well as the user and identifier of the tweet being censored. We show a sample censorship request in Figure 2.1.

Figure 2.1: Sample withholding request from Russia, retrieved from Chilling Effects in 2013.

In 2014, Twitter changed their procedures and began hiding some information about the government agencies requesting the censorship (see Figure 2.2 for a sample Turkish notice). To build our database of Chilling Effects notices, we converted PDF to RTF and manually extracted tweets, user-names, and other fields. This was sufficiently robust,
across multiple languages, that it served our needs.

Figure 2.2: Sample of withholding notice from Turkey - Chilling Effects extracted in 2015.

2.3.2 The “Country_Withheld” Process

As described before, Twitter allows several mechanisms for countries to request that tweets be censored, including email, a web form, and hardcopy notices [18]. If Twitter agrees with the request, the withheld tweet will then be greyed-out with a message: “This tweet from @username has been withheld in Country Name” (see Figure 2.3). The user may or may not receive a notice that their tweet was being censored, but the censored tweet is generally visible from outside of the censoring country. Twitter offers a similar mechanism to censor an entire user, rather than just an individual tweet (see Figure 2.4).
2.3.3 How Twitter Appears from Different Countries

To view how Twitter looks from different countries, we used Tor as a proxy service, allowing us to choose the country of origin for our traffic by selecting different Tor exit nodes. As an example, Figure 2.5 shows how a Russian-withheld tweet appears respectively from Russia and the US.
2.3.4 Twitter APIs

Twitter offers a large collection of public REST and Streaming APIs\(^*\). All API responses return JSON structures; we present an abbreviated response below:

*Twitter API https://dev.twitter.com/docs/api/1.1
When a tweet is withheld, the corresponding REST API JSON response will include a field called “withheld_in_countries” with a list of two-letter ISO country codes where the tweet is to be censored; otherwise the field is absent. We inspected all the collected tweets from Chilling Effects by using Tor with a variety of exit nodes, and validated that the “withheld_in_countries” field is present in the JSON structure if the tweet is withheld in the target countries. We note that we get the same response, regardless of the country from which we make the request. This essential observation means that we can collect tweets from our home servers without needing to use any sort of in-country proxy servers. When
a user’s home country happens to have withheld a tweet, their users won’t see it. Figure 2.3 shows how withheld tweets are rendered. (See also, Section 2.7 for ramifications on this observation toward bypassing Twitter’s censorship mechanisms.)

### 2.3.5 Withheld Accounts

The Twitter documentation states that when an account is withheld, the “text” field will be assigned a specific value with a specific structure. For example: “text”: “@chatty’s account is withheld in: Greece, Hong Kong, Malaysia”, and that the field “scope” will be assigned the value “user”. We found this to be rather inconsistent and in some cases invalid. Throughout our experiments, we found that 72% of withheld accounts that we identified did not have the “user”:“scope” field format, but still generated tweets containing a “withheld_in_countries” field, rendering them indistinguishable from those generated by non-withheld accounts. In addition, the “text” field had the original tweet string being withheld, contrary to what the documentation suggested.

Ultimately, the only reliable method we discovered to determine whether an entire account is withheld, rather than just individual tweets, was to simply enumerate all of that user’s tweets. If they were all withheld, that was a reliable signal.

### 2.4 Case Study: Turkey

While it would be desirable to collect all withheld content, worldwide, the sheer volume of this would be impractical. Alexa reported that by January 2013 Twitter had 500 million registered users and that the service generates 500 million tweets daily [28]. Twitter en-
forces rate limits that make it impractical to collect this much data from their service, even when using the obvious bag of tricks (multiple Twitter accounts, multiple IP addresses, etc.—See Section 2.5.1 for additional details).

Instead, we decided to focus our study on one country with a reputation for censorship. Which one? We selected Turkey, due to its apparently vigorous use of Twitter’s censorship mechanism and its generally hostile behavior toward journalists and dissenting political speech [29]. Also, we have access to native speakers to assist us when automated translation falls short.

Notably, in May 2013, the Taksim Square riots [30] led to an “Occupy Gezi” movement (Gezi is a park next to Taksim Square, in Istanbul). Following this movement, Twitter achieved significant popularity in Turkey, gaining over a million new accounts [31].

Subsequently, in March 2014, then-Prime Minister (now President) Recep Tayyip Erdogan ordered Twitter blocked in Turkey for Twitter’s failure to implement Turkish court orders seeking removal of some links posted on Twitter. He also demanded that Twitter establish an office in Turkey to ease take down requests and improve Twitter’s accountability under Turkish laws. Perhaps unsurprisingly, Twitter declined [32]. During this event, Twitter instructed its users (via tweet) to continue tweeting via SMS. The hashtag #twitterisblockedinTurkey was trending amongst protesters and their followers. Later that month, Turkey’s highest court rejected the ban as a violation of freedom of expression [33].

In August 2014, a news report claimed that Turkey and Twitter were scheduled to meet for the third time to discuss the establishment of a new office for Twitter representatives in Istanbul [34]. Clearly, negotiations were afoot so Turkey could keep Twitter around, and Twitter could accommodate Turkey’s censorship needs.
It would be useful to estimate the volume of tweets per year in Turkey. According to Fox [35], 3.0% percent of Twitter’s active users are in Turkey. Sysomos [36] doesn’t list Turkey anywhere in its list of top countries although Baronchelli et al. [37] found Turkish to be the 10th most popular language on Twitter in 2013. Twitter does not disclose any such data itself. Ultimately, this means that any measurements we make of Twitter censorship can only be treated as a *lower bound* on the total volume of Twitter censorship. We unfortunately have no way to assure that our data collection will be a *representative sample* of the total traffic.

### 2.4.1 Disclosed Censorship

Starting in 2014, we noticed an increase in the number of Turkish requests posted to Chilling Effects. In addition, the Twitter Transparency Reports published in 2014 showed an increase in withheld Turkish notices following the unblocking of the Twitter service; Twitter itself reports 183 withheld tweets and 17 withheld accounts in the (Jan 1, 2014-June 30, 2014) report, and 1820 withheld tweets and 62 withheld accounts in the (July 1, 2014-December 31, 2014) report. Table 2.2 shows the distribution of number of withheld tweets by reporting period. Clearly, once Twitter was no longer blocked to Turkish citizens, the Turkish government availed itself of Twitter’s censorship mechanisms.

### 2.4.2 Collecting Censored Tweets

As we described previously, Twitter posts withholding notices to the Chilling Effects website. A sample Turkish notice is shown in Figure 2.2. This scanned document, and many more like it, is blurry and partially redacted by Twitter. We extracted all the Turkish Chill-
Table 2.2: Distribution of withheld tweets in Turkey, as reported by Twitter.

<table>
<thead>
<tr>
<th>Date</th>
<th>Number of Withheld Tweets in Turkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012: Jan 1 - Jun 30</td>
<td>0</td>
</tr>
<tr>
<td>2012: Jul 1 - Dec 31</td>
<td>0</td>
</tr>
<tr>
<td>2013: Jan 1 - Jun 30</td>
<td>0</td>
</tr>
<tr>
<td>2013: Jul 1 - Dec 31</td>
<td>0</td>
</tr>
<tr>
<td>2014: Jan 1 - Jun 30</td>
<td>183</td>
</tr>
<tr>
<td>2014: Jul 1 - Dec 31</td>
<td>1820</td>
</tr>
</tbody>
</table>

ing Effects documents posted through March 30, 2015, extracted the tweet IDs from the notices, fetched the tweets with Twitter’s REST API, and stored them in a local database. Overall, we identified 2896 tweets from Turkey with this method. Some tweet IDs appeared in multiple notices, so we removed duplicates as well as some apparently malformed responses, ultimately ending up with 2,473 unique tweets, of which 1,340 were still present on Twitter. The remaining tweets were either removed or were perhaps associated with “protected” users, whose tweets are normally only visible to permitted users rather than the whole world; for these tweets, each user would need to grant us permission to see their tweets in order for us to confirm their withholding status. (We decided not to pursue such permissions.) Of these remaining tweets, we confirmed 1,155 withheld Turkish tweets by examining the “withheld_in_countries” JSON field; this also shows that at least 86% of the Turkish government’s withholding requests for non-protected tweets
were approved by Twitter. We cannot estimate the approval rate for “protected” tweets, but assuming their actions on “protected” accounts are consistent with their actions on “public” accounts, we can confirm that Twitter seems to approve most of the withholding requests that it receives from Turkey.

As an interesting aside, it’s worth posing a question we cannot answer: how is the Turkish government managing to censor anything from “protected” users? These tweets are not visible to the world yet they appear in withholding requests. This implies several possibilities. Perhaps the censors are requiring keyword or hashtag-based censorship. Perhaps they’re demanding access to protected accounts. Or, perhaps the simplest answer is that these user accounts were once public but are now protected. We have inadequate information to determine what happened.

2.4.3 Collecting Censored Accounts

As above, we wish to identify censored user accounts, not just censored tweets. We found a total of 80 Chilling Effects requests for user accounts to be censored. Of these, 40 accounts appear to be withheld in Turkey. Of the remaining 40 accounts, 23 accounts appear to be suspended or deleted.

2.5 Broader Turkish Data Collection

At this point, we hypothesized that the visible Turkish censorship of Twitter was just the tip of the iceberg. To quantify the degree of censorship, we would need to collect much more data.
2.5.1 Data Sources

There are many ways of acquiring or purchasing large-scale datasets of tweets:

**The Twitter/Gnip Firehose**  The “firehose” is a special API, only offered to paid customers, that returns the entire live stream of tweets, as they occur. In addition, it allows historical retrieval of old tweets from an archive including those that are deleted. We contacted Gnip, a third party tweet-reseller that was recently acquired by Twitter. We received a costly quote of $12,000 per 1 million tweets. They explained that we could purchase Turkish tweets by filtering on the language. However, they could not guarantee to find all tweets containing the “withheld_in_countries” field as censorship may occur after the tweet was captured. They requested that we complete special forms disclosing our research goals to obtain an internal approval from Twitter. We declined to pursue this relationship.

**Twitter Free Public APIs**  Many researchers, like us, are ultimately forced to use Twitter’s free public Streaming and REST APIs. These only return 1% of the total public stream, with no particular explanation of how the 1% are sampled from the overall population of tweets [28]. Furthermore, these APIs allow only 180 calls per 15 minute interval and requires that we register with a set of OAuth credentials. Ultimately, we procured many sets of these credentials, allowing us to at least partially overcome these rate limits, but we would certainly be unable to fetch every single tweet, much less revisit selected users every few minutes as Zhu et al. [1] did while looking for censorship on Weibo.

Morstatter et al. [28] found that the volume of tweets obtained using Twitter Stream-
ing API, versus Firehose, depended on the coverage of the streaming API data. The more complete the filtering criteria is, the less coverage we might get. More usefully, they found that when the geo property was used, the coverage was almost complete. Consequently, this was the route for data collection that we pursued:

- Sampling random Turkish tweets using geo coordinates
- Follow Turkish “sensitive” users

2.5.2 Data Collection

Our general methodology was to collect large volumes of tweets, using these free mechanisms, then revisit them occasionally to discern if any had been censored. In contrast with Zhu et al.’s goal of identifying how fast censorship occurs, we were more interested in collecting as many censored posts as possible. Given the manually intensive review process that Twitter appears to enforce on governments, there does not seem to be any useful information in more precise timing, while volume measurements are still quite valuable.

**Phase I: Streaming API using Turkish geo coordinates** The Twitter `POST statuses/-filter` Streaming API includes an optional parameter called “location” that takes a set of geo bounding boxes (latitude/longitude) to stream tweets geographically. These appear to be set by Twitter’s smartphone clients, although we did not make a detailed examination. Instead, we queried with geobounds corresponding to three major Turkish cities: Izmir, Ankara and Istanbul.
We ran our streamer from October 2014 through January 2015, and collected 17 million tweets.

**Phase II: Revisiting Tweets, looking for censorship** Since there is a time gap between sending the government-request and actually withholding the requested tweets by Twitter, we decided to wait before re-inspecting the collected tweets to see if they became withheld. In February and March 2015, we revisited the previously collected 17 million tweets using the REST APIs, and found 3,258 withheld tweets. Our data clearly shows that there are far more censored tweets than 1,155 we found on Chilling Effects. We also observed that we managed to capture some censorship events of tweets prior to our phase I data collection. These corresponded to users we were following (more on that below), so they are not a representative sample of censorship events, but they are an existence proof of censorship events reaching well into the past.

**Phase III: Friends of sensitive users** In some cases, Twitter’s APIs will not only return a stream of tweets from a set of users being queried, but will also return replies, mentions, and retweets. Since followers and friends of censored users are perhaps more likely to be censored themselves, this means that, like Zhu et al. [1], we can spider outward from a small set of known-censored users and derive a larger set of “interesting” users, then collect all of their tweets. Following this process, with our original set of censored tweets as a baseline, we ultimately collected 689 unique user IDs that have been subject to withholding. Ultimately, we collected almost 1.7 million tweets (i.e., an average of 82 tweets per day per user; these are very active tweeters) from these users in March 2015. Of these,
46,769 were withheld.

Repeating the process again, we expanded to the followers of this larger set of censored users, ultimately yielding nearly 85,000 user IDs. We then scanned for each user’s tweets, ultimately yielding a total of 171,652 withheld tweets. Curiously, 386 of those tweets were not from Turkey, with the bulk coming from a Brazilian user, unhappy at having been fired from his job, he claimed, due to his sexual orientation. We also found 3 Russian tweets and 13 whose country was listed as “XY”, which we understand to refer to tweets removed for DMCA violations, along with a corresponding “withheld_copyright” field. Removing all of the non-Turkish tweets, we ultimately found 171,266 withheld tweets in Turkey.

We also looked at censored retweets, leading us to almost 3000 new user IDs and yielding an additional 45,114 withheld tweets beyond the 171,266 above.

Altogether, we discovered 266,407 censored tweets (3258 + 46769 + 171266 + 45114).

2.5.3 Analysis

Using these methods, as discussed above, we discovered a grand total of 266,407 censored tweets generated by 7,642 distinct users. Of these, we estimate that 46 of the user accounts are, themselves, withheld in their entirety. Of the 266,407 tweets, we identified 205,451 tweets that were not generated from the 46 withheld accounts. This represents two orders of magnitude greater censorship of tweets in Turkey than disclosed by Twitter through its Transparency Report and the Chilling Effects website.

This difference is no statistical anomaly. Twitter is deliberately underreporting the volume of withholding in Turkey, perhaps as a result of legal requirements made of it by
the Turkish government as part of allowing them to continue operating in Turkey. No such agreements have been publicly disclosed.

We next consider two simple analyses over our data: tweet censorship volume per user and deduplication. We apply machine learning techniques toward topic analysis in Section 2.6.

**Censorship Volume per User**

We extracted the user accounts with the largest number of withheld tweets, and manually examined their profiles. We found that nine of the ten users tend to tweet political anti-government content, and appeared to be influential given the large number of their followers, and one account tweeted hacking information and appeared non-influential. Figure 2.6 shows the probability distribution function (PDF) of withheld tweets by user, where the users are sorted from the most withheld to the least withheld. The distribution is heavily skewed to the left, indicating that a small number of users generated a large number of withheld tweets. This shows that the top 100 censored users generate 44% of the censored tweets, and the top 500 censored users generate 73% of the censored tweets. This observation also corresponds to our withheld account vs. non-withheld account observation, in which withheld accounts are influential accounts tweeting frequently with their whole accounts being withheld. The non-withheld accounts are normal persons who sometimes tweet withheld content.

The long-tailed aspect of the censorship distribution (see Figure 2.7) has important ramifications for understanding Turkey’s Twitter censorship. If we saw a short tail, then this would imply that Turkey was focusing its attention on the biggest rabblerousers, but
instead we see a broad focus over a large population of users. This also implies that Turkey’s censors are using search tools to discover these long-tail tweets that would not otherwise rise to their attention via organic means such as having been posted by celebrity users or having gone viral via retweeting.
The long-tail also implies that popular users cannot simply discard their identities and start over to fly below the radar. As an example of this, we manually dug deeper into our data to identify censored users. Users clearly become aware that they are being censored, then create new accounts which again suffer censorship. Figure 2.8 shows one example of a withheld user with two accounts and many followers. The older account seems to be the one with over 75k followers while the newer one has 12k followers.

Figure 2.8 : Same user with two profiles withheld in Turkey.

Duplicate Text Analysis

An interesting question, now that we have established such a large number of withheld tweets, is the extent to which they duplicate one another. Our prior analysis considered tweets with unique IDs, but users can “retweet” one another’s tweets, either through official mechanisms (which generate metadata headers) or by simply cutting-and-pasting the text from an old tweet into a new tweet, perhaps prepending the text with in-line metadata, typically “RT @” and the username of the original author.

Stripping out the “RT” prefixes, ignoring retweet metadata (both inline and in the rest of the JSON structure), and focusing strictly on the message body text for the 266,793
withheld tweets, we ultimately yielded 88,276 unique strings. In other words, we found 178,517 retweets of various styles. Digging a little deeper, it appears that all but 15 of them originated with the standard “retweet” button, with those 15 being the result of cutting-and-pasting raw text.

We extracted the top 10 most duplicate tweets and found that all 10 tweets had bad things to say about Recep Tayyip Erdogan, the current Turkish president, and appeared to be generated by supporters of *Fethullah Gulen*†.

Additionally, we looked if there are users that reposted or retweeted the same thing more than once; we found 295 users posted a total of 1,503 tweets in this category. We manually reviewed the top-ten most prolific users in this set. Two users regularly post anti-government topics and have large number of followers, suggesting that they are influential. The third account appeared to be a marketing bot for a software product; it has a low number of followers, but generates a large number of tweets.

We finally ask the question of whether censorship of a retweet has any bearing on censorship of the original. We note that the JSON structure for a retweet contains the ID of the original, so we collected the withholding status of each original tweet for every withheld retweet. We ultimately discovered that 92% of these original tweets are also withheld. What about the remaining 8%? Half of them survived, uncensored, while the other half belong to withheld accounts (see Section 2.3.5). This suggests that, through whatever mechanism Turkey is directing its remarkable volume of Twitter censorship, there is either some amount of human discretion involved, or the mechanism has some

†*Fethullah Gülen*: a Turkish preacher and founder of the Gülen movement [38].
degree of inaccuracy in its targeting.

2.6 Censorship Topics with tf-idf and NMF

Now that we have established a remarkable volume of Turkish censorship of Twitter, the next question is to understand what topics the Turkish censors are interested in. This sort of analysis is valuable for understanding the political aims of the Turkish censors. It is also pragmatically valuable to determine hashtags and topics that would help in discovering additional censored tweets that our earlier methods may have missed.

Topic extraction and clustering is a standard feature of natural language processing systems. The key concept behind automatic topic extraction is to assign weights to terms and sentences based on their frequency of appearance. We applied non-negative matrix factorization ($NMF$) combined with term frequency–inverse document frequency ($tf-idf$) to extract hot topics. $tf-idf$ is a weighting schema widely used in document classification. The $tf-idf$ value of a word increases proportionally to the number of times it appears in the document, but is also offset by the frequency of the word in the corpus. The advantage of $tf-idf$ over a simple word frequency is that it can effectively adjust weights of words that are very frequent but not informative [39].

$NMF$ is a technique to factor document-term matrix into a term-topic and a topic-document matrix. The topics are derived from the contents of all tweets. $NMF$ has been widely used in text mining related applications including clustering on email message, scientific journals and Wikipedia articles [40]. The method has been shown to be effective in monitoring underlying semantic features (topics) in a general way [41].
2.6.1 Hot Topic Clustering

We first tokenize each censored tweet into a list of words, eliminating Turkish stop words. After this, we applied tf-idf, built the document-term matrix, and used non-negative matrix factorization on the document-term matrix to extract the top 5 topics with 10 words for each topic. The result of this process is a series of Turkish words that may be best interpreted by a Turkish speaker, although Google’s Translation service is very helpful. See Table 2.3 for the hottest censored topics.

(This method could potentially be improved in a variety of ways, e.g., exploring the quality and stability of topic clusters as a function of the number of topics, using n-grams rather than unigrams, or varying the number of words in the tf-idf document representation. Nonetheless, our method still yields interesting results.)

Topic 0 is about cursing Vatan and Hurriyet media owned by Aydin Doğan, who owns largest media cooperation in Turkey, and calling them dogs. These media groups are known to favor the leading Justice and Development Party (AKP). Topic 1 is related to the word şapşık, which does not have a direct interpretation in the Turkish language. This word was used by the Kurdish party leader once in his speech, and suddenly the word became popular, and people started making fun of it. This word increased his popularity, and people find it sympathetic. This topic is also related to Şekerbank, a financial institution in Turkey, and a person name İbrahim Karaca.

Topic 2 discusses the Koç family in Turkey. Vehbi Koç, who founded Koç Holding A.S. in 1963, is a Turkish entrepreneur and philanthropist. His son Rahmi Koç took his

---

‡Google has a nice library for stop-word removal. [https://code.google.com/p/stop-words/](https://code.google.com/p/stop-words/)
father’s position as the chairman of the company in 1984 and retired in 2003. Aydin Doğan was mentioned in topic 0. Some people claim that Aydin Doğan is Vehbi’s secret son, and that Vehbi Koç’s father is Haim Nahum, a Jewish rabbi. Following the crackdown of the Taksim Gezi Park protests, Erdogan went after the Koç family as they had criticized his actions [43].

Topic 3 discusses the controversy ignited by Lütfi Elvan, head of Minister of Transport, Maritime and Communication in Turkey. Elvan proposed to establish Turkish own web protocol using “ttt” as a prefix instead of “www”. This suggestion has led to broad criticism inside Turkey.

Topic 4 is about people’s disappointment and condemnations of current Turkish Prime Minister Ahmet Davutoglu. The topical words found by our analysis appear to reference
his anti-Jewish rhetoric, and summarize the use of vulgar language to describe him.

Evidently, strongly worded and vulgar political discourse are top on the minds of Turkey’s censorship authorities.

### 2.6.2 Hot Topics from Withheld Accounts

We applied a similar topic analysis methodology toward tweets from withheld accounts (i.e., accounts where every single tweet has been withheld versus accounts where only some tweets are withheld). There are 46 such accounts in our dataset. However, the algorithm failed to extract meaningful topics due to the apparently diverse contents of the withheld tweets. The resulting keywords were mostly usernames and hashtags. Consequently, we performed a manual inspection of the withheld accounts by looking at their Twitter profiles and their timelines tweets. Table 2.4 summarizes our impressions of these accounts.

Users from the “Politics” category constantly post anti-government comments. These

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>36</td>
</tr>
<tr>
<td>Pornography</td>
<td>2</td>
</tr>
<tr>
<td>Advertising Bots</td>
<td>1</td>
</tr>
<tr>
<td>Unidentified</td>
<td>3</td>
</tr>
<tr>
<td>Not Found</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2.4: Topics discussed in withheld accounts.
user also openly criticize current Turkish president Erdogan by attacking his personality or posting political caricatures. They also tend to have a large number of followers.

Users from the “Pornography” group, constantly posts explicit images and video clips with links to other pornography-related Twitter accounts or websites. Users from this group also have a reasonably large number of followers.

Users from the “Advertising Bot” category repeatedly send tweets with similar or identical contents. Each tweet has a link at the end that will redirect to a company’s website.

We classified three users in an “Unidentified” category. They did not appear to be engaged in political speech or anything else which would appear to be worthy of censorship.

Finally, users from the “Not Found” category were originally in our dataset but when we later looked at their profiles, we could not find their accounts anymore. These accounts could have been deleted by the users’ own actions or through the actions of Twitter (perhaps on the behest of the Turkish government, perhaps not).

2.7 Bypassing Censorship

In April 2015, we followed a group of withheld accounts in Turkey and noticed that at least 7 users were still tweeting from inside the country despite having their entire accounts withheld. We manually inspected each profile and found that these users tweeted topics political in nature, specifically criticizing Erdogan’s leadership. We also noticed that all of these users had large number of followers, and were thus presumably quite influential.

We note that these users were originally found using the Streaming API with geobounded boxes set to Turkish cities. Presumably, they all live physically in Turkey and
wish to be read by a Turkish audience. So how are they being seen?

We found that Twitter manages location information by a cookie set in the browser. We tested this by viewing a known-censored tweet using a Turkish proxy server and a regular Internet browser from our home institution. In both cases the known-censored tweets in Turkey were not presented as having been withheld. Conversely, using a proxy did not help us bypass censorship. However, when we changed the location setting in the Twitter application and set it to “Turkey”, surprisingly the tweet appeared withheld despite our physical location outside of the country. It is possible that this trivial hack is used inside Turkey. All a Turkish user needs to is simply set their “country” to be elsewhere and they can see everything. We searched the web both in English and Turkish languages for instructions on bypassing censorship in Turkey, to see if this seemingly obvious advice was described anywhere on the Web. We found many Turkish descriptions on how to use Tor or paid VPN services to evade censorship (see, e.g., [44, 45]). However, we found no results on evading Twitter censorship using our method (i.e., changing the location setting). It is likely that this method will gain popularity after the publication of this paper.

Perhaps Twitter is being deliberately simplistic with its censorship mechanism, doing what it interprets to be the bare minimum necessary in order to operate its service in countries that might otherwise ban it altogether. Certainly, Twitter is attempting to walk a fine line to allow its users to have their tweets as widely visible as possible, despite different nation-states having different legal policies for what must be censored in their respective jurisdictions.
2.7.1 Censorship Escalation

Consider what might occur if Twitter gets more serious about censorship. With basic IP geolocation techniques, which Twitter appears to already employ, Twitter could base censorship decisions on these geolocations rather than on its easily-changed cookie. In response, users could use proxy or VPN services (Tor, etc.) to obfuscate their locations. If this continues, the logical conclusion is that Twitter will have no reliable signals that indicate a user’s location.

What then? Countries may attempt to strong-arm Twitter into replacing its “withholding” mechanism with a more draconian deletion policy, or else ban them from the country. In such a circumstance, what would it mean for a user in one country, criticizing another country, to find their tweet censored worldwide through no action of their home government? (Example: if Twitter cannot reliably distinguish an American writing a tweet about Turkey from a native writing the same tweet, and the Turks press their case, they may demand the ability to censor the American’s tweet.) This draconian progression is the seemingly inevitable path that Twitter will be forced to follow. Google is fighting against requirements that it follow a similar path with respect to European “right to be forgotten” restrictions on its search results [46]. Both Twitter and Google’s mechanisms of withholding content in one country while preserving it in others are fundamentally fragile and would seem unlikely to survive in the face of insistent government-sponsored censorship.
2.7.2 How Many Users Use Tor in Turkey

As discussed above, Turkish users are already aware of Tor as a mechanism for overcoming censorship. How popular, then, is Tor in Turkey? The Tor project has a public tool “Tor Metrics-Direct users by country” that estimates the daily number of directly-connecting clients to the Tor network from any country, and also displays “indications of censorship events” [47]. Figure 2.9 shows a Tor metrics graph for Turkey. Red dots indicate possible censorship events and blue dots indicate a “release” of censorship events. We also plotted the distribution of all withheld tweets we detected and those we collected from Chilling Effects, ordered by the “created_at” month as shown in Figure 2.10.

Both graphs follow roughly the same time period, from 2013 through 2015. There are two noticeable peaks on the Tor graph, in May 2013 and March 2014, which correspond to the end of the Taksim Gazi Park protests (2013) and a brief ban on Twitter across all of Turkey (2014). The withholding graphs, based on our own data collection, have peaks as well, which seem to correlate with these same two events, but the overall trajectory of the withholding is very much upward, versus the apparently otherwise steady-state usage of Tor. This suggests that, to the mind of the Turkish Twitter censors, the withholding mechanism is “good enough” for their needs. The Tor volume suggests that the Turkish population agrees. The y-axis of the Tor graph is the number of daily unique users of Tor, so the Turkish government may be willing to allow 20,000 people to have unfettered Twitter access so long as the larger population’s ability to see and spread information is sufficiently squelched.

§https://metrics.torproject.org/userstats-relay-country.html
Figure 2.9: Tor Metrics: users connecting from Turkey.

Figure 2.10: Distribution of withheld tweets by month (log scale).
2.8 Summary

In this work, we collected data on the frequency of Turkey’s use of Twitter’s censorship mechanisms. Through Twitter’s own transparency reporting, Turkey is one of the countries with the most censorship so we focused our attentions there and found an estimated one order of magnitude more censorship than Twitter officially reports. The actual numbers might be even higher, but it is difficult to be certain due to Twitter’s rate limits and other restrictions on external groups who want to crawl their content. Certainly, our results make it clear that Twitter’s transparency reports are entirely unrepresentative of the actual scale and scope of censorship that Twitter enables.

There are a variety of directions for future work related to this study. Expanding this work to study Twitter censorship in other countries is an obvious direction, particularly in countries with no reported censorship requests. A full country-by-country study would require significantly greater resources than we could muster from our home institution. While we certainly could scale up with modern cloud services, engineering a full-scale web crawler dedicated to downloading each and every tweet as it is posted, with periodic followups to check every tweet’s withholding status, would represent a non-trivial load on Twitter’s service; Twitter would presumably then take technical and/or legal steps to block such research crawler. Alternatively, Twitter could allow researchers access to its “firehose” without restrictive agreements as to how they may or may not use the data, although this seems politically unlikely.

Twitter ultimately faces a stark challenge. By doing business in countries that require censorship of tweets, including the requirement that Twitter wildly distort its reporting
of *statistics* over those tweets, while simultaneously allowing users outside of those censorship regimes to read and respond to those same tweets, Twitter is trying to placate its censors while still maximizing what free speech might be leftover. This is not a long-term stable business strategy, and we fear freedom of speech may be the loser in this battle.
Chapter 3

Modeling Influential Users and Users’ Communities

3.1 Influential Users

Graph metrics and network analysis are commonly used by researchers to study social network characteristics and extract meaningful users’ information \[48,49\]. In this chapter, we show how we identifying influential users and user communities in dynamic twitter graphs. Within the context of our work, we define influential users as users with the highest out-degree (users who generate most original data), whose tweets reached most users in their social graph and are considered to be the data sources. We propose a simple algorithm that generates a complete dynamic data-flow graph to represents a snapshot of the direction of data communication among Twitter users who engage in censorship-likely topics. Where traditional approaches construct social graphs using friendship relationships and retweets only \[49,48\], we are instead interested in modeling all relationships including retweets, replies, and mentions, and users who are indirectly connected.

3.2 Related Work

Yamaguchi et al. \[50\] showed methods for identifying authoritative users (i.e. influential users) in Twitter by examining information flow in Twitter using data link analysis by
assigning an authoritative score to each user in the social graph. They argue that their approach is unlike the existing methods that only deal with following and follower static relationships [51][52]. Twitter users can be categorized as “information generators” who post important tweets and tend to have many followers, and “information seekers” who are users that follow a large number of users and are interested in reading more than posting tweets [50]. However, these static following/follower relationships do not tell us much about the level of interaction between these users because many users follow back other users as a courtesy. Huberman and Romero [53] found that the vast percentage of users in the following/follower relationship do not interact with each other, so examining the link between thousands of users (in many cases millions of users) is not useful. Moreover, over the years Twitter, attracted a large number of automated accounts known as bots that tend to randomly follow a large number of users and retweet them in hoping to get followed back [54]. For these reasons, using the static followers/following relationship to identify influential users is not very useful. The Yamaguchi et al. rating algorithm considers the relationship from a user \( u \) to other users that \( u \) follows, which is a static relationship that can be obtained from Twitter’s API. Next, they consider the retweet dynamic relationship. The social graph they built consists of nodes and edges, where nodes are user-nodes and tweet-nodes. The edges can be one or more of the following bi-directional relationships: a link from user \( u \) to a tweet that \( u \) posted, a link from user \( u \) to a user that \( u \) is following, a link from tweet \( t \) to its retweet. After they obtain the graph, they get an authoritative score based on the calculated weight of the edges. Twitter allows users to mention, reply, and retweet other users that are not directly connected (not following each other). This could occur for example, when the tweet is found in the public feed, or posted by some
intermediary users. We argue that although their approach can be better than the existing methods, it only considers users that are directly connected via following/follower relationships, and ignores users that received the tweet from their timeline but are not directly connected to the user who generated the tweet. This means that their approach missed many users that could also be influential. Additionally, their approach only considers the retweet relationship, and ignores the links between users that are mentioned or replied-to.

3.3 Tweet’s JSON Fields

Twitter offers many free APIs for collecting public tweets, including the Streaming API[1] which captures continuous live tweets up to 1% of total Twitter feed, and the REST API[2] for retrieving historical tweets, such as tweets of specific users, or by tweet-ID. In this section we provide a brief description of some of the important fields in the tweet data block returned by some of the Twitter APIs that are relevant to our analysis. Some of the fields we present below could be absent or NULL in the returned JSON file (i.e. the returned value from the API resource). The values of these fields help us determine when tweets are: retweets, replies, mentions, original, or censored[3].

1. “user”: This field contains attributes of the tweeting user, including, user-name, screen-name, user-id, profile description, and many other attributes.

2. “retweeted_status”: This field when present, means that the tweet is a retweet, and

---

[1] Streaming API: https://dev.twitter.com/streaming/overview
it contains the representation of the original tweet including attributes of the original user and the original tweet. Note that a retweeted-retweet does not contain attributes of the intermediary retweet, but only the original tweet. For this reason, we consider the relationship between the retweet and the original tweet a logical relationship, because the users are not directly connected in the social graph.

3. **“user_mentions”**: Is a sub-field that contains a list of users being mentioned in the actual text.

4. **“in_reply_to_screen_name”**: This field indicates that the tweet is a reply. When present, it contains the screen-name of the original user.

5. **“withheld_in_countries”**: This field is only present if the tweet is censored. When present, it contains a list of ISO-country codes in which the tweet is censored. For example, “withheld_in_countries”: [“TR”], means that the tweet is censored in Turkey only. In our previous work [55], we validated that when this field is present, the tweet is greyed-out when viewed from inside of the censoring country.

As we mentioned earlier, tweets may have slightly different data block structures depending on the API being used. Specifically, we discovered during our experiments that a retweet collected using the Streaming API (version 1), does not include the "retweeted_status" field, appearing only as an original tweet. When the same tweet is recollected using the REST API, it includes the "retweeted_status" field, identifying the original user attributes. This finding is important as it allows us to find additional users that may be of interest to the Turkish censors. We also note that as Twitter continues to update its APIs, this feature
may change in the future.

3.3.1 Constructing Users’ Influence Data-Flow Graph

We introduce a simple algorithm for constructing a complete data-flow graph using a wide variety of metadata embedded in tweets to follow data paths of censored tweets, instead of the traditional approach of constructing such graphs using friendship relationships and retweets only as we explained in the previous section. To model the data-flow between users in our Twitter dataset, we converted our set of censored tweets to a directed graph $G(E,U)$, where $U$ is the set of all users in our data, and each user is a graph vertex. $E$ is the set of directed edges, representing the data flow links between users. Edges are added to the graph by examining each tweet’s JSON data structure returned by the Twitter API resource, and applying the appropriate rule defined in Table 3.1. The rules are described as following: (1) When a tweet is original ($O$), the user is a singleton, and no edge is added. (2) When the tweet is original ($O$) but contains a user_mentions subfield, edges are added from the original user to each of the mentioned users, this is because when users are mentioned in a tweet, the mentioned users receive the tweet, and the data flows from the original user to those being mentioned. (3) If the tweet contains a retweeted_status field, an edge is added from the original user to the retweeting user, this is because the source of the data is the original user. (4) When the tweet contains retweeted_status field and the embedded original tweet (source) contains user_mentions subfield, rule (3) is applied, then edges are added from the original user (source) to each of the mentioned users. Finally, (5) When the tweet contains in_reply_to_screen_name field, an edge is added from the replying user to the original user, this is because a reply tweet generates new data which
flows from the replying user to the original user. Consequently, we are capturing a more robust set of social relationships including those of users that are disconnected in a static relationship than we might arrive at by examining retweets or following only.

<table>
<thead>
<tr>
<th>Num</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If $O$, then no edge, it is a singleton</td>
</tr>
<tr>
<td>2</td>
<td>If $O&amp;M$, then, $\forall M$ add edge($user_O$, $user_M$)</td>
</tr>
<tr>
<td>3</td>
<td>If $RT$, then, add edge($user_O$, $user_{RT}$)</td>
</tr>
<tr>
<td>4</td>
<td>If $RT&amp;M$, then, add edge($user_O$, $user_{RT}$), and $\forall M$, add edge($user_{RT}$, $user_M$)</td>
</tr>
<tr>
<td>5</td>
<td>If $R$, then, add edge($user_R$, $user_O$)</td>
</tr>
</tbody>
</table>

Table 3.1: Data flow rules for adding directed edges

Note that because Twitter allows users to mention, reply, and retweet other users that are not directly connected (not following each other), our algorithm allows us to captures additional users from the metadata that we were not specifically crawling, while also identifying new users who can also be influential. Users with the highest out-degree number (users who generated most data) are users whose tweets reached most users in our graph and are influential.

### 3.3.2 Community detection

After constructing the social data-flow graph using the above process, we apply a standard data clustering metric to identify user communities, known as modularity. Modularity
is a widely used metric for extracting network structures [56] and studying communities in social networks [48]. Studying communities and their members is practically useful for conveying important information about common topics and users’ characteristic. In social networks, specifically, friends tend to communicate with each other more often than they do with other users that are members of other communities, therefore each module or community consists of “densely connected nodes with scarce connections” to other modules [57]. Modularity defined by Newman [56], is explained as follows: When $G$ is a directed graph, modularity $Q$ is defined as the number of edges that exists in communities in $G$, minus the same number of edges expected if the edges were distributed randomly. Modularity $Q$ is noted by Newman in Formula (3.1)

$$ Q = \frac{1}{2m} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{2m} \right] $$  

(3.1)

Where $m$ is the total number of edges in directed graph $G$, $v$ and $w$ are two random nodes, and $k_v$ and $k_w$ are the out-degree for $v$ and $w$ respectively. When there is an edge between $v$ and $w$, then $A_{vw}$=1, otherwise it is 0 [48]. Dividing $G$ into two communities, the algorithm is then repeated recursively until $Q$ is maximized. The Louvain modularity method, which we chose to use, is known to run faster by using greedy optimization [58], which is appropriate for processing very large networks.

In the chapters to follow we demonstrate how we apply these methods to detect influential users and their communities from censored tweets.
Chapter 4

The Dynamics of Social Media Censorship in Transitioning Democracies

4.1 Introduction

Given the increasing importance of social media in social and political life, scholars across a number of disciplines have been fascinated with understanding not simply how people use this medium, but also how governments regulate users, content, and usage [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Censorship is one form of regulation that is of particular interest since it has such profound implications for democratic institutions and processes, social movements, and democratization. While most observers might expect China, Iran or some other dictatorial regime to rank first on the list of countries that censor social media, the fact that Turkey, a hybrid or semi-democracy, actually holds this position likely comes as somewhat of a surprise. In fact, according to Twitter Transparency Reports, Turkey accounted for 88% of global censored accounts and 84% of global censored tweets between 2014-2015 [12]. And, Brazil another transitioning democracy, ranked second on Twitter Transparency’s list. Despite the prevalence of social media censorship in transitioning democracies [13], we know very little about the nature or dynamics of censorship in these states. The reason is that most existing studies have focused almost exclusively on authoritarian regimes like China and Syria. This work finds that Internet censors do not
necessarily target social media posts that solely criticize the government. Instead government censors in authoritarian regimes focus on social media posts that represent, reinforce, or spur collective action events [59, 60]. As King and his colleagues argue, the reason is because the government perceives collective action events as a major threat to the stability and order of the country, and hence, its survival [61].

Is the fear of collective action also what motivates government censors in Turkey and other transitioning democracies? If not, what does? Who and what are the targets of Internet censorship in these semi-democracies and to what extent, if any, does censorship in non-dictatorial regimes differ from that in dictatorial regimes? This study addresses these questions, by focusing on Turkey as a case study. While Turkey’s unprecedented record of Twitter censorship makes it a relatively extreme case to examine, it provides the best insights to behavior that is likely to be modeled and adopted by other semi-democracies as they continue to develop the technological capacity to better regulate and control social media usage among their citizenry, and if and when incumbent parties face potential challenges that cannot be addressed using more democratic means.

Data for this study in this chapter, include 25 million Turkish tweets posted in Turkey between late-2014 and late-2015. Using these data, which include 712,218 censored tweets, we empirically investigate what and who is being censored in Turkey. We do this by first testing whether posts referencing collective action events are censored. If they are, this suggests that censorship in semi-democracies may not be that different from censorship in dictatorial regimes. However, we also test a new hypothesis based on insights from

*We applied the same methods from Tanash et al.’s (2015) to collect and identify censored tweets, and used some of their data.*
our theoretical framework. Focusing more explicitly on the incentives and constraints of incumbent governments in semi-democratic states, we posit that in contrast to authoritarian regimes, electoral constraints in semi-democracies encourage incumbent parties to prioritize the censorship of policy-related posts that have the potential to undermine the party’s approval ratings and reelection prospects. Using topic-clustering techniques, we test to see if Turkish censors target social media content that communicates users’ dissatisfaction with issues or problems that are highly salient to the public and particularly sensitive for the ruling party. Finally, to analyze who is being censored in Turkey, we use network analysis techniques to identify top influential users in our censored data. We expect that influential users promote important issues and are followed by a large number of audience who share the same opinions, and hence, are targeted by the Turkish censors.

Our analysis shows that a perceived threat of collective action does not explain censorship in Turkey. Instead, we find support for our hypothesis that social media censorship in transitioning democracies is targeted at sensitive public policy issue topics—those perceived by the public as important, but on which the ruling party was unwilling or unable to address. In the time period we analyzed, in 2014-2015, these issues included the corruption scandals of Turkey’s ruling AKP party and escalating violence between the government and Kurdish groups in southeastern Turkey.
4.2 Explaining the Impetus and Dynamics of Censorship Beyond Dictatorial Regimes

Our investigation of the impetus and dynamics of censorship in transitioning or semi-democracies begins with a consideration of the incentives and constraints of incumbent governments in authoritarian and semi-democratic countries. While incumbent governments in both sets of countries are concerned with their survival, they face different threats and challenges when it comes to holding on to power. A key difference is the role of elections and public opinion. While elections do take place in most authoritarian regimes, they are typically not fully free or fair, and thus have little influence on the selection of ruling elites or the policy priorities of government \[62\]. Because citizens “cannot throw the bums out” \[63\], elections pose little threat to the ability of incumbent parties in authoritarian regimes to stay in power. This in turn gives ruling parties less incentive to worry too much about citizens’ criticisms of their policies or public opinion more generally. On the other hand, because citizen protests, demonstrations and civil unrest can undermine the legitimacy of authoritarian regimes and weaken their political support they pose real threats to the survival of these regimes.

The same is not true in transitioning democracies. Here, the existence of competitive elections means that incumbent parties must pay more attention to campaigns, elections, and public opinion \[64, 65\], since failure to do so may risk their reelection prospects \[66\]. Building on Riker (1962) \[67\], a large body of work has examined how parties strategically mirror citizen priorities in their legislative behavior or party platforms \[65, 68, 69, 70\]. These studies find a widespread congruency between the issue areas perceived by the pub-
lic as highly important and the issue areas prioritized by parties. Research also finds that parties that fail to respond to shifts in citizen priorities are more likely to lose public support to other parties [71] [66]. For example, the emergence of and increasing voter support for green parties in Europe is partially explained by this phenomenon [72]. Hence, in democratic systems, even transitioning ones, parties are often more likely to win elections when they propose platforms and policies that address the issues and problems that voters perceive as important.

Of course, parties cannot be responsive to every problem and issue the public cares about, and oftentimes, it is the government itself that is the target of the public’s dissatisfaction. For example, corruption scandals, unpopular wars, inflation, and economic downturns are frequently at the center of voters’ discontent. When incumbent parties are unwilling or unable to act in ways that address public concerns about these and other issues, they may pursue alternative, less democratic means to get reelected. One approach might be for the party to engage in patronage or clientelistic practices in the hopes that voters would be willing to extend support to the incumbent party in exchange for material rewards and benefits. While this has long been the strategy of many ruling parties in transitioning democracies, economic restructuring and increasing globalization have significantly increased the costs of clientelism [73]. In addition, if parties campaign on more programmatic platforms, as they tend to do as democratic consolidation progresses, voters come to expect more and are therefore less likely to be content with clientelism. Indeed, since clientelism is associated with more authoritarian regimes, it tends to undermine not strengthen democratization.

Another approach is the manipulation of media, which can occur in different forms.
One is direct control of newspapers and television stations by the state. Another is pressuring privately owned media organizations or prohibiting the publication of their work if they dissent [74].

However, the expansion of the Internet and social media makes it impossible for parties or governments to fully control the media agenda (and hence manipulate citizen priorities) unless it has some way to censor the content of social media. It is precisely because social media messages are able to spread so rapidly to such large swaths of electorate that they pose such serious threat to incumbent parties. Not only do governments fear losing control of social media messages, but if left unchecked, the spread of posts critical of the government can also undermine popular support, particularly when the posts are about issues salient to the public and sensitive to the ruling party.

In such situations, ruling parties in transitioning democracies find it difficult to resist tampering with the flow of information on social media. Therefore, we hypothesize that in transitioning democracies the overwhelming majority of social media censorship focuses on national issues that citizens perceive as top priorities. In other words, the major target of the censors are the social media conversations and the individuals that promote certain issues, which are perceived as the most important national problems by the public.

4.3 Censorship Mechanism in Turkey

Turkey has been consistently categorized as a transitioning democracy by the Freedom House democracy index for almost two decades. Its overall liberty score has ranged from 3 to 4.5 since 1999 when Freedom House published its first report on Turkey. We se-
lected Turkey as a case study not only because of its unprecedented record of social media censorship, but also because it provides important insights about behavior that is likely to become more widespread as ruling parties in other semi-democracies advance technologically and develop greater capacity to regulate and control social media usage among their citizenry or a democratically elected leader decides to resort using less democratic instruments in order to deal with decreasing approval rates. The recent rise in right-wing populism in the Western societies (as seen in Brexit referendum and 2016 US elections) may possibly trigger the occurrence of such events in transitioning democracies. Notably, Turkey consistently ranks at the top in Twitter Transparency Reports, which provide list of countries that censor the most [12]. For example, in the first and second half of 2014, Turkey was responsible for 90% of all censored tweets and 72% of all censored accounts in their entirety worldwide. During the first and second half of 2015, Turkey was again the leader in Twitter censorship, accounting for 82% of all censored tweets and 91% of all censored accounts worldwide. In the 2016 first bi-annual report, Turkey accounted for 80% of all censored tweets and 44% of all censored accounts globally [12]. Turkey manages this massive social media censorship effort using an institutionalized (or at least quasi-institutionalized) mechanism, which is outlined in Law #5651 [75]. This level of institutionalization differs from the censorship processes in dictatorial regimes where an army of censors blocks sites and posts without any accountability. The process of censorship in Turkey starts with the initiation of a court case by an individual or agency claiming that certain content violates slander laws, is against national interests, or contravenes a provision defined in the law. Individual citizens can apply to a court directly or through their lawyers (see Al Jazeera 2014).
Management (BGİGM, Başbakanlık Güvenlik İşleri Genel Müdürlüğü) was created to coordinate litigation for censorship initiated by government agencies. The designated court for Internet crimes is the Penal Judgeship of Peace (Sulh Ceza Hakimliği). According to Law #5651, the court must make its decision within 24 hours; if the court accepts the claim, the decision is sent to the Telecommunication and Communication Agency (TIB, or Telekomunikasyon ve İletisim Başkanlığı), which is known as the Internet watchdog of Turkey [76]. TIB is responsible for faxing court decisions to the company headquarters of the social media network in question†(such as Twitter Inc.). If the company declines to censor the post, TIB may choose to block the entire website permanently or temporarily. This has happened on multiple occasions in Turkey, including on July 22, 2015, when Twitter was temporarily blocked after a bombing in Ankara that killed 32 people. Twitter occasionally shares these court rulings on the Lumen Database, previously known as the Chilling Effects website https://lumendatabase.org/

4.4 Data

To test the above theories on social media censorship we utilize three types of data. First, we obtained public opinion data in order to detect what are the most important problems perceived by Turkish citizens. Second, we collected a data on nationwide popular street protests in Turkey. Third, we used our Twitter data, which includes censored and uncensored tweets between 2014 and 2015. We test the collective action potential theory using the Twitter data and the street protests data, while we test the theory on most important

† After the coup attempt of July 15, 2016, the government abolished the TIB entirely.
problems using the public opinion data and the Twitter data. The following sub-sections briefly describe each of the data.

4.4.1 Data on Citizens’ Top Priorities

To measure Turkish citizens’ priorities regarding public policies and issues, we rely on public opinion data collected by Socio-Political Tendencies Research, which has been conducting the same survey since 2008. This survey includes questions on “the most important problems facing the country”, which are routinely used to gauge what citizens care most about and how public concerns change over time in many different contexts [65, 68, 66, 70]. We are particularly interested in the years 2014 and 2015 due to the range of our Twitter data collection. The public opinion data show that overwhelming majority of the citizens selected one of the macroeconomic problems categories as the most important issue in Turkey every year since the survey was first implemented. These categories are unemployment, inflation, economic crises, and income inequality. For example, in 2014, 58.2% of the respondents selected one of the above categories as the most important issue. Another highly important problem perceived by the citizens in 2014 was corruption with 14.2% response rate. This is likely due to the corruption scandal, which many commentators identify as the biggest in Turkish history [77, 78], that broke out in

---

‡ Conducted annually since 2008 by Kadir Has University in Turkey, the survey includes face-to-face interviews with about 1000 respondents, who are older than age of 18. For more information, see: http://www.khas.edu.tr/news/1119

§ Pew Global Attitudes Project 2014 survey also show that macroeconomic problems were the top concern while the corruption issue is the second top concern for Turkish citizens in 2014.
late 2013 and the associated probe against government officials, their family members, and businessmen [79, 78]. The survey also showed that the remaining issues are minor as none of them received attention of 5% of the respondents. In 2015, the total percentage of responses for the macroeconomic problems was 57.7. The next most important problem in the same survey turned out to be the terrorism/Kurdish issue with response rate of 13.9%[4]. This is partly because of the escalating war between the Kurds and the ISIS in northern Syria near the Turkish border and the reaction of the Kurdish separatist groups in Turkey to the government’s Syria policies. Again, the survey showed that no other issue category received 5% of the respondents’ attention or more. In summary, between 2014 and 2015 the most important concerns of the public were the economy (2014 and 2015), corruption (2014), and terrorism/Kurdish issue (2015). However, we need to note that macroeconomic problems have been almost always found as the most important problem across public opinion surveys around the world (See [80, 65, 68, 70]). Hence we assume citizens may have natural tendency to select this category. Therefore, governments may be less likely to care for this particular category for censorship if there is no serious financial crisis, whereas they would be more likely to censor posts that refer to the other most important issues such as corruption and terrorism/Kurdish issues.

[4] Many other public opinion survey projects support this finding. For example see the survey done by the IPSOS Social Researches Institute for 2015. During March and April of 2015, terrorism/Kurdish issue was the second top priority of the public, while the macroeconomic problems were again the leading issue. The other top issues were refugees, democracy, and the laws. Corruption was not a top issue at all for 2015.
4.4.2 Street Protests Data

To collect data on nationwide popular street protest, we first scrutinized the first pages of the Turkish national newspapers for the periods of our first and second datasets: October 15, 2014 to January 15, 2015 and June 3, 2015 to June 26, 2015. We found that there are 29 street protests during the first period and 4 street protests during the second period. Next we created a list of keywords that are associated with each of these collective action events (See Appendix A.2 for the list of street protests, the associated keywords for each protest and the volume of tweets for each protest type). We then searched our entire database for tweets containing any of these keywords, and estimated how many are censored and how many are not.

4.4.3 Twitter Data

In 2012, Twitter launched a new tool to allow governments to request that tweets or entire accounts be censored within the geographical boundaries of specific countries [55]. If Twitter complies with a request to withhold a tweet, the tweet becomes greyed-out while it remains visible in the rest of the world, the content becomes censored when viewed from inside of the censoring country[4]. Similarly, an entire user account can be withheld, resulting in the censorship of the user’s entire timeline and all future tweets. Following Tanash et al.’s (2015) [55] crawling methodology, we collected an additional 5 million tweets from Turkey. Combining the new data with Tanash et al.’s (2015), we analyzed two time intervals, obtaining two distinct datasets. The first dataset was collected from Octo-

[4]Throughout this paper we use “to censor” and “to withhold” as synonymous.
ber 15, 2014 until January 15, 2015 using the Twitter streaming API, the geo-bounding box parameters for Ankara, Izmir, and Istanbul, and the user screen-names parameters for users we found on the ChillingEffects.com website. For the second dataset, which was collected from June 3, 2015 to June 26, 2015, we ran a streamer, however this time with only the geo-bounding box parameters. For each dataset we use the REST API with the “id” parameters to revisit the same tweets from the Twitter server after several weeks by providing the tweets’ ids as input parameters. We repeatedly revisit these tweets hoping to observe censorship events. Censorship is readily detected since Twitter helpfully provides a “withheld_in_countries” field in the tweet data block, identifying a list of countries where the tweet should not be shown. If we detect this field, we know the tweet is censored in Turkey (Tanash et al. 2015) [55].

We also used the Twitter’s REST API to crawl users’ timelines for those with censored tweets found in our streamed dataset, their friends, and users who originally retweeted them. Overall we identified 712,218 unique censored tweets including retweets, i.e., each having a unique tweet id, of which 535,759 had unique strings[**] 96% of the posts were retweets (containing a retweeted_status field) and were associated with 13,056 distinct users. These numbers exclude tweets posted by 258 Turkish “withheld account” because we could not distinguish which individual tweets were selected for censorship and thus could not easily include them in our subsequent clustering and topic analysis. We also ignore tweets that were withheld due to copyright infringement claims (e.g., DMCA com-

[**We followed a straightforward “deduplication” process, first deleting retweet (“RT”) notations and shortened URLs.]}
plaints††). We implemented a series of data pre-processing techniques in order to make the data ready for quantitative text analysis. This process includes de-duplication of the tweets that appeared multiple times such as the case of retweets, which we also previous explained in 2.5.3 and stemming the words. (Also see the Appendix A.3 for the details)

4.5 Testing Collective Action Theory

We first test the collective-action potential theory proposed by King et al. (2013) [60]. They posit that Chinese censorship policies mostly target social media posts with a potential for collective action in the form of street protests. More specifically they argue that the overwhelming majority of social media posts referring to a collective action event are censored, particularly around the date of protest. Using the data on street protests described above, we estimated the total number and percentages for the withheld and non-withheld tweets using collective action related keywords during the period of our datasets. We present the results in Table 4.1 which shows that only 8.33% of the tweets using the protest-related keywords were withheld from Oct 15, 2014 to January 15, 2015, while only 2.95% of all tweets using the relevant keywords were withheld from June 3, 2015 to June 26, 2015. This is contrary to the expectation based on the collective-action potential theory.

In Figure 4.1 we further investigate the percentage of withheld versus non-withheld

††When a tweet is withheld due to copyright infringements, the tweet’s JSON data block will contain a withheld_copyright field; the withheld_in_countries field then contains the value “XY” rather than the ISO country code such as “TR”. For additional details see
tweets across days during the period of our second dataset, (see Appendix A.1 for the graphs belonging to the first period) which represents protests in the month of June from our second dataset. In each graph the y-axis illustrates the number of tweets and the x-axis illustrates the days. The first graph is based on the data for the workers’ protest. The number of uncensored tweets roughly ranges from 500 to 1,000 daily while the number of censored tweets is mostly below 50. This pattern is robust even around the protest date, which was illustrated by the vertical line. The next three graphs in Figure 4.1 presents similar results for the women’s protest, miners’ protest, and Gezi Park anniversary protest respectively. In all of the graphs the volume of censored tweets is substantively smaller than the volume of uncensored tweets on the same topic. All these results suggest that Turkish censors are not interested in stifling social media conversations on street protests.

<table>
<thead>
<tr>
<th>Data Range</th>
<th>Number of Protests</th>
<th>Withheld tweets with keyword</th>
<th>Non-withheld tweets with keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Oct 2014-</td>
<td>29</td>
<td>8,901 (8.33%)</td>
<td>97,992 (91.67%)</td>
</tr>
<tr>
<td>15 Jan 2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Jun 2015-</td>
<td>4</td>
<td>2,198 (2.95%)</td>
<td>72,200 (97.05%)</td>
</tr>
<tr>
<td>23 Jun 2015</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Withheld vs. non-withheld Tweets on street protests.
Figure 4.1: Daily comparison of withheld vs. non-withheld tweets.

4.6 Testing the Sensitive Issues Hypothesis using Topic Modeling

The above analysis showed that the Turkish government does not have a special aim to stifle social media conversations of street protests. But, then, what is the aim of the government censorship in Turkey, and what differentiates the censored tweets from the uncensored ones? To respond to these questions, we employed tf-idf and NMF for topic clustering [81].

We applied the topic-clustering algorithms on both censored and uncensored tweets from both the first and the second datasets, and set the number of topics to be extracted \( n \) to 10 topics. Following Fu et al. [4], we limit our analysis only to users whose tweets are partially censored. For every tweet, we compared its id with all withheld ids in our
nearly one million withheld tweets database. If there is no match, the tweet is regarded as non-withheld. It is important to note that a withheld status is temporary and reversible on the Twitter server side. Not only will new tweets become withheld but withheld tweets can potentially later become un-withheld. The changing nature of tweet statuses poses a challenge for categorizing tweets, so we decided to use all the information we have up to the end point of our data collection. In other words, users whose accounts were entirely withheld in Turkey or users who has zero censored tweets are excluded from this analysis. In total, our dataset contains 4,319 Twitter accounts in Turkey with censored tweets. The following section discusses the findings of this analysis. (See Appendix A.3 for details)

4.6.1 Results

Tables 4.2 and 4.3 present the results of the NMF analysis for the first dataset from both the non-withheld tweets and the withheld tweets (tweets from October 15, 2014 to January 15, 2015), and Table 4.4 and Table 4.5 present that for the second dataset (tweets from June 3, 2015 to June 26, 2015).

The first column in Table 4.2 shows the popular topics from the non-withheld tweets. The list includes several phrases from daily language (such as good morning, thanks, to you as well, Sunday, life), keywords on Sufism in Islam (such as Rumi, Şemsi Tebrizi, Hussein, Ali, Fatima, Omar, Allah, Muhammad), hashtags with nationalistic contents (such as Cyprus belongs to Turks and it will remain belonging to Turks, being Turkish is an honor, If you concern your nation you should be with the Nationalistic Movement Party) and finally list of major districts/provinces in Turkey (Ankara, Bursa, İstanbul, Beşiktas, Kadıköy). The set of topics for the withheld tweets in Table 4.3 are substantially different
<table>
<thead>
<tr>
<th>Topic no</th>
<th>Most Salient Topics from Non-withheld Tweets in the First Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic #0</td>
<td>throw ‘step down’ Turkic İstanbul İzmir new cafe İstanbul Ankara Kocaeli ‘KP everywhere’ amp Kadıköy restaurant İzmit coffee Bursa Buca markafoni Beşiktaş</td>
</tr>
<tr>
<td>Topic #1</td>
<td>be no day pity holy unity passed power health happy beautiful island feast nation prayer thing name republic pleased ally</td>
</tr>
<tr>
<td>Topic #2</td>
<td>day ‘good morning’ happy beautiful follow week ally everybody beneficial morning language thanks Sunday life Aydın last ‘to you as well’ young tranquility</td>
</tr>
<tr>
<td>Topic #3</td>
<td>Blacksheep media Hurriyet broadcast ban corruption Euphrates Tigris silent Turks Cyprus bastard PKK dishonored dogs martyrs parliament miners dying MHP</td>
</tr>
<tr>
<td>Topic #4</td>
<td>good night evening ally day look beneficial morning everybody week bad ‘to you as well’ life happy thanks know sleep how</td>
</tr>
<tr>
<td>Topic #5</td>
<td>love not know thanks none think here understand do go write explain brother ‘like that’ time one ally see</td>
</tr>
<tr>
<td>Topic #6</td>
<td>love think look see do understand stay how go forget pray one write burn take come</td>
</tr>
<tr>
<td>Topic #7</td>
<td>very good correct love thing rose see thanks gratitude yes day importance little human tell country net weather speak sweet life</td>
</tr>
<tr>
<td>Topic #8</td>
<td>do Allah give for not come absent do human do is it take time look AKP country CHP work day eat</td>
</tr>
<tr>
<td>Topic #9</td>
<td>blessed God Muhammad Ali Allah Omar Rumi Rumi pbuh AbuBakr Rumi Husain Fatima deed oh ŞemsiTebrizi Muhammad prophet</td>
</tr>
</tbody>
</table>

Table 4.2 : Most salient topics from non-withheld tweets in the first dataset
<table>
<thead>
<tr>
<th>Topic no</th>
<th>Most Salient Topics from Withheld Tweets in the First Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic #0</td>
<td>be no necessary human know thing pray island see time against how ‘like this’ country relaxed only big day</td>
</tr>
<tr>
<td>Topic #1</td>
<td>‘good morning’ moonshine day ally beneficial happy brother morning life beautiful good Chief halvah eat AKsiğr citizen sister Aydın steal</td>
</tr>
<tr>
<td>Topic #2</td>
<td>Allah Quran blessed prophet polytheist Sheikh meditation word servant slander Islam believe hadith verse absent invitation fabricated do Mahmud</td>
</tr>
<tr>
<td>Topic #3</td>
<td>do operation work situation palace monkey necessary know steal explain corrupt news is it action work for meeting home</td>
</tr>
<tr>
<td>Topic #4</td>
<td>thief Tayyos killer AKP steal traitor ban AK all very blessed catch corrupt rob bigot religion forget nail Tayyip leave</td>
</tr>
<tr>
<td>Topic #5</td>
<td>vulg polytheist love understand die think boy dick age go know very one rose beautiful dynasty home</td>
</tr>
<tr>
<td>Topic #6</td>
<td>is it religion fog ask not ‘like that’ send pervert police stotrial is it possible understand none Muhammad island necessary not for hadith Islam think thing (originally in English) anonymous Charliehebdo Turkey Ferguson Twitter Kurds</td>
</tr>
<tr>
<td>Topic #7</td>
<td>human rights wiki leaks the Erdoğan be in Paris Shooting Corruption to Redhack of is ftp FBI new</td>
</tr>
<tr>
<td>Topic #8</td>
<td>do for give absent do Turki new Turkish take very AKP come state PKK police last work Erdoğan continue</td>
</tr>
<tr>
<td>Topic #9</td>
<td>look good is it Erdoğan see come read how write night let’s beautiful this one work period country eye evening</td>
</tr>
</tbody>
</table>

Table 4.3 : Most salient topics from withheld tweets in the first dataset
than the other set of topics. To begin with, there are quite a number of keywords using vulgar language gathered mostly in Topic #5. Additionally, there are words like “bigot” or “AKşığır” (i.e., “AKcattle”), which are used to insult supporters of the governing AK Party. Secondly, there is a set of originally English key words, which are exclusively in Topic #6 and Topic #7. Most of them refer to recent important crises that took place outside of Turkey, such as the Charlie Hebdo attack in France, the Ferguson issue in the U.S., the Paris shooting, and WikiLeaks activities. Thirdly, and most importantly, significant number of keywords refers to the corruption scandal. In this regard, certain verbs related to corruption (such as rob, steal, corrupt, thief, traitor) frequently appear together with the words related to the government such as Tayyip (middle name of President Erdoğan), Tayyoş (a rude way of saying Tayyip), Chief (Reis in Turkish, refers to the President Erdoğan), palace, AKP (the ruling party, Adalet ve Kalkınma Partisi), and AK. The frequent emphasis on the corruption related keywords within the set of censored tweets supports our earlier expectations on the government censorship strategies during which corruption is one of the most important public concerns.

Table 4.4 and Table 4.5 present the output of the topic modeling analysis for the second dataset which includes tweets from June 2015.

The results show that the top 10 most salient topics for non-censored topics are composed of mostly, again, everyday language (such as good morning, my friends, thanks), words from standard Turkish political terminology (such as party acronyms like AKP and CHP, vote, election), and the names of the major provinces/districts (such as İstanbul, İzmir, Ankara, Diyarbakır, Kadıköy, Beşiktaş etc.). None of these words are aggressive or sharp-tongued.
<table>
<thead>
<tr>
<th>Topic no</th>
<th>Most Salient Topics from Non-withheld Tweets in the Second Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic #0</td>
<td>in at Istanbul Izmir Beşiktaş Carşı Kadıköy Turkey cafe house</td>
</tr>
<tr>
<td>Topic #1</td>
<td>AKP CHP MHP coalition coalition voting public Bahçeli governing party</td>
</tr>
<tr>
<td>Topic #2</td>
<td>good morning beautiful good my friends good morning (Kurdish) day</td>
</tr>
<tr>
<td>Topic #3</td>
<td>peaceful happy nights friends</td>
</tr>
<tr>
<td>Topic #4</td>
<td>for good thanks vote not now time man human new</td>
</tr>
<tr>
<td>Topic #5</td>
<td>very beautiful thanks I do thank thing big vote such</td>
</tr>
<tr>
<td>Topic #6</td>
<td>not exist necessary other nothing thing not no longer Beşiktaş after only</td>
</tr>
<tr>
<td>Topic #7</td>
<td>HDP Diyarbakır vote long live (Kurdish) (electoral) barrier Amed CHP</td>
</tr>
<tr>
<td>Topic #8</td>
<td>election today</td>
</tr>
<tr>
<td>Topic #9</td>
<td>is it not ‘is it?’ good vote now ever to you like that real</td>
</tr>
<tr>
<td>Topic #10</td>
<td>(list of Kurdish phrases, mostly stopwords) bi ji jî li xwe min me re ez te</td>
</tr>
<tr>
<td>Topic #11</td>
<td>Turkey election new AKP ISIS Syria not Izmir party</td>
</tr>
</tbody>
</table>

Table 4.4: Most salient topics from non-withheld tweets in the second dataset
<table>
<thead>
<tr>
<th>Topic no</th>
<th>Most Salient Topics from Withheld Tweets in the Second Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic #0</td>
<td>last minute minute information situation day mn. traveller pass word</td>
</tr>
<tr>
<td>Topic #1</td>
<td>twitterkurds Syria Kurdistan Rojava (Syrian Kurdistan) ISIS YPG, Kobani Kurds Obama HPG</td>
</tr>
<tr>
<td>Topic #2</td>
<td>be human country president passed good support come</td>
</tr>
<tr>
<td>Topic #3</td>
<td>(originally in English) in the of ypg turkey to and turkish by kurdish</td>
</tr>
<tr>
<td>Topic #4</td>
<td>do police Kurdish state public child kill continue do soldier</td>
</tr>
<tr>
<td>Topic #5</td>
<td>HDP vote PKK YPG building party Demirtaş MHP attack CHP</td>
</tr>
<tr>
<td>Topic #6</td>
<td>do for Erdoğan president war peace massacre election demonstration</td>
</tr>
<tr>
<td>Topic #7</td>
<td>Cizre public neighborhood Şırnak resist walk voice delegation police</td>
</tr>
<tr>
<td>Topic #8</td>
<td>Cizre under attack forget killed civilians killing in cizre president</td>
</tr>
<tr>
<td>Topic #9</td>
<td>this old these one protest</td>
</tr>
</tbody>
</table>

Table 4.5 : Most salient topics from withheld tweets in the second dataset
When we look at the topics from withheld posts, the picture is quite different. The words directly refer to government policies on terrorism/Kurdish issue with particular attention to the ISIS attacks against the Syrian Kurds near the border and Turkish governments’ attacks against PKK in Cizre. The words associated with the government leadership (such as AKP, Erdoğan, government, president, soldier, police etc.) frequently appeared together with terms related to violence (such as massacre, kill, attack, war etc.). The words also include the names of the locations where the armed conflicts took place (such as Syria, Kobani, Rojova, Kurdistan, Cizre), and the names of the fighting groups (YPG, HPG, PKK, ISIS). The predominance of keywords related to the terrorism/Kurdish issue confirms our theoretical expectations one more time.

Consistent with the public opinion survey data on the most important problems facing the country, the Turkish censors were mostly concerned with a top public concern during June 2015.

The above analysis showed which topics are most likely to be censored. But we still do not know which users are most likely to be censored. If our initial expectations are correct we would find users who frequently tweet on sensitive topics. The following section presents our network graph analysis for detecting the most influential users and their respective communities.

### 4.7 Detecting Influential Users and Visualizing Communities of Users

We use the data-flow graph rules that we proposed in Chapter 3 Section 3.3.1 to convert the censored tweets in to a data-flow graph, which we present in Figure 4.2, we use this
graph to identify top influential users, and community-based censored topics. We found that when applying the *tf-idf* and *NMF* standard topic extraction algorithms at a community level, we were able to achieve more precise topics extraction with meaningful references, this is because users in each module/community tend to communicate with each other more often than they do with users that are members of other communities, and they share similar attributes and interests, (See Chapter [3] Section [3.3.2] for more details). Applying modularity, we identified a total of 25 communities, five of which have thousands of distinct users in them. We plot these top five communities using Gephi [www.gephi.org](http://www.gephi.org), an open source visualization tool that uses a 3D engine to display graphs in real time and detects underlying patterns in the data. We used its Atlas Force layout. A layout algorithm sets the graph shape so it is more aesthetically pleasing, Force Atlas layout sets the nodes that are connected closer and the nodes that are not connected gets pushed away. We found that Gephi is more efficient when working with large networks comparing to other open source visualization tools such as the popular networkx [https://networkx.github.io/](https://networkx.github.io/) and it provides useful features such as filtering and clustering. Figure 4.2 shows the top five communities excluding small communities with member size ranging from 386 to 2 users. Each cluster is colored with a different color: red, green, blue, purple, and yellow to provide a better visualization.

With so many users represented in the graph, we wanted a way to emphasize some users over others. We decided to highlight users as a function of their vertex out-degree, which is a reasonable proxy for a user’s influence in the social graph (See Chapter [3] Section [3.3.1] for more details). Using the visualization tool Gephi, we adjusted the vertex size

---

[https://networkx.github.io/](https://networkx.github.io/)
Figure 4.2: Top Users’ Communities of Censored Tweets, and Influential Users.
for each user to correspond with its out-degree, letting us visually spot the most influential users. Notice also that the purple community is set further away from the remaining communities in this layout. This is because the nodes in the purple community are less connected to the other communities. The red, green, yellow, and blue communities appeared tightly connected and closer to each other, which suggests that users from these communities share common attributes, such as social contacts and topics of interest.

4.7.1 Analysis of Influential Users in the Top 5 Communities

To confirm that these users are influential, we qualitatively examined their profiles. The community in purple is the largest community. The most influential users are fuatavnifuat, HARAMZADELER333, BASCALAN, csagir2015, mehtabyuceel, and TheRedHack. One important common feature of all these accounts is their active social media involvement during the corruption scandal of December 2013. For example, Fuat Avni (@fuatavnifuat) acted like a government insider leaking Erdoğan’s strategies since the scandal was uncovered. HARAMZADELER333 is the account that leaked corruption evidence such as audio recordings of conversations between politicians and businessmen [76].

The communities with red, green, and blue colors are the next three largest communities. The influential users in these communities are overwhelmingly advocates of the Kurdish movement. These include AjansaKurdi1, AjansaKurdi2, Kurd24M, Diyarbakir7, curdistani, and ROJOVA. AjansaKurdi1 and AjansaKurdi2 are the twitter accounts of a Kurdish news site (www.ajansakurdi.com). Kurd24M is also a news source. Curdistani is a pro-Kurdish user who regularly promotes the PKK and other Kurdish groups.

These results also confirm our earlier expectations on the government’s censorship. An
overwhelming majority of the users from censored tweet data belong to the communities who frequently tweet on sensitive issues like government corruption and terrorism/Kurdish issue.

The above findings provide important evidence for the validity of our expectations regarding the link between the most sensitive concerns of the public and the censorship behavior of the government. However, they do not provide any evidence of large scale censorship of tweets discussing an important sensitive issue of the macroeconomic problems, which is seen as the most important problem Turkey is facing during the period of our study. This can be explained by three possible reasons. First, as we noted above, since there may be a natural tendency among citizens to select macroeconomic problems as the most important issue across many countries around the world, governments may be less likely to perceive this as a threat for their survival particularly if there is no significant financial crisis. Secondly, perhaps, anti-government users rarely tweet about macroeconomic problems, so the government does not need to pay special attention to this issue. Thirdly, due to the existence of due process as a requirement of the censorship mechanism in Turkey, each censorship request needs to provide a reason for violation of law. Accordingly, there may well be a large volume of Turkish discussion of macroeconomic problems, but as long as they did not violate existing laws, the Government did not attempt to censor these tweets. In the corruption cases, the Government perhaps argued that these tweets (or users) have published secret court documents, or have published fabricated evidence used by the court, all of which are against current laws. For example, see a Turkish court order on censoring Fuat Avni’s Twitters account [82].

Similarly, HARAMZADELER333 and BASCALAN accounts were withheld as the
government claimed that they violate the laws on national security and individual privacy [83]. Additionally, in the terrorism/Kurdish issue, the government argued that those tweets (or users) promote terrorism and act against national interests.

4.7.2 Censored Adult Content

In addition to the top five communities, we identified one specific community with 387 nodes, shown in Figure 4.3. The graph representing this community takes the shape of a star network, in which all of the nodes in this community are directly connected to the focal node orgidee79 which acts as a hub and information source. We categorized this as a porn community of censored tweets because the content and the screen names contained sexual language. This suggests that Turkish censors also target pornographic topics and users.
4.7.3 Singleton Users

Our analysis also revealed singletons nodes, which we define as users with no communication links to other users in our dataset (e.g. retweeting, replying etc.). Although these are not influential in our graph, we found that Turkish censorship authorities still censored their tweets. We analyzed the content of the 547 censored singleton-tweets to understand the potential cause for the Turkish censorship authorities to withhold them. As depicted in Figure 4.4, these users have no links to other users in our graph. We found that the content of these tweets is mostly about certain individuals (pro- or anti-government), with the overwhelming majority containing insulting language. It is likely that people who were insulted were the ones to have initiated the court process for censorship. The largest set of tweets (233 tweets) in this category consists of tweets targeting the businessman
Aydın Doğan, who is also the owner of a large media group [84]. Although his media group recently lowered the tone of its government criticism, during and after the Gezi park protests, Dogan’s media group was one of the most outspoken groups criticizing the government. This suggests that many of these tweets were created by users who supported the government. We have also identified 116 tweets that insult members of the AKP government, such as Recep Tayyip Erdoğan (current Turkish president), Ahmet Davutoğlu (former Turkish Prime Minister), and Lutfi Elvan (former Turkish Minister of Transport, Maritime, and Communication). The remaining 198 tweets target less prominent individuals, but still use insulting language.

Figure 4.4 : Users without edges (singleton).
4.8 Summary

Social media has become an increasingly popular source of information across the world, replacing the role of traditional news sources in many cases. Additionally, the recent uprisings in the Middle East have shown that Twitter is a highly effective tool to coordinate citizens against the suppressive regimes. The consequent increase in the political importance of social media has resulted in a dramatic increase in the volume of the scholarly work on social media. In this study we examined the dynamics of censorship in transitioning democracies, which is widely neglected by the earlier studies. Previous studies on dictatorial regimes claimed that incumbents are primarily concerned with their survival and they perceive collective action events in the form of street protests as the most important threat for their survival. Therefore, such governments primarily target stifling conversations on collective action events. We believe that incumbents in semi-democracies are also concerned with their survival. However, building on the literature on electoral behavior and responsiveness in non-dictatorial regimes, we believe that incumbents in semi-democracies perceive declining popularity and approval rates as the most important threat for their survival. Hence such incumbents need to care about what citizens care about most, and employ censorship policies if they cannot respond to citizens concerns using more democratic ways. We hypothesized that these incumbents scrutinize what are the important problems that citizens care about most and shape their censorship strategies based on those concerns. Using public opinion data on the most important problem a country is facing, we found that during the 2014-2015 period, corruption and the terrorism/Kurdish issues were among the top priorities of the citizens in Turkey. Our analysis of
the censored tweets shows that the overwhelming majority of the censored tweets are on these topics, while only a very small portion of the tweets that refer to the street protests were censored.
Chapter 5

The Decline of Social Media Censorship and the Rise of Self-Censorship after the 2016 Failed Turkish Coup

5.1 Introduction

The short lived Turkish military coup attempt on July 15, 2016, left the streets of Ankara in distress, with over 265 people killed [85]. The Turkish president, Recep Tayyip Erdoğan, was quick to blame Fethullah Gülen and his followers for the coup-plotter [14]. Fethullah Gülen is a former political figure who was previously close to President Erdoğan [86], and the founder of the Islamic movement Hizmet (a.k.a., the Gülen movement) [87], also known for his support for science, and thousands of schools and universities around the world [88]. Amid the coup, 140 journalists were arrested over ties to the Gülen-movement, 29 news outlets were shut down despite some being anti-Gülenist, and 21,000 academics were fired [89, 90, 91].

This coup is not unique in Turkish history. Five other coups were carried between 1960 and 1997 [92, 93]. Although self-censorship and censorship of the press have become commonplace following previous Turkish coups [94], self-censorship of social media is a new twist on an old story. Many Turkish Twitter users started self-censoring their Twitter accounts by switching to protected mode, presumably to avoid punishment for the public expression of prohibited content [14]. Considering this coup as a turning point in Turkish
politics, we expect its effect to be reflected in our data. Our study is first to examine and quantify users’ self-censorship of Twitter, using a systematic approach to label and process millions of Turkish tweets and users accounts. Additionally, we empirically analyze the coup’s impact on government-censored topics, and the volume of government censorship, by comparing posts after the coup to posts we collected before the coup during the Turkish general election of 2015.

In our work, we try to answer the following questions:

1. Because of the political instability in Turkey after the failed coup, and the fear of government punishment or retribution, will we detect fewer government-censored tweets generated from inside of Turkey due to users’ self-censorship?

2. Are there new government-censored topics post-coup? If so, are they related to the coup?

3. Can we systematically classify user accounts in our dataset based on self or government censorship?

5.2 Related work: Censorship in Turkey

Tanash et al. [55] detected government-censored tweets by scraping Twitter for tweets containing a withheld_in_countries field in Twitter’s API responses. They notably showed an order of magnitude higher censorship than disclosed in Twitter’s “Transparency Reports.”

Tanash et al. [95] also proposed methods for precisely extracting topics from censored Twitter data, using metrics on the social graph (e.g., user out-degree) as well as standard
machine learning topic clustering algorithms. Applying their methods to censored tweets collected between late-2014 and mid-2015, they found that the Turkish government’s most censored topics related to Kurdish issues and government corruption. In our work, we borrow from these methodologies to collect and identify censored tweets, then extract community based topics.

There are no known studies of social media self-censorship in Turkey, aside from some news reports [14], however, Arsan [94] studied censorship and self-censorship of the Turkish press. In his survey, he found that 96% of respondents agreed that journalists apply censorship when reporting on general interest topics, mainly due to internal pressure, and media owners’ financial interest. As a result, 55% of the time, journalist do not convey important information that concern the general public.

5.3 Data

Twitter’s public APIs allow external crawlers to follow specific users, search within geographic constraints, follow popular hashtags, etc. While Twitter makes it difficult to extract a full feed of every tweet, a more constrained search, such as in our work in Turkey, allows us to gather something much closer to a complete sample.

Pre-coup data: Previously, we collected 5.6M tweets during the June 2015 Turkish general election, for a duration of 24 days starting on 3 June 2015 with geo-parameters set to three major provinces in Turkey: Ankara, Istanbul, and Izmir. Using methodology similar to Tanash et al. [55], we expanded our sample by identifying users with at least one censored tweet who are not withheld-accounts, then crawling their timelines and their
friends’ timelines. Our overall sample contains 513,719 censored tweets.

Post-coup data: We streamed Twitter using the same geo-parameters collecting roughly 8.5M tweets posted by 342,650 distinct users across for 75 days. We began collecting this data roughly three hours after hearing news of the coup attempt on 15 July 2016.

5.4 Hypotheses

In this section, we propose several hypotheses concerning Turkish Twitter censorship.

5.4.1 Censorship Size

Hypothesis 1: The dynamics of censorship shifted after the failed coup. We hypothesize that we will detect less censored tweets generated from inside of Turkey, perhaps because users will be less likely to tweet “sensitive” topics.

Coups, by their nature, can be violent affairs. We would expect people to fear government retribution for tweeting opinions contrary to the ruling party. Therefore, many users might then switch their Twitter accounts to private mode, or deleted tweets or accounts entirely, resulting in a reduced volume of public tweets, and thus less content to be censored by the government.

Prior to the coup in 2015, we detected 513,719 censored tweets from non-withheld users by processing 5.6M tweets collected from Turkey. To test Hypothesis 1, we applied the same methods to our 8.5M tweets collected post-coup, and finding 142,492 distinct censored tweets from non-withheld users, which is 72% fewer censored tweets post-
coup, compared to pre-coup. Does this mean that people are tweeting less? Note that
the 5,644,284 pre-coup tweets were streamed over 24 days, compared to the 8,543,856
post-coup tweets streamed over 75 days, as shown in Figure 5.1. Normalizing the counts
into 24-days bins, we see 51% fewer streamed tweets during the first 24 days post-coup,
and an estimated 43% decline in the overall collection relative to the Twitter volume we
observed in 2015. We also noticed an incremental decline in the streamed volume in each
consecutive periods; 2% decline in August, and 5% in September.

![Figure 5.1: Volume of streamed Tweets post-coup 2015 vs. pre-coup 2016, 24 days bins.](image)

Investigating possible causes for this decline, we ruled-out the possibility of Twitter
being throttled by the Turkish ISPs [96, 97, 98]. Network-level throttling apparently oc-
curred during the initial days of the coup. Such network censorship, as well as the use of
Tor to work around it, can be observed with Tor’s “Directly Connecting Users Tor metric”\(^*\) presented in Figures 5.2 and 5.3. These graphs show the volume of Turkish Tor users before and after the coup, with a clear *decline* in Tor usage afterward. This data is more consistent with users curtailing their use of Tor and Twitter, than with being thwarted from using them.

![Directly connecting users from Turkey](https://metrics.torproject.org)

**Figure 5.2**: Daily Tor connection in Turkey - July 2016.

\(^*\)Tor Metrics: [https://metrics.torproject.org](https://metrics.torproject.org)
Figure 5.3: Daily Tor connection in Turkey - June 2015.

Perhaps there was more Twitter data that we just did not see? In our collection, we identified only one rate-limited tweet\(^\dagger\) in our post-coup data. As best we can tell, we collected virtually every Turkish Twitter post from our post-coup time period.

Deutsche Welle \cite{14} reported that the government made several arrests over social media posts praising the coup, and that “people are scared,” and are “self-censuring themselves,” fearing government punishment. Is the decline in our streamed data caused by users’ self-censorship? To test this, we must classify all 342,650 users accounts in our post-coup data as either protected, deleted, suspended, or active with deleted-tweets.

1. We first identify tweets in our post-coup data that are no longer public. We do this

\(^\dagger\)Twitter includes a special *limit* field in its streamed JSON data that indicates the number of skipped tweets.
by revisiting each tweet using the REST API‡. From 8,543,856 tweets, we found that 19% of tweets, from 139,656 users, are no longer reachable by the API. This means that these tweets are either individually deleted, or their author’s account is protected, deleted, or suspended.

2. Next, we label the 139,656 users accounts. We first check the status-code value by requesting \texttt{twitter.com/intent/user?user_id=xxx} if the code is 400, the account is deleted by the user, otherwise we call the REST API by user-id, determining if the account is suspended, or active-protected.

Our results are summarized in Table 5.1. We found that 41% of users in our post-coup data, excluding suspended users, voluntarily removed 18% of all tweets either by switching their accounts to protected, deleting tweets, or deleting their accounts entirely.§ Note that the largest set are “Active users” with some deleted tweets, followed by “Protected.”

‡If requesting a tweet by id from the REST API returns an empty string or error, it means that the tweet is no longer reachable.

§Twitter deletes accounts after 6 months of inactivity (https://support.twitter.com/articles/15362). Our initial and follow-on samples were at most five months apart, avoiding this issue.
<table>
<thead>
<tr>
<th></th>
<th>Tweets</th>
<th>%</th>
<th>Users</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total streamed tweets</strong></td>
<td>8,543,856</td>
<td>-</td>
<td>342,650</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total unreachable tweets</strong></td>
<td>1,582,632</td>
<td>19%</td>
<td>139,656</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Suspended users</strong></td>
<td>29,971</td>
<td>2%</td>
<td>614</td>
<td>0.4%</td>
</tr>
<tr>
<td><strong>Deleted users</strong></td>
<td>200,523</td>
<td>13%</td>
<td>6,893</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Protected users</strong></td>
<td>662,319</td>
<td>42%</td>
<td>30,612</td>
<td>22%</td>
</tr>
<tr>
<td><strong>Active users with deleted tweets</strong></td>
<td>689,819</td>
<td>44%</td>
<td>101,537</td>
<td>73%</td>
</tr>
<tr>
<td><strong>Changed by user</strong></td>
<td>1,552,661</td>
<td>18%</td>
<td>139,042</td>
<td>40.6%</td>
</tr>
</tbody>
</table>

Table 5.1: Post-coup user-accounts labeling results
We conclude that there are fewer government-censored tweets from inside of Turkey after the coup. This finding is consistent with Turkish people being afraid to speak in public, and consequently taking steps to hide prior speech and self-censor future speech.

### 5.4.2 Censored Topics

In this section, we examine which topics are censored by the Turkish government post-coup. We start by proposing the following hypothesis.

**Hypothesis 2:** Censored tweets post-coup will contain new topics related to the coup.

Tanash et. al [95] established that the Turkish government censored two main topics; government corruption and Kurdish issues. To identify newer topics, we convert our 2016 censored tweets to a *Data-Flow-Influence* graph (as in Tanash et al. [95]), representing data communication between users. User influence is a function of nodes highest out-degree. Applying a graph modularity metric, we extract user’s communities, then manually examine the top influential users from each community. Figure [5.4](#) illustrates the results of users communities and users influence.
Figure 5.4: Users communities and influential users of censored tweets post-coup.

Figure 5.4 shows two users’ clusters. We asked a native Turkish speaker to manually examine the top influential users from each cluster. We found that the left clusters’ (pink and red) influential users are mainly Gülen supporters, such as “Yagizefe”, and the right cluster’s influential users are Kurdish users, such as “ANF_TURKCE_ANF”. To extract popular topics from each cluster, we first apply several steps to remove noise from our data and achieve more precise topics extraction, as in Tanash et al. [95]. This includes removing tweets from users with bot-like behavior, stripping URLs and stop-words, and then using the widely-used zemberek Turkish-language stemmer [99]. After this data tuning, we apply tf-idf and NMF [81], both standard machine learning algorithms for topic retrieval from documents and tweets [100] [55].
Our results show that the topics from each community cluster correspond to the profiles of the top influential users. The cluster with the Gülen supporters contained topics with Islamic references, references to popular Twitter Gülen supporters and Erdoğan critics, as well as specific terms that are clearly anti-government, e.g., “torture”, “human rights”, “dictator”, “arrests”, “military”, “Erdogan soldiers” (originally Turkish, shown here in English). Our other cluster is clearly Kurdish activists. Several topics contained the string “kurd”, references to popular Kurdish Twitter usernames, including “red hack”: a hactivist group. We similarly saw words like “guns” and “fascism”.

The Red Hack group of hactivists has drawn a lot of government attention. They’re clearly visible in Figure 5.4 as “theRedHack”. In September 2016, the Turkish government blocked Github, Google Drive, and other content sharing sites to specifically prevent the Red Hack group from leaking government secrets [101], which they threatened to do if Turkey did not release imprisoned Kurdish politicians [76].

In summary, our findings confirm our second hypothesis. The Gülen movement was not a topic of censorship prior to the coup, but afterward became a clear focus of Turkish government attention, both online and offline.

5.4.3 Self-Censorship of Gülen Topics

Turkish police began arresting people with ties to the Gülen movement, including social media posts praising the coup. Naturally, to protect themselves, we would expect people to both self-censor their old postings as well as make strong public performances of their loyalty [14]:
**Hypothesis 3**: Pro-Gülen tweets from before the coup will be self-censored.

**Hypothesis 4**: Public anti-Gülen tweets will occur more often post-coup.

To test these, we conducted two experiments:

1. **Topic Clustering**: We first identified Gülen-related tweets using a bag-of-words approach against both the public and the unreachable tweets in our 2016 sample, excluding users who are suspended by Twitter. The results are summarized in Table 5.2. Notably, the unreachable rate for Gülen-related tweets is twice the background rate. Next, we extract popular topics from each set using **tf-idf** and **NMF**, asking a native Turkish speaker to label each topic’s sentiment as either pro-Gülen, or anti-Gülen. We found zero pro-Gülen topics in the public tweets, supporting Hypothesis 3. Conversely, we found 70% of the unreachable Gülen tweets were pro-Gülen, supporting Hypothesis 4.

<table>
<thead>
<tr>
<th>Status</th>
<th>Unreachable</th>
<th>Public</th>
<th>Unreachable%</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tweets</td>
<td>1,582,632</td>
<td>6,961,224</td>
<td>23%</td>
</tr>
<tr>
<td>Gülen tweets</td>
<td>3,599</td>
<td>25,538</td>
<td>14%</td>
</tr>
</tbody>
</table>

Table 5.2: Gülen-related vs. all tweets, collected post-coup, counting how many remain public vs. unreachable.

2. **Sample Labeling**: We randomly sampled 40 tweets from each set, exclusively from the first three days post-coup, expecting to find more political content closer to the coup event. We asked a native Turkish speaker to classify the tweets as pro-Gülen, anti-Gülen, or
neutral/unrelated. We summarize the results in Figure 5.5. As above, we found support for Hypothesis 3: pro-Gülen tweets appear **only** in the unreachable set, with none in the public set. We also found 40% more anti-Gülen tweets in the public set, supporting Hypothesis 4.

![Figure 5.5: Labeling of Gülen related Tweets from public vs. unreachable tweets.](image)

At this point, we have strong support for Hypotheses 3 and 4. What is less clear is whether some of the anti-Gülen discussion is “natural” or is a sort of public performance intended to defend the poster against otherwise baseless pro-Gülen accusations.
5.4.4 Self-Censoring Users

Self-censoring users are clearly responding to an external stimulus. They don’t want to get caught up in the Turkish government’s anti-Gülen witch-hunt, so it’s sensible that they would take steps to polish their public Twitter timeline. We might expect this sort of self-censorship to be performed broadly across all Twitter users, as opposed to just being the province of dedicated activists, who presumably would have a harder time hiding obscuring their political speech, whether online or elsewhere. This lead us to our final hypothesis.

**Hypothesis 5**: Self-censoring users’ tweets are more likely to be politically neutral, for the most part, with only a few “sensitive” tweets.

To test this hypothesis, we extract popular topics from all tweets posted by users who voluntarily changed their profiles status, or removed some of their old tweets. Examining the top 10 topics, we found that 9 of 10 topics contained neutral language, such as “I love it very much”. The only political topic contained two political terms: “soldier” and “democracy”, which clearly relate to tweets posted during the first week after the coup. Repeating the same experiment using only tweets from the active-users set who deleted some tweets, while keeping their profiles public, accounting for 44% of total unreachable tweet, we found similar results.

Next, we conducted a comparative analysis to quantify political and non-political tweets per user for each user category, and found that on average, active, protected, and deleted users, tweeted only 6%, 5%, and 6% political tweets, respectively. Figures 5.6
show this distribution for the top 200 users sorted based on the highest number of political tweets a user generated in our dataset for each of the user groups. All of these graphs demonstrate that these users largely do not discuss Gülen or other political topics, supporting Hypothesis 5.

![Graph showing political tweets distribution](image)

Figure 5.6: Percent of political tweets of the top 200 users, from the highest percent to the lowest

### 5.5 Twitter Transparency Report

In 2015, Tanash et al. [55] observed one order of magnitude more censored tweets than the data reported by Twitter’s Transparency Reports. Similarly, we tried to compare the number of censored tweet in our post-coup dataset to the number reported in Twitter’s
July 1, 2016 – December 31, 2016 Transparency Report [102], in which Twitter reported 489 censored tweets in Turkey from non-withheld accounts. Our 142,492 censored post-coup tweets were posted between July 15 through November 1, 2016 from non-withheld accounts, of which 96% are retweets, as reported by Twitter’s JSON metadata. Deduplicating these retweets, we identified 6,402 unique censored tweets, which contrasts with the 489 tweets reported by Twitter. As with Tanash, we find an order of magnitude more censored tweets than Twitter reports. Consequently, we caution other researchers from treating Twitter’s reporting as a reliable source.

5.6 Summary

The 2016 coup attempt in Turkey provided us with an unusual opportunity to measure the impact of a singular event like this on both government-driven censorship as well as self-censorship. We were able to compare a dataset of 8.5 million tweets, collected in 2015, with a new dataset of 5.6 million tweets, collected in the immediate aftermath of the coup. Our data shows clear evidence that users are self-censoring their post-coup posts, particularly anything they might have said positively about Gülen, accused by the government of masterminding the coup. Similarly, they are limiting what they write going forward, with some evidence of users even deliberately writing pro-Gülen tweets, perhaps as a public performance of loyalty to the government.
Chapter 6

The State of Democracy in Turkey

6.1 Freedom of Speech in Turkish History

According to Gareth Jenkins, a Turkish analyst \[103\], and the data published by the Freedom House (FH) in 2014 \[78\], suppression of speech existed in Turkey long before the Justice and Development Party (AKP) came to power, and that as a result of the 1960-1980 coups, Turkey was under a secular military “guardianship”, which strictly prohibited discussions of “ethnical and religious identities” to combat the Islamist Welfare Party. Jenkins believes that this effort was also supported by the main Turkish media, resulting in a poor state of freedom of speech in the region as well as media self-censorship. When the AKP party ran for election in 2002, they promised to reform old martial laws to improve the declined state of democracy and relax restrictions on freedom of speech \[78\]. Desperately seeking change, the majority of people voted for AKP in the 2002 general election. Many liberals and minorities who were interviewed by FH attest that the AKP brought notable improvement during their first presidential term compared to the military guardianship with regards to the rights of the Kurdish minorities, Islamic religion rights, and even speaking of the Armenian genocide \[78\].

However, despite the AKP’s claim to advancing democracy, the party’s promises insidiously started to collapsed as they consolidated their power and became less tolerable
to criticisms that could threaten their power. This became evident in the numerous arrests of Kurdish politicians, journalist, and activist who are critical of the government. These broad and sweeping arrests raised concerns with human rights organizations [103], such as Reporters Without Boarders∗ (RSF), a French based non-profit organization that monitors and tracks freedom of information on the Internet and censorship, as well as abuse cases of journalists and the media globally [104]. RSF regularly publish a list of countries rankings to reflect the level of its freedom. According to Jenkins [103], RSF reported that Turkey’s ranking declined from 138th in 2010 to 148th in 2012 out of 179 countries that were monitored. Most recently, in 2017, Turkey’s index dropped to 155 out of 180 countries, dropping 4 places since 2016 [105]. Freedom House reported [78] that by 2010 the Turkish main media was largely owned and operated by the AKP, which often fired outspoken columnists, and used their control over the media to spread its propaganda. This was done mainly to control collective public attitude to achieve its re-election agenda and remain in power. For media that was not within its control, the AKP used economic pressure by either withholding licenses and permits, or launching tax investigations, the AKP also used reward-tactics to encourage media outlets to write in favor of the government [78]. Jenkins argues that the media is to blame for its poor state of freedom, for accepting such pressure, as a result of its lack of journalism-standards. Turkish media is tribal, and mainly consists of columnists and panelists who present opinions instead of verified facts, and are willing to present rumors and spread conspiracies to harm its opponent instead of practicing "investigative journalism" [103]. On the other hand, when

∗https://rsf.org/en
journalists are faced with the alternative of losing their jobs, they are left with no option but to comply and self-censor.

Some major events in the recent years, attributed to the drastic state of decline in Turkish democracy. First was the riot of the Gezi park in March 2013 [106]. The protest started as a peaceful demonstration to protect the green Gezi park in Istanbul, then the protest escalated and turned violent between the police and civilians and spread to another 70 cities. A report published on Gnip’s website [107] reported that Twitter gained popularity in Turkey after the Gezi event, generating over one million new Twitter accounts. This is because people felt that the information posted on Twitter was more trust-worthy than what was reported by the Turkish mainstream media. Realizing how powerful Twitter and other social media were in mobilizing people and spreading news, Erdoğan ordered blocking Twitter in Turkey a year following the Gezi riot. He used the excuse of Twitter’s rejection to implement Turkish court orders seeking removal of some links posted on Twitter. Erdoğan also asked Twitter to establish an office in Turkey to ease take down requests of tweets to hold Twitter accountable under Turkish laws, which Twitter declined [32]. The block only lasted one week, after Highest Court ruled the ban a violation of freedom of expression [33], leaving Erdoğan frustrated for failing to control social media or ban it in Turkey.

The next major event that some argue weakened AKP’s position with the general public, was the December 2014 election corruption scandal. Some independent media associated with the Gülen movement, and others owned by Doğan Media Group found their voices and decided to cover the corruption stories [78]. In retaliation, the government used economic pressure to force them to shut down, such as the case of the Milliyet paper
owned by Doğan Media Group, which was one of the largest papers in the country \cite{78}. (Note that those are the same type of media outlets targeted by AKP post the 2016 failed coup). The AKP also arrested numerous journalists who talked about the corruption scandal, in additional to many Kurdish political figures and activists. We note that we found artifacts of these actions in our previous findings, where we identified two main censored topics targeted by the Turkish government related to the corruption scandal and the Kurdish issues.

It is clear that the state of freedom of expression and journalism in Turkey remains poor even under the control of the AKP government despite their promises. On the other hand, controlling social media remains a challenge for the AKP party. Unlike social media in China, which is solely controlled by the government, popular social media in Turkey such as Twitter and Facebook are managed by western entities with no control by the Turkish government. In recent years, however, many content sharing services implemented procedures to allow governments to request certain posts to be removed or redacted in their countries such as the Twitter case we presented earlier in this work. We found that Twitter censorship in Turkey spiked in volume starting in 2014, where the number of censored tweets and accounts increased from zero in 2013, to 2,003 censored tweets and 84 censored accounts in 2014 as presented in Section \ref{sec:twitter2014} and they continue to rise in following years. The AKP also went to the extent to cancel previous appeals to reform laws which it fought for during the election in 2011 \cite{78}, and instead used the same laws to punish violators and issue Twitter removal requests under some general antiterrorism and criminal defamation laws, as we found this evident in some of the censored posts in Chapter \ref{ch:chapter2} Section \ref{sec:twitter2014} which contained vulgar language, and Kurdish and terrorism references.
6.2 Summary

We argue that winning this battle of controlling western social media in Turkey is a tough game, because posts spread quickly and exponentially, and users continue to find ways to circumvent the censors via changing their location or using VPNs. The AKP realized this challenge especially after it failed to block social media entirely in the country, resulting in a shift of its propaganda strategy to utilize fear and punishment to deter people from using social media, as we found in our post-coup 2016 results, where the volume of tweets drastically decreased by 43% over all attributed in part to users’ self-censorship. We also note that a new report by International Amnesty reports [108] that “one third” of total imprisoned journalist in the world are in Turkey, and the total number of Turkish citizens impacted by Erdoğan’s aggressive reaction to the 2016 coup is 300,000, including journalists, academic, secularists and other groups [109]. We foresee that if the state of democracy continues to declines in Turkey as the AKP gains more power, we would not be surprised if laws change in Turkey banning all social media in the country similar to Iran and China.
Chapter 7

Conclusion and Future Work

7.1 Conclusion

Scholars across a number of disciplines have been fascinated with understanding not simply how people use social media, but also how governments regulate users, content, and usage. Most existing studies have focused almost exclusively on authoritarian regimes like China and Syria. In this work, we study western social media censorship in semi-democracies, by focusing on examining the frequency of Turkey’s use of Twitter’s censorship mechanisms, and the content being censored.

Our study considers over 35M Turkish tweets that we collected between 2014 – 2016. Our work established that unlike authoritarian regimes where censors focus on social media posts that represent or reinforce collective action events, we found that social media censorship in transitioning democracies is targeted at sensitive public policy issue topics—those perceived by the public as important, but on which the ruling party was unwilling or unable to address. The 2016 coup attempt in Turkey provided us with another unusual opportunity to measure the impact of a singular event on both government-driven censorship as well as self-censorship. Our measurements following the coup show a significant decline in publicly identifiable government-censored tweets. We attribute this, in part, to the decline in overall Twitter usage in Turkey and in part to users’ self-censorship.
We also found an estimated one-order of magnitude more censorship in Turkey than what Twitter officially reported, raising the possibility that similar trends hold for censored tweets from other countries as well. Consequently, we caution other researchers from treating Twitter’s reporting as a reliable source.

7.2 Future Work

There are a variety of directions for future work related to these studies. Expanding this work to study Twitter censorship in other countries is an obvious direction, particularly in countries with no reported censorship requests. The recent rise in Twitter censorship in the United Kingdom, in tandem with the Brexit issue, particularly suggests an opportunity for comparative research of censorship between the two countries. A full country-by-country study would require significantly greater resources than we could muster from our home institution. While we certainly could scale up with modern cloud services, engineering a full-scale web crawler dedicated to downloading each and every tweet as it is posted, with periodic follow-ups to check every tweet’s withholding status, would represent a non-trivial load on Twitter’s service; Twitter would presumably then take technical and/or legal steps to block such research crawler. Alternatively, Twitter could allow researchers to access its firehose without restrictive agreements as to how they may or may not use the data.

Another direction may include conducting a large-N time series analysis to test the relationship between citizens’ concerns and governments censorship policies. Going forward, we note that social networks change in popularity over time, and Twitter may not
always represent a reliable barometer of public opinion. Twitter is valuable for conducting this sort of research because it’s easy to scrape and most content is public. Conducting similar research on Facebook or elsewhere, where user’s default security settings limit their posts’ visibility to their friends, would represent a significantly greater challenge, particularly without the social network’s cooperation. If that cooperation cannot be assured, then other tactics, ranging from browser plugins to server-side apps, with their own issues of low user adoption, may become necessary to understand and measure this sort of application-level censorship as it occurs.
Appendix A

A.1 Comparing Withheld vs. Non-Withheld Tweets

The blue solid lines in Figure A.2 and Table A.1 represent the number of non-censored tweets while the red dashed lines represent the number of censored tweets for each day. The vertical line illustrates the date of protest event. We have not displayed the data for protests with less than ten non-withheld tweets.
Figure A.1: Daily comparison of withheld vs. non-withheld tweets (Oct 21 to Dec 5)
Figure A.2: Daily comparison of withheld vs. non-withheld tweets (Dec 9 to Jan 10)

A.2 Protests and Keywords

The results in Table A.2 show that many of the events had zero censorship rates while some others (such as anti-corruption protest on January 4, 2015) have 100% censorship rate. At this point we need to note that although the number of censored tweets reflects very accu-
<table>
<thead>
<tr>
<th>Protest Type</th>
<th>Protest Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protest against mosque construction</td>
<td>validebağ, çamlıca konakları, cami inşaati, koruluğu</td>
</tr>
<tr>
<td>Anti-corruption protest</td>
<td>Adaletin kara günü, 17 Aralık, rüşvet, yolsuzluk, 17-25</td>
</tr>
<tr>
<td>Saturday mothers’ protest</td>
<td>Cumartesi anneleri</td>
</tr>
<tr>
<td>Protest against mosque construction</td>
<td>validebağ, çamlıca konakları, cami inşaati, koruluğu</td>
</tr>
<tr>
<td>Miners’ protest</td>
<td>madenci, maden faciası, maden, soma, ermenek, maden işçisi, maden kazası</td>
</tr>
<tr>
<td>Protest for solidarity with Kobane</td>
<td>İŞİD zulmū, kobane, suruç</td>
</tr>
<tr>
<td>Protest against mosque construction</td>
<td>validebağ, çamlıca konakları, cami inşaati, koruluğu</td>
</tr>
<tr>
<td>Students’ protest</td>
<td>yok protestosu, yok’ü protesto, ankara üniversitesi</td>
</tr>
<tr>
<td>Environment protest</td>
<td>HES, termik santral, çimento, siyanür, Trabzon, derelerin kardeşliği</td>
</tr>
<tr>
<td>Protest against mosque construction</td>
<td>validebağ, çamlıca konakları, cami inşaati, koruluğu</td>
</tr>
<tr>
<td>Doctors’ protest</td>
<td>sağlık bütçesi, sağlık bakanı, sağlık bakanı, bütçe tasarısı,</td>
</tr>
<tr>
<td>Workers’ protest</td>
<td>plan ve bütçe komisyonu, sağlık emekçileri</td>
</tr>
<tr>
<td>Women’s protest</td>
<td>25 kasmı, kadına şiddet, kadın erkek eşit</td>
</tr>
<tr>
<td>Workers’ protest</td>
<td>taşeron işçi, yol-ış, iznerji, tced, kesk</td>
</tr>
<tr>
<td>Miners’ protest</td>
<td>madenci, özelleştirme, yatağan, termik santral, maden-ış, tes-ış, sattırmayacağız</td>
</tr>
<tr>
<td>Hospital workers’ protest</td>
<td>maltepe üniversitesi, hastahanesi, dev sağlık iş, sendika, işçi</td>
</tr>
<tr>
<td>Government employees’ protest</td>
<td>memurlar, emekçileri, bütçe görüşmeleri, meclis önü</td>
</tr>
<tr>
<td>Doctors’ protest</td>
<td>aile hekimleri grev, acil hastaları</td>
</tr>
<tr>
<td>Soccer Fun Club solidarity protest</td>
<td>Taraftar grubu çarşı, çarşı darbe, çarşı dava</td>
</tr>
<tr>
<td>Anti-corruption protest</td>
<td>hırsız var, 17 Aralık, rüşvet, yolsuzluk, 17-25</td>
</tr>
<tr>
<td>Miners’ protest</td>
<td>hattat holding, işten atma, işten çıkarma, hema maden, gmis, ttk</td>
</tr>
<tr>
<td>Teachers’ protest</td>
<td>eğitim-ış, laik eğitim, emeğe sayıış yürüyüşü</td>
</tr>
<tr>
<td>Workers’ protest</td>
<td>birleşik metal-ış, mess, metal işçileri</td>
</tr>
<tr>
<td>Students’ protest</td>
<td>faşist saldırdı, ögrencilere saldırd</td>
</tr>
<tr>
<td>Gezi Park Solidarity protest</td>
<td>Kent savunması, Haziran Hareketi, Gezi, Birleşik Haziran, gerici eğitim</td>
</tr>
<tr>
<td>Workers’ protest</td>
<td>Ulker işçileri, 70. günü, hastahanesi işçileri, demokratik kitle</td>
</tr>
<tr>
<td>Environment protest</td>
<td>iztuzu plajı, caretta, yaşam alanı</td>
</tr>
<tr>
<td>Charlie Hebdo solidarity protest</td>
<td>Je suis Charlie, Charlie Hebdo</td>
</tr>
<tr>
<td>Protest against religious education</td>
<td>haziran hareketi, laik eğitim, gerici eğitim</td>
</tr>
<tr>
<td>Renault Factory Workers</td>
<td>Renault, işçi, direniş, metal, ego, patron, talep, işten atma, Türk metal, sendika, protesto</td>
</tr>
<tr>
<td>Women’s Movement</td>
<td>Özgecan Aslan, isyan, protesto, cinayet, dava, katil, ceza, tecavüz</td>
</tr>
<tr>
<td>Coal Miners</td>
<td>soma, ermenek, katliam, facia, maden, kazası, madenci, kömür</td>
</tr>
<tr>
<td>Gezi Park Anniversary Protests</td>
<td>haziran hareketi, protesto, gezi park, gezi parkı, görülmemiş hesap</td>
</tr>
<tr>
<td></td>
<td>kalmayacak, birleşik haziran, görülmemiş, taksim, protesto, direniş</td>
</tr>
</tbody>
</table>

Table A.1: Protests and the Associated Keywords
<table>
<thead>
<tr>
<th>Date</th>
<th>Protest Type</th>
<th>Number of withheld tweets with keyword</th>
<th>Number of non-withheld tweets with keyword</th>
<th>Percentage of withheld tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/19/14</td>
<td>Protest against mosque construction</td>
<td>16</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>10/21/14</td>
<td>Anti-corruption protest</td>
<td>2974</td>
<td>4688</td>
<td>39</td>
</tr>
<tr>
<td>10/26/14</td>
<td>Saturday mothers’ protest</td>
<td>1</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>10/28/14</td>
<td>Protest against mosque construction</td>
<td>16</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>10/30/14</td>
<td>Miners’ protest</td>
<td>534</td>
<td>18762</td>
<td>3</td>
</tr>
<tr>
<td>11/2/14</td>
<td>Protest for solidarity with Kobane</td>
<td>183</td>
<td>620</td>
<td>23</td>
</tr>
<tr>
<td>11/3/14</td>
<td>Protest against mosque construction</td>
<td>16</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>11/9/14</td>
<td>Students’ protest</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11/10/14</td>
<td>Environment protest</td>
<td>430</td>
<td>18164</td>
<td>2</td>
</tr>
<tr>
<td>11/14/14</td>
<td>Protest against mosque construction</td>
<td>16</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>11/21/14</td>
<td>Doctors’ protest</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>11/22/14</td>
<td>Workers’ protest</td>
<td>16</td>
<td>9588</td>
<td>0</td>
</tr>
<tr>
<td>11/24/14</td>
<td>Women’s protest</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>11/25/14</td>
<td>Workers’ protest</td>
<td>29</td>
<td>12377</td>
<td>0</td>
</tr>
<tr>
<td>12/2/14</td>
<td>Miners’ protest (continued 4 days)</td>
<td>484</td>
<td>7888</td>
<td>6</td>
</tr>
<tr>
<td>12/9/14</td>
<td>Hospital workers’ protest</td>
<td>243</td>
<td>1198</td>
<td>17</td>
</tr>
<tr>
<td>12/11/14</td>
<td>Government employees’ protest</td>
<td>196</td>
<td>545</td>
<td>26</td>
</tr>
<tr>
<td>12/13/14</td>
<td>Doctors’ protest</td>
<td>22</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>12/17/14</td>
<td>Soccer Fun Club solidarity protest</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12/18/14</td>
<td>Anti-corruption protest</td>
<td>3006</td>
<td>4688</td>
<td>39</td>
</tr>
<tr>
<td>12/20/14</td>
<td>Miners’ protest</td>
<td>1</td>
<td>1810</td>
<td>0</td>
</tr>
<tr>
<td>12/21/14</td>
<td>Teachers’ protest</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12/23/14</td>
<td>Workers’ protest (continued 2 days)</td>
<td>445</td>
<td>11237</td>
<td>4</td>
</tr>
<tr>
<td>12/26/14</td>
<td>Students’ protest</td>
<td>20</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>12/29/14</td>
<td>Gezi Park Solidarity protest</td>
<td>196</td>
<td>6061</td>
<td>3</td>
</tr>
<tr>
<td>1/4/15</td>
<td>Workers’ protest</td>
<td>11</td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>1/5/15</td>
<td>Environment protest</td>
<td>0</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>1/10/15</td>
<td>Charlie Hebdo solidarity protest</td>
<td>44</td>
<td>279</td>
<td>14</td>
</tr>
<tr>
<td>1/12/15</td>
<td>Protest against religious education</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Table A.2: Protests from October 15, 2014 to January 15, 2015
Table A.3 : Protests from June 11, 2015 to June 23, 2015

<table>
<thead>
<tr>
<th>Date</th>
<th>Protest Type</th>
<th>Number of withheld tweets with keyword</th>
<th>Number of non-withheld tweets with keyword</th>
<th>Percentage of withheld tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-Jun</td>
<td>Renault Factory Workers</td>
<td>457</td>
<td>16593</td>
<td>2.68</td>
</tr>
<tr>
<td>12-Jun</td>
<td>Women’s Movement</td>
<td>932</td>
<td>22805</td>
<td>3.93</td>
</tr>
<tr>
<td>15-Jun</td>
<td>Coal Miners</td>
<td>419</td>
<td>4249</td>
<td>8.98</td>
</tr>
<tr>
<td>21-Jun</td>
<td>Gezi Park Anniversary Protests</td>
<td>390</td>
<td>28553</td>
<td>1.35</td>
</tr>
</tbody>
</table>

rate scores, we believe the number of non-withheld tweets should have been even higher if we were able to collect all the tweets in the twitter database. During the time we collected this particular dataset we did not particularly interested in the non-withheld tweets. Our priority was exclusively to detect the censored tweets. Hence we downloaded and/or saved only small number of non-withheld ones. However, even in this case, we observe that the percentage of withheld tweets with collective action keywords is very small (8.33%). And it supports our expectation that tweets with collective action potential is not highly likely to be censored in semi democratic regimes. This data collection problem does not apply to our second dataset.

Table A.3 shows the comparison of the censored and uncensored tweets for our second dataset. Women’s movement protest on June 12 is the protest with the highest number of censored tweets, which is 932. However, percentage-wise the most salient censorship (with 8.98%) took place after the coal miners’ protest on June 15. June 21 protests for the Gezi park anniversary have the lowest percentage, 1.35%, and the highest number of non-withheld tweets (28553) using the associated keywords. All in all, this data shows
that government is not aiming to kill the conversation mentioning a real world collective action event.

A.3 Preprocessing the Data for Topic Modeling Analysis

Before feeding tweet texts into our topic classification model, we applied a sequence of text processing techniques to ensure validity, accuracy, and relevance of our data. Our first step is extracting original tweets from retweets. As a popular social media platform, Twitter feed consists of a large number of retweets. Indeed, in our dataset 96% of all posts are retweets. If a tweet is a retweet, its JSON structure will have a non-empty retweeted_status field. We found that the top five communities have an average of 99% retweets, and since retweets are endorsements of original tweets and add no additional information, we extracted the original embedded tweet from its JSON structure, removing “RT” (retweet) prefixes and other such metadata.

Not all “retweets” use Twitter’s standard representation, so they might have the textual structure of a retweet without the associated retweeted_status metadata. We removed all RTusername text strings that appear at the beginning of a tweet and hashed every tweet to ensure uniqueness.

At this point, our tweet texts are still not ready to be analyzed due to high variability between similar words. For example, when a human sees three words university, University and university’s, he knows that they are all related to university. For a computer however, these three words are coded differently and therefore are distinct from one another. A common text processing technique is to make all letters lowercase. Besides doing
<table>
<thead>
<tr>
<th>Tweet text shown on Twitter’s web page</th>
<th>Example One</th>
<th>Example Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example One</td>
<td>»AKPTERÖRÖRGÜTÜ«</td>
<td>RT PartizanChe: Ambulans ve sağlık ekiplerinden önce Toma geldi. Yazıklar olsun böyle ülkeye!  . #SuruçtakatliamVar <a href="http://t.co/n9P5oPHQ2">http://t.co/n9P5oPHQ2</a></td>
</tr>
<tr>
<td>Raw text, as it arrives in our system</td>
<td>&gt;&gt;&gt;&gt;&gt;&gt;AKPTERÖR ÖRGÜTÜ&lt;&lt;&lt;&lt;&lt;&lt;&lt;&lt; &lt;&lt;&lt;&lt;&lt;&lt;&lt;&lt;&lt;</td>
<td>ambulans ve sağlık ekiplerinden önce toma geldi yazıklar olsun böyle ülkeye suruçetakatliamvar</td>
</tr>
<tr>
<td>Filtered text, used for subsequent processing</td>
<td>akpterörgütü</td>
<td></td>
</tr>
</tbody>
</table>

Table A.4 : Examples of transforming tweets in preprocessing phase

this, we removed common Turkish stop words as well as all occurrences of &gt; and &lt;, which are merely encodings for “<” and “>” symbols (see Example One in Table A.4).

We removed any letters immediately after an apostrophe because they are used only for grammatical completeness. We also removed URLs appended at the end of each tweet, since we are ultimately interested in clustering tweets based on words, not URLs. As an illustration of this process, Example Two in Table A.4 shows a raw tweet text and its transformed version after data preprocessing.

### A.4 Stemming

We apply a stemming algorithm as a part of the preprocessing. Stemming is the process of removing prefixes and suffixes (also called morphemes) so as to reduce derivational differences amongst words with the same root meaning. For instance, stemming collapses the words university and universities into one word regardless of their grammatical tenses or pluralities. A variety of open source Turkish stemmers are available on the Internet,
and each has pros and cons. Packages such as snowballstemmer\(^*\) are lightweight and easily customized, but do not natively handle the Turkish language. We decide to use the zemberek-nlp [99] package because it is both widely used and well researched. One limitation of this package however, is that the program returns an empty string if there is no match for the given input word. In our Twitter texts there are names (e.g. Erdoğan) and hashtags (e.g. #AjansaKurdi) that are neither stemmable nor in the package default dictionary. As a workaround for this, we modified the package to return the original word if there is no match found.

\(^*\)Python Package. [https://pypi.python.org/pypi/snowballstemmer](https://pypi.python.org/pypi/snowballstemmer)
Bibliography


Expires=1495597797&Signature=0g4ynSb6RBsTrp%2FaVLoLJp%2BjUgA%3D&response-content-disposition=inline%3B%20filename%3DAuthoritarianism_Online_What_Can_We_Learn.pdf


http://dl.acm.org/citation.cfm?id=2663720.


http://science.sciencemag.org/content/345/6199/1251722.


https://scholar.google.com/scholar?q=Why+Parties+and+Elections+in+Authoritarian+Regimes&hl=en&as_sdt=0&as_vis=1&oi=scholart&sa=X&ved=


[89] J. Bohannon, “Turkish Academics Targeted as Government Reacts to Failed


[101] D. Pauli, “Turkey Blocks Drive, Github, OneDrive in Bid to Kill RedHack Leaks,”
The Register, 2016. https://www.theregister.co.uk/2016/10/10/turkey_blocks_drive_github_onedrive_in_bid_to_kill_redhackLeaks/.


https://www.washingtonpost.com/opinions/global-opinions/the-turkey-i-no-longer-know/2017/05/15/bda71c62-397c-11e7-8854-21f359183e8c_story.html?utm_term=.2c8a0161b997