

Spatiotemporal Analysis of Censored Content on Twitter

Onur Varol*

Center for Complex Networks and Systems Research
School of Informatics and Computing, Indiana University, Bloomington, USA

ABSTRACT

Social media have become vehicles for instantly disseminating and accessing information on a global scale. Beside such positive contributions, social media also enable malicious activities such as recruitment for terrorist groups or coordinate orchestrated campaigns. Censorship is one way of limiting user activities, but applying it fairly is not easy, as exemplified by site-blocking censorship by governments. To avoid complete site-blocking, some social media sites have complied with requests by governments for content removal or partial censorship. In this study, we analyzed our collection of more than 100,000 tweets that were either censored tweets or retweeted censored content. We showed variability in audience location using time zone and language preferences as proxies, which is not bounded by geographic location of the censorship. We show that most of the time content find its way to disseminate and reach broader audience even with the censorship.

CCS Concepts

•Information systems → *Web mining*;

1. INTRODUCTION

Ease of access to online services creates opportunities to freely discuss and share opinions and to debate different points of view. Political discussions are the most influential for individuals and consequential for society [3, 5]. In some societies, governments have responded to the growing phenomenon of political mobilization by either terminating access to the online services or developing laws to restrict the exchange of information [6]. Countries like China, Iran, North Korea, and Turkey are applying internet censorship widely and monitoring social media and news to control online discourse. If discussions turn to sensitive topics, concerned governments intervene and attempt to control information dissemination [4, 1]. The majority of the related

previous research focused on Weibo platform [2, 7], since the Great Firewall of China prevents foreigner social media services.

Social media platforms are receiving an increasing number of requests for content removal and account closures. When these requests are rejected, some governments simply terminate access to these services. To respond to censorship requests, Twitter developed a system to withhold tweets and users from particular locations based on the internet protocol (IP) addresses of users. Twitter's approach limits access to content from particular locations as requested by governments while protecting the rights of other users to access content.

2. DATA COLLECTION

Twitter is a popular micro-blogging platform that is available to millions of people all over the world. Some of the content shared on Twitter might not be legal under applicable laws in various countries such as copyright, pornography, threatening messages, and insults to other users. Twitter receives requests for removal of content and users from various governments and law enforcement agencies. If removal requests are submitted properly by authorized entities, Twitter grants censorship to these requests. Other than content removal, Twitter can limit access to a particular tweet or user when requested by some country. *Withheld tweets* are censored only by the country that makes such a request to Twitter. To apply content censorship, Twitter determines the user locations based on IP addresses.

To study censored content on Twitter, we first extracted all withheld tweets and their retweets from our collection (approximately a 10% random sample of all public tweets streamed in real time) starting from June 2013 to December 2015. This collection consists of 29,619 unique users who were tweeting and retweeting censored content and 716 of these accounts created at least one censored tweet. In total, we analyzed 53,028 censored tweets and 99,643 retweets received by those.

3. RESULTS

3.1 Temporal analysis of censorship

Twitter has been accepting requests for content removals starting from 2012. Since that time several countries have been submitting requests. Consequently, thousands of tweets have been withheld temporarily or permanently. We collected those censored tweets and their retweets. We observe consistent statistics with Twitter transparency reports as

*Corresponding author: ovarol@indiana.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WebSci '16 May 22-25, 2016, Hannover, Germany

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4208-7/16/05.

DOI: <http://dx.doi.org/10.1145/2908131.2908208>

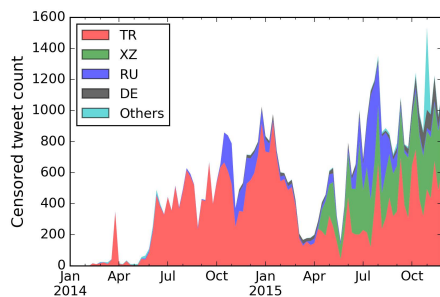


Figure 1: Time series of weekly frequencies of withheld content.

Turkey (TR), Russia (RU) and Germany (DE) are listed as top countries.

To characterize temporal changes of censorship, we study the volume of censored content and their corresponding countries. In Fig. 1, we observe a rapid increase in the volume of censored tweets starting from January 2014. The number of censored tweets reveals an increasing trend. We also observe seasonal changes for the amount of withheld content due to the activity rate of some popular accounts. For instance Turkey’s censorship requests were highest in January 2015 and Russia had more censorship requests during July 2015.

3.2 Geographical censorship

To investigate the effectiveness of IP-based censorship, we investigated the relationship between censored countries and time zone and language preferences of retweeting users as a proxy to user location. In Figure 2, we show co-occurrence between censored country codes and the primary language of users. In this analysis, we can see a relationship between countries and languages due to their political and cultural relevance. Similarly in Fig. 2, users have diverse time-zone preferences regardless of where the content is being censored. For instance in Turkey and Russia, majority of the retweets are created by users from UTC+2 and UTC+3 time zones, which correspond to these countries’ local time zones.

These analysis of language and time-zone preferences indicate that the audience of censored content is not bound by geographic location. It is known that citizens of countries with strict internet regulations adopt strategies against censorship by using VPN services or changing DNS settings to access censored sites.

4. CONCLUSIONS

Our initial analysis focused on evaluating the effectiveness of the current censorship system on Twitter. Controlling access to the content from particular geographical area is not an efficient way since many users integrate VPN services and anonymous browsing using tools like Tor with their daily online activities. Internet users living in the countries practicing serious internet censorship learn how to regain their online freedom using the available services. We also observed variability in local time and language settings of retweeting users as well as large volume of activities from users in censored countries, which might imply evidence for VPN usage and anonymous browsing.

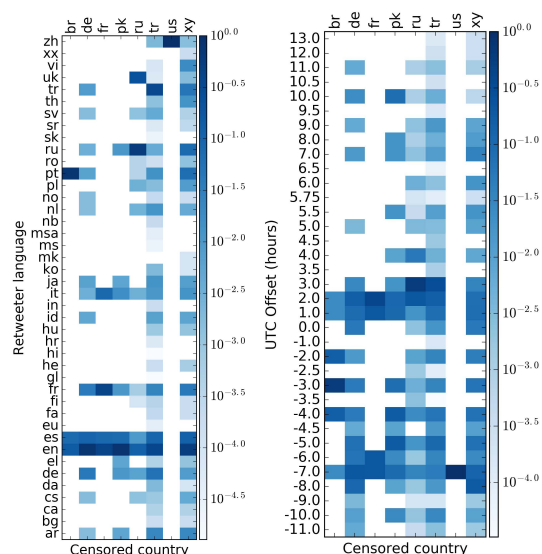


Figure 2: Co-occurrence relations for censorship countries (columns) shown for retweeting user’s language (left panel) and utf-offset (right panel). Observed frequencies are normalized by shared countries to highlight the distribution of retweeting users.

5. ACKNOWLEDGMENTS

This work was partially supported by NSF (grant CCF-1101743), DARPA (grant W911NF-12-1-0037), and the J.S. McDonnell Foundation.

6. REFERENCES

- [1] S. R. Ali and S. Fahmy. Gatekeeping and citizen journalism: The use of social media during the recent uprisings in iran, egypt, and libya. *Media, War & Conflict*, 6(1):55–69, 2013.
- [2] D. Bamman, B. O’Connor, and N. Smith. Censorship and deletion practices in chinese social media. *First Monday*, 17(3), 2012.
- [3] A. Choudhary, W. Hendrix, K. Lee, D. Palsetia, and W. Liao. Social media evolution of the egyptian revolution. *Communications of the ACM*, 55(5):74–80, 2012.
- [4] G. King, J. Pan, and M. E. Roberts. How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 107(02):326–343, 2013.
- [5] O. Varol, E. Ferrara, C. L. Ogan, F. Menczer, and A. Flammini. Evolution of online user behavior during a social upheaval. In *Proc. of Web Science*, pages 81–90, 2014.
- [6] L. L. Zhang. Behind the ‘great firewall’ decoding china’s internet media policies from the inside. *Convergence: The International Journal of Research into New Media Technologies*, 12(3):271–291, 2006.
- [7] T. Zhu, D. Phipps, A. Pridgen, J. R. Crandall, and D. S. Wallach. Tracking and quantifying censorship on a chinese microblogging site. *arXiv preprint arXiv:1211.6166*, 2012.