Double-edged Swords: The Good and the Bad of Privacy and Anonymity in Social Media

Mainack Mondal
MPI-SWS
Germany
mainack@mpi-sws.org

KEYWORDS

anonymity; privacy; social media; Whisper; Twitter; pattern recognition; hate speech; social sensors

Reference format:
Mainack Mondal. 2017. Double-edged Swords: The Good and the Bad of Privacy and Anonymity in Social Media. In *Proceedings of SIDEWAYS '17, Prague, Czech Republic, July 04, 2017,* 2 pages, CEUR-WS.org.

1 NEED FOR PRIVACY AND ANONYMITY IN ONLINE SOCIAL MEDIA SITES

Online Social Media sites (OSMs) like Facebook and Twitter drastically changed the way users communicate with each other and share content. OSMs provide inexpensive communication medium that allows anyone to quickly reach millions of users. Consequently, in these platforms anyone can publish content and anyone interested in the content can obtain it, representing a transformative revolution in our society. However, the very strength of OSMs to easily reach millions of users also indicates that there might be dire consequences to the content publishers if it reaches wrong people. For example, A content creator might lose her job simply because her rant in Facebook might be viewed by her boss [3] or even worse, Government organizations might press criminal charges against an activist solely based on her political opinions expressed in OSM posts [2]. To that end, over time, as OSMs gradually become a medium for freedom of expression, specially in times of unrest [9], there is a strong need for protecting users and their freedom of expression. In other words, there is a need for more privacy and anonymity to the users in OSMs so that they can preserve their right to free speech without fear of repercussions from their Government or other authorities. However, there is a cost—more private and anonymous platforms can also be abused to hurt other users, e.g., via spreading hate, cyber bullying or trolling.

In this talk, we will emphasize that we can leverage the OSM data as social sensors to understand privacy and anonymity practices in social media and improve upon them. Moreover, we argue that privacy and anonymity are double-edged swords—needed by many (e.g., activists during Arab Spring) but also abused by some to harm others. To that end, we argue that these social sensors might enable the OSM operators to stop the abuse too. We would next expand upon how social sensors can be used to (i) understand the need for privacy and anonymity and (ii) limit the abuse of these technologies to harm others. Finally we will conclude this talk with a call for action to explore future research directions in this space.

SIDEWAYS '17, July 04, 2017, Prague, Czech Republic

2 UNDERSTANDING PRIVACY AND ANONYMITY NEEDS VIA SOCIAL SENSORS

We start with pointing out that OSM data acts as a valuable microscope to look into the privacy and anonymity needs for millions of users. Traditionally, these needs are explored in the Human Computer Interaction (HCI) community via semi-structured interviews and user surveys. However, OSMs provide us a tremendous opportunity in the form of social sensors to scientifically measure how millions of users are using (or abusing) privacy and anonymity in a real world setting.

Understanding privacy via social sensors: Plethora of legal, sociology, psychology and even philosophical scholars aimed to understand concrete aspects of privacy. However, their definitions provides us understanding of *what is privacy*, but not *how privacy is enforced and used in practice*. Using OSM data we can partially bridge this gap, particularly in social context. We propose exposure control [5], an improvement over current privacy management models. Exposure is simply defined as *who actually views the content* and controlling exposure satisfy many privacy needs of OSM users today. However exposure control is a theoretical model and we need to specifically understand how exposure is controlled in real world and what we can do to improve them.

To that end, social sensors enable us to measure how users are actually controlling their exposure today and point out the limitations of current mechanisms. We have leveraged social sensors to identify the usage of social access control lists (SACLs) in real world [6]. Using real world SACL usage data we propose a simple cache-based mechanism to make SACLs more usable. Further, we looked into how users are protecting their longitudinal privacy by changing privacy settings of their historical data [7]. We found that a surprisingly high number of users are controlling longitudinal exposure of their content—more than 30% of social content posted 6 years back is withdrawn by overs. However, using this same OSM data we identify some key problems with longitudinal mechanisms in OSMs today and propose improvements of longitudinal exposure control mechanisms. Our work demonstrates the usefulness of social sensors for understanding and improving privacy; However, there is a need to further understand other aspects of exposure in different social scenarios (e.g., privacy violation via social search); we note that our object in improving privacy can be achieved by further leveraging the enormous behavioral data from OSMs.

Understanding anonymity via social sensors: Anonymity is another need for OSM users that is becoming more and more important in recent years. E.g., during a political turmoil, activists or whistle-blowers want to reach millions of fellow citizens, but don't want to face the wrath of authorities who monitor OSMs for finding these activists and silence them. To that end, anonymous OSMs like Whisper, Yik-Yak or 4chan is becoming popular as mediums to exercise freedom of expression. However, it is important to understand the usage of these anonymous platforms to detect if millions of users (and not only a handful of activists) indeed use these platforms to post content which need anonymity (i.e., personal experiences or strong opinions). To that end, we collected large scale data from Whisper [1] and compare this content with non-anonymous OSM, Twitter. Using these datasets as sensors we found that anonymity sensitivity of most whispers (posts from Whisper), unlike tweets (posts from Twitter) is not binary. Instead, most whispers exhibit many shades or different levels of anonymity. The content of whispers ranges from posting confession to opinions on LGBTQ. We also find that the linguistic differences between whispers and tweets are so significant that we could train automated classifiers to distinguish between them with reasonable accuracy. Our findings shed light on human behavior in anonymous media systems that lack the notion of an identity. Among other implications, these social sensors also open an exciting venue for us to understand the disinhibition effect, where users post content in presence of anonymity which they otherwise will not post.

3 LIMITING ABUSE OF PRIVACY AND ANONYMITY VIA SOCIAL SENSORS

Privacy and anonymity, however, have a dark aspect too, which cannot be ignored in the current world. When OSMs enable people to express themselves privately and anonymously, there are always some users who abuse the systems and hurt others. Particularly, OSMs have become a fertile ground for inflamed discussions, that usually polarize 'us' against 'them', resulting in many cases of insulting and offensive language. There are cases where individuals are mentally scarred forever by public shaming on online media or received death threats. The situation is becoming so worrisome that many Governments are now taking active steps to stop online abuse. For instance, in UK, 43.5% of children between the ages of 11 and 16 were bullied on social sites [8].

We argue that we can leverage OSM data as sensors to detect abusive atrocities and thus this very data can be used to limit the abuse of OSMs. Specifically, we note that abusive acts like cyber bullying, trolling or hate speech take place on OSMs and thus, there is a chance to automatically detect and limit them right when they are posted. We present a proof-of-concept example for this idea: understanding hate speech in social media [4]. We use sentence structures to create a high precision dataset of hate speech in OSMs. Using this dataset we investigate the types of hate that propagates in OSMs. We found that hate speech based on race, physical or behavioral features are common in OSMs. Moreover there is intra as well as inter country differences in the types of posted hate speech. Our findings demonstrate the effectiveness of sensing abuse using OSM data and hints at the possibility to improve upon abuse detection mechanisms.

4 FUTURE DIRECTIONS: A CALL FOR ACTION

Finally, we would like to conclude this talk with a call for action: leverage the available social sensors for improving privacy and anonymity in OSMs as well as keeping the abuse of these platforms at bay. We point out 3 high level directions:

Understanding privacy and anonymity requirements of users: OSMs provide researchers an unique opportunity to analyze astronomical amount of user generated data; The data can be leveraged as sensors to understand and improve upon aspects like privacy and anonymity. Specifically, this data can be used to find the mechanisms that users employ to control exposure of their data and check the effectiveness of those methods. Further data from anonymous OSMs like Whisper also can be used to understand the behavior of users in anonymous social media sites and can help in understanding the anonymity requirements. For e.g., an important question to investigate would be: how to measure and satisfy different anonymity requirements of users for different types of content?

Limiting abuse of OSMs leveraging big data: Another field that traditionally received less research focus than privacy and anonymity is to limit the abuse of OSMs. It is safe to say that, although becoming more and more important in recent years, research on detecting and limiting online abuse is still at its nascent stage. For example, what are different classes of online abuse? What are their concrete definitions and characteristics? Are there enough social signals in OSM data to detect abusive behavior? Can we build effective systems to limit these abuses in real time? In fact, a very first step might be methodological—how to automatically detect different types of abuse in online social media? We strongly feel that OSM data can help tremendously in solving both of these challenges and correctly leveraging this data paves a way towards a safer online environment.

REFERENCES

[1] Denzil Correa, Leandro Araùjo Silva, Mainack Mondal, Fabracio Benevenuto and Krishna P. Gummadi. 2015. The Many Shades of Anonymity: Characterizing Anonymous Social Media Content. In *ICWSM'15*.

[2] Facebook activism 2017. Iran social media activists held on 'security' charges. https://english.alarabiya.net/en/media/digital/2017/04/12/Iran-social-media-activists-held-on-security-charges-.html. (2017). Accessed on May 2017.

[3] Facebook fired 2016. 20 Tales of Employees Who Were Fired Because of Social Media Posts. http://people.com/celebrity/employees-who-were-fired-because-of-social-media-posts/. (2016). Accessed on May 2017.

[4] Mainack Mondal and Leandro Araújo Silva and Fabrício Benevenuto. 2017. A Measurement Study of Hate Speech in Social Media . In *HT'17*.

[5] Mainack Mondal, Peter Druschel, Krishna P. Gummadi, and Alan Mislove. 2014. Beyond Access Control: Managing Online Privacy via Exposure. In *USEC'14*.

[6] Mainack Mondal, Yabing Liu, Bimal Viswanath, Krishna P. Gummadi, and Alan Mislove. 2014. Understanding and Specifying Social Access Control Lists. In *SOUPS'14*.

[7] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. 2016. Forgetting in Social Media: Understanding and Controlling Longitudinal Exposure of Socially Shared Data. In *SOUPS'16*.

[8] nobullying.com. 2017. The Startling Facts about Cyberbullying in the UK. https://nobullying.com/cyberbullying-in-uk/. (2017). (Accessed on May 2017).

[9] Twitter arab stpring 2012. Twitter Revolution: How the Arab Spring Was Helped By Social Media. https://mic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media. (2012). Accessed on May 2017.