

Who Wants To Get Fired?

Ricardo Kawase
L3S Research Center
Hannover, Germany
kawase@L3S.de

Bernardo Pereira Nunes
L3S Research Center
Hannover, Germany
nunes@L3S.de

Eelco Herder
L3S Research Center
Hannover, Germany
Eelco.Herder@L3S.de

Wolfgang Nejdl
L3S Research Center
Hannover, Germany
nejdl@L3S.de

Marco Antonio Casanova
Department of Informatics
PUC-Rio
Rio de Janeiro - Brazil
casanova@inf.puc-rio.br

ABSTRACT

Microblogging services like Twitter have witnessed a flood of users and short updates (tweets). Although this phenomenon brings new possibilities of communication, it also brings dangerous consequences. From time to time, people post tweets guided by strong emotions. By default, tweets are public and anyone, anywhere can instantly see your updates, creating high exposure and lack of awareness about privacy issues. In many cases, this may lead to consequences that can be harmful to one's personal and professional life. In this paper, we investigate the posting behavior of people who tweet that they hate their jobs and bosses and their responses to alerts about the potential damage that such a tweet may cause. We show that, in many cases, people are not aware about the dimension of their audience, and once alerted, they often regret what they have publicly said. Our analysis leads us to believe that many users could benefit from a 'give a second thought before posting' tool that may save their jobs.

Author Keywords

Twitter, privacy awareness, user issues

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation:
Miscellaneous

General Terms

Human Factors

INTRODUCTION

Recently, the microblogging service Twitter has become one of the most popular social networks available on the Web, reaching almost 300 million active users [1]. This rapid growth comes with a great concern about privacy

of information, since users are not always aware of where and to whom these data will be available.

Although Twitter updates are meant to be publicly available (at most restricted to the authors' followers), a lot of private and sensitive information is leaked via tweets. In fact, not every user realizes the consequences of such information leaks, while some others are not aware of the potential audience of their tweets.

Wang et al. [2] have demonstrated through a qualitative study that, indeed, several users regret posts written on Facebook - specially when under the influence of drugs, alcohol or emotion instability.

In this paper, we focus on identifying the unawareness of Twitter users regarding their privacy. We choose to study those users who put their jobs at risk by publicly announcing their discontentment with their works or their bosses. Our main goal is to raise attention to the fact that many users are not aware of their audience. We achieve this by exposing statistics of their potentially risky behavior that it is publicly available. Additionally, we built an online system that alerts users when a tweet might compromise their jobs. On top of the feedback collected, we conclude that many users could benefit from a 'give a second thought before posting' tool.

PRIVACY ISSUES

Social media sites, such as Twitter and Facebook, are designed to share information - and other content, such as pictures, videos and links - among users. Apart from relatively harmless updates, such as sharing a link or other types of public content, messages on Twitter and Facebook may contain highly personal information such as geolocation or email. For this reason, social media sites typically offer their users several ways to indicate the intended audience of shared messages. First of all, there are *default* settings - which can be adapted by the user. Second, users can overrule these default settings for specific messages. Third, in many cases it is possible to delete, hide or edit a message post hoc.

However, as indicated by several studies (e.g. [5]), users often do not inspect or adapt the default settings offered

Permission to make digital or hard copies of all or part 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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'13, April 27–May 2, 2013, Paris, France.

Copyright 2013 ACM 978-1-XXXX-XXXX-X/XX/XX...\$10.00.

by the system; thus, most messages are sent with the default settings. Due to this behavior, messages often have a wider audience than intended or expected by the user. According to a recent report from the Pew Internet & American Life Project [6], particularly males and young adults have posted content that they regret; not surprisingly, these are also the users with the least restricted privacy settings. However, due to the raising awareness of privacy issues and their implications, more and more users actively manage their privacy settings and prune their profiles.

There are many risks associated with content that is unknowingly disclosed to the public. Some of these risks - including mobbing, loss of reputation, family problems and lost career opportunities - are summarized in [7, 8]. A remarkable initiative to raise attention for these issues is the site PleaseRobMe¹, which aggregates and shows tweets of users who report to be away from home. In addition, the user is informed via a (public) tweet.

In many cases, users believe that they have an *anonymous online presence* for those who do not know them personally. They assume that if they do not fill information like real names, emails, location, among others, their anonymity will not be compromised. However, among various other studies, Hecht et al. have proven that, by analyzing the users' stream and social graph, it is possible to identify a user's location 'with decent accuracy', since users implicitly reveal their location [4].

FIREME!

In order to address privacy issues and sensitive information leaks on social networks, we chose to tackle specifically those public updates on Twitter where users express their disappointment regarding their jobs and bosses. To emphasize the recklessness of some people when posting updates about their working environment, we called our framework FireMe!

In FireMe!, we track every twitter update that mentions, inappropriately, the authors' working environment. We chose a set of 13 sentences that identifies the user dissatisfaction with their jobs or employers. For example, sentences like 'I hate my job', 'I hate my boss', 'I have the worst job' and other sentences that include harsh profanity. Note that our goal is not to identify all possible dissatisfactions, but to address those that we discovered. For the remainder of this paper, we call the author of such tweet a *'hater'*.

We divided the work around FireMe! into two stages. First, we perform a data analysis to characterize the profile of a hater. Further, to address the real state of awareness of Twitter users, we built the online FireMe! alert system that warns these users about a tweet that may put their jobs at risk - we assume that no boss would be happy to be publicly profaned online or to find out that their employees hate their jobs.

¹<http://pleaserobme.com/>

Data collection

Before deploying the FireMe! as an alert system, we first collected as many *haters* as we could during one week, between June 18 until June 26, 2012. In this period we gathered a total of 21,852 *haters*, which correspond to almost two reckless tweets per minute. On other hand, during the same period, we also collected tweets from what we call *'lovers'*, people who posted delightful updates about their jobs, such as 'I love my job', 'my boss is the best'. We found twice as much *lovers*(44,710) than *haters*.

From these two sets, we randomly selected 10,000 users from each group for further analysis. For each of these users we collected the past 200 tweets on the users' stream, summing up to approximately 2 million *haters'* tweets and 2 million *lovers'* tweets.

In addition, we polarized each tweet using the www.sentiment140.com natural language processing API based on the Maximum Entropy classifier [3], a state-of-the-art method for classifying sentiment of tweets. For each given tweet, the service classifies it as positive, negative or neutral. Finally, we counted the number of profanity words in each tweet.

DATA ANALYSIS

Haters versus Lovers

A closer look on the data collected (Table 1) reveals interesting characteristics of *haters* in comparison to *lovers*. The first thing to notice is that *lovers* are better connected within the social graph. In our sample, *lovers* have 3 times as many followers and around 20% more friends. On the other hand, *haters* seem to be more active in terms of tweeting speed, posting twice as many tweets per day than *lovers*. As expected, *haters* are less careful regarding profanity on their language: they curse more often than *lovers*. Finally, we verified that *lovers* are rather more positive in their comments.

In Twitter, users can add their personal Website information (link a external Website to their profiles). Around 36% of the *lovers* added the Website information, while 22% of *haters* did the same. From those who had such information, we checked the profiles that contained a link to their Facebook profile (potentially exposing even more information about the user). In 20% of the cases, *lovers* linked their Facebook to their Twitter account, while *haters* did it with a 31% rate.

Who wants to get fired?

The online FireMe! ² alert system monitors online Twitter updates from users that expose discontentment with jobs and bosses. After FireMe! recognizes a *hater-user*, the system sends an alert tweet to this user with a warning message. Each alert tweet also contains a unique link where users can access and visualize their FireMe! score

²<http://fireme.l3s.uni-hannover.de>

Table 1. Averages characteristics of *haters* and *lovers*. Speed is tweets/day.

	Followers	Friends	Tweets	Speed	ReTweets	Profanity	Negative Tweets	Neutral Tweets	Positive Tweets
Haters	446	368	9599	7.0	42.9	14.7	27.0	137.3	32.1
Lovers	1214	444	10118	3.8	43.8	8.3	20.5	129.7	45.4



Figure 1. FireMe! alert interface.

(FireMeter!). Unfortunately, due to Twitter API limits, we were not able to alert every single identified *hater*.

The FireMeter! score is a mixed computation that considers job discontentment messages, profanity and job mentioning in the past 100 users' tweets. For the sake of entertainment, we present to the users *'how likely is her chance of being fired, in case her boss found out about her Twitter account'*.

Once the user accesses the link, we show the *hate-tweet* and the FireMeter! score. Before explaining the users the reason of the score, we first ask them what they will do about it. Three options are given: 'Delete tweet', 'Check privacy settings' or 'Don't care'. Once the users clicked on their probable action, they can access a history page (last 100 tweets) where we highlight all job mentions and profanities. In addition to that, we allow any user to check their FireMeter! score and provide a *leaderboard* where users can compare their scores with others³.

Alert Impact

In order to evaluate the impact of FireMe! on Twitter's users, we collected the outcomes after a short period of three weeks, between August 16 and September 7, 2012. During this time, the system sent in total 4304 alert

³Only users who have been checked in FireMe! enter the leaderboard.

messages to unique *haters*. From those users that got an alert message regarding their potentially harmful tweets, 914 of them (around 21%) accessed the link to FireMe! and checked their FireMeter! scores.

A likely explanation for the relatively low percentage of people actually accessing the Website is that the exposure of users' personal data - and specially their (not so polite) tweets - may give them second thoughts to further engage in our Website. Nevertheless, in this short period, we got 243 replies to the question *'What you gonna do about it?'*.

In total, 101 users answered that they would delete the *hate-tweet* (around 42%), 45 users claimed they would check their privacy settings (18%) and 97 just do not care about it (40%). In the end, almost 60% of the users who gave us feedback are actually concerned about their personal data and the impact that it may have if the wrong person finds out about it. To validate the users' feedback, we directly accessed *haters* and *hate-tweets* to check if any action was indeed taken.

From the 4304 alerts sent (response for *hate-tweets*), 249 *haters* deleted their original *hate-tweet*. We checked the tweet status two hours after the alert message was sent. We noticed that many users deleted their *hate-tweets*, but did not visit our Website (only 69 of them actually accessed FireMe!). We believe that just the warning message (e.g. *'Hi @hater.user Do you think your boss will like this?'*) was enough to make the user realize her imprudence.

Finally, some users interacted with FireMe! only via Twitter. We received 241 replies in response to our messages, 144 mentions and 88 retweets. Most replies were not very friendly and people were annoyed that we were monitoring their activity. We also got tweet replies from users who appreciated our work, even from those who do not care about consequences:

- @WhyFired THANKS BUDDY
- @WhyFired *It's true. You're right, I was immature reason but no one is perfect in this world and I made a mistake and I apologize.*
- @WhyFired *Oooh...i'm so scared. but i commend your use of twitter though.*
- @WhyFired *they already know that I hate my job. I'm in the process of leaving. Cheers for the heads up though!*

Who Cares and Who Does Not

Of the users who indicated that they intend to delete the compromising tweet, about 45% actually did so. A lower, but still quite high percentage of users who indicated that they did not care, actually deleted their tweet (33%). Interestingly, the group of users who indicated that they intended to change their privacy settings, had the lowest rate of deletions (28%). This may indicate that the latter group was unsuccessful in finding the relevant privacy options in Twitter.

It is a likely assumption that authors of tweets with a high FireMeter! score are more likely to delete their tweet than those with a lower score. The reverse turned out to be the case: users who indicated that they did not care had a significantly ($F = 10.02; p < .05$) higher score (72%) than those who planned to delete their tweet (62%) or to change the privacy settings (59%). This suggests that an inappropriate tweet is not an exception in a user's 'oeuvre', but rather an element of the user's overall behavior - this result is in line with our analysis in 'Haters versus Lovers' section.

As we expected that the toning of the alert would have an impact on the user's reactions, we experimented with three different types of alerts: *neutral* messages (e.g. 'if you hate your job, why not simply quit?'), *aggressive* messages that might embarrass the user (e.g. 'tweeting this in public is rather stupid #fail') and messages that suggested the user to take some *action* (e.g. 'even if you hate your job, you'd better delete or hide this tweet'). Users who received an 'action' message deleted their messages more often (13% of the cases) than those who received a neutral or aggressive message (both 6%).

Discussion and Conclusion

In this paper, we investigated users' awareness of the bandwidth of the audience and the potential consequences of negatively loaded, personal tweets. We focused on users who wrote tweets that they hate their jobs or their bosses. By monitoring Twitter with only a couple of English-language *hate-queries*, we were able to identify over two *haters* per minute. An analysis of their profiles showed that 'haters' tweet more than regular users and that their tweets are more negatively loaded; haters are typically less connected than others. Such aspects may be useful for proactively identifying potentially reckless users.

In a period of three weeks, we sent alerts to over 4000 'haters'. Of the users who received an alert, around 21% accessed FireMe! Website. Additionally, many others interacted with us directly via Twitter (replies, mentions and retweets). Interestingly, users with lower FireMe! scores were more inclined to delete the concerning tweet than those whose tweets were ranked as more reckless. This may indicate that, in particular, people who already care about their privacy would benefit from an alert system that motivates users to revert potentially harmful actions.

We experimented with three types of alerts: alerts that explicitly suggested the user to take some action turned out to be more effective than messages that did not contain such a suggestion. Still, less than 45% of the users who indicated that they wanted to delete their tweet actually did so. This issue can be solved by providing explicit instructions on *how* to delete a message or to provide a one-click function to do so.

Alerts about potentially harmful behavior may be considered annoying to some users, in particular to those who generally care less about their online privacy. We think that 'reckless' users should be given the opportunity to turn these alerts off. However, particularly inexperienced or young users would benefit from post-hoc privacy alerts. Potential dangers of personal, negatively loaded tweets remain abstract for most users, until the damage has been done.

REFERENCES

1. Twitter now the fastest growing social platform in the world. <http://globalwebindex.net/thinking/twitter-now-the-fastest-growing-social-platform-in-the-world/>, Jan. 2013.
2. A. Acquisti. I regretted the minute i pressed share: A qualitative study of regrets on facebook. volume 28, pages 169–185, 2011.
3. A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *Technical report, Stanford*.
4. B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 237–246, New York, NY, USA, 2011. ACM.
5. Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 61–70, New York, NY, USA, 2011. ACM.
6. M. Madden. Privacy management on social media sites. Technical report, Pew Internet and American Life Project, 2012.
7. C. Rose. The security implications of ubiquitous social media. *International Journal of Management and Information Systems*, 15(1), 2011.
8. E. Toch, Y. Wang, and L. Cranor. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, 22:203–220, 2012. 10.1007/s11257-011-9110-z.