

# Listen to Me if You can: Tracking User Experience of Mobile Network on Social Media

Tongqing Qiu  
Georgia Tech  
Atlanta, GA  
tongqqiu@cc.gatech.edu

Junlan Feng  
AT&T Labs – Research  
Florham Park, NJ  
junlan@research.att.com

Zihui Ge  
AT&T Labs – Research  
Florham Park, NJ  
gezihui@research.att.com

Jia Wang  
AT&T Labs – Research  
Florham Park, NJ  
jiawang@research.att.com

Jun (Jim) Xu<sup>\*</sup>  
Georgia Tech  
Atlanta, GA  
jx@cc.gatech.edu

Jennifer Yates  
AT&T Labs – Research  
Florham Park, NJ  
jyates@research.att.com

## ABSTRACT

Social media sites such as Twitter continue to grow at a fast pace. People of all generations use social media to exchange messages and share experiences of their life in a timely fashion. Most of these sites make their data available. An intriguing question is can we exploit this real-time and massive data-flow to improve business in a measurable way. In this paper, we are particularly interested in tweets (Twitter messages) that are relevant to mobile network performance. We compare tweets with a more traditional source of user experience, i.e., customer care tickets, and correlate both of them with a list of major network incidents. From our study, we have the following observations. First, Twitter users and users who call customer service tend to report different types of performance issues. Second, we observe that tweets typically appear more rapidly in response to network problems than customer tickets. They also appear to respond to a wider range of network issues. Third, significant spikes in the number of tweets appear to indicate short term performance impairments which are not reported in our current list of major network incidents. These observations together indicate that Twitter is an attractive, complementary source for monitoring service performance and its impact on user experience.

## Categories and Subject Descriptors

C.2.3 [Computer System Organization]: Computer-Communication Networks—*Network Operations*

## General Terms

Management, Measurement, Performance

<sup>\*</sup>This work was supported in part by NSF grant CNS-0905169, funded under the American Recovery and Reinvestment Act of 2009 (Public Law 111-5), and NSF grant CNS-0716423.

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, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IMC'10, November 1–3, 2010, Melbourne, Australia.

Copyright 2010 ACM 978-1-4503-0057-5/10/11 ...\$10.00.

## Keywords

Social Media, Twitter, User Experience

## 1. INTRODUCTION

Monitoring network performance is one of the key tasks in network operation. We detect service performance issues from both a systems perspective and from a customers' perspective. From the systems perspective, we infer the issues based on measurement of network delay, packet loss, etc. From a customers' perspective, we identify issues based on users' feedback. The traditional way to learn user feedback is through customer calls or emails. When a customer complains about a problem, we investigate and resolve it. In this paper, we propose to go beyond customer care data to exploit a different channel – online social media, for tracking user feedback regarding service performance.

Online social networks (OSN) have gained significant popularity during recent years [18]. Microblogging services such as Twitter, are one of popular means through which users share information and experiences on the web. Comparing with Facebook, LinkedIn, MySpace, YouTube, and other social networking services, messages on Twitter are short (less than 140 characters). Twitter messages are widely referred to as *tweets*. It takes only a few seconds for a user to write a tweet and have it distributed to the public and his followers. Users can send and receive tweets through various applications (such as web and instant messaging) and devices (such as mobile phones, TV and computers). According to comScore, Twitter finished 2009 with nearly 20 million visitors to its website, up from just 2 million visitors in 2008 [1].

In this paper, we analyze tweets related to one of the largest mobile service providers in the United States. We first identify tweets that relate to service performance issues, and compare them with customer care trouble tickets. Second, we correlate these two sources of customer feedback with a report of major network incidents. Our findings are threefold: (1) Issues reported on Twitter are complementary to customer care calls. (2) Twitter users are faster to report service performance issues compared to customers who call and complain to the customer care center. (3) Tweets report some short term and/or less severe problems, which are not recorded in the major network incidents report.

The remainder of the paper is structured as follows. Datasets for analysis are discussed in Section 2, followed by our presentation of results in Section 3. Section 4 reviews related work and Section 5 concludes the paper.

## 2. DATASETS

In this section, we discuss the methodology used to collect data and describe the data used in our work. Section 2.1 discusses Twitter data, Section 2.2 discusses customer care call data, and Section 2.3 discusses the major network incidents report data.

### 2.1 Twitter Data

We used Twitter APIs to retrieve publicly available data relevant to our task. We emphasize here that only information that was shared publicly by Twitter users was obtained and analyzed. First, we manually selected a few keywords that were deemed as good queries for retrieving tweets relevant to the mobile service provider in question. Second, we used the Twitter search API to obtain tweets and we archived the retrieved tweets along with the associated meta data. The meta data include the time that the tweet was submitted, the user who authored the tweets, etc. Third, we fetched user information through the Twitter REST API for those who authored the tweets that we archived. User information consists of the user profile such as user location to help localize issues reported.

After data archiving, our next step was to identify tweets related to mobile network performance issues. We used a few heuristic rules: (1) Tweets must contain mobile related words such as phone, mobile, 3G, edge, etc; (2) Tweets must contain performance related words such as slow, drop, intermittent, doesn't work, etc; and (3) Tweets should not contain words indicative of advertising, such as "Ads" and price symbols \$. To verify the effectiveness of these rules, we randomly sampled 100 tweets, and manually annotated whether they were related to mobile performance issues. We observe an 87% agreement between rule-based prediction and human annotation. This is an acceptable level of accuracy<sup>1</sup> for our study. There are potential methods via natural language processing and machine learning which would likely improve the performance of this step. We leave this as future work.

### 2.2 Customer Care Calls

We obtained customer care tickets that had been created in response to customer calls for the mobile service provider. Note that these tickets are anonymized; no customer identification information is used from the tickets in this analysis. These tickets are tagged with the types of issues, i.e., billing and accounting, calling plan and features, mobile devices, service coverage, performance impairments and service outages. In this paper, we focus on the customer tickets relating to service impairments and outages. In addition, each ticket contains information regarding the type of service, the time that the call was received by the customer care team (the trouble ticket is often issued at the same time), the location, and a description of the performance issues that the customer experienced. The description usually provides detailed information on when and where the customer experienced the performance impairment as well as the device and application that the customer was using when the performance impairment occurred.

In our data, the performance related customer trouble tickets include issues such as no coverage, cannot make or receive calls, call disconnected/dropped, and poor voice quality, etc. It is important to note that not all customers will call the customer care team and report the performance impairments that they experienced. In addition, we will show later that customers may not call the customer service center immediately after they experience the performance impairments.

<sup>1</sup>In statistical scene, the estimation result 87% is within 95% confidence interval, with a margin error less than 0.1, given that we randomly selected 100 samples.

### 2.3 Major Network Incidents Report

The service provider has a top tier of operators that oversee the entire operation of the cellular network and service. They maintain a so called *Major Network Incidents Review* in a collaborative editing system. The network incident review keeps track of information regarding major network or service incidents as they are reported, diagnosed, resolved and concluded. The review report serves as a channel to communicate summary-level information about major incidents among the team members and senior management members.

The incidents in the report include a wide variety of network and service issues including hardware failures, major maintenance activities, outages due to adverse weather, congestion due to flash crowds (e.g., highway accident causing traffic congestion and unexpected high cellular network congestion), etc. Reading through the incidents report, we first filtered out those non-customer-impacting events from our study, such as planned system upgrade where redundant capacity is in use during the upgrade. We have also excluded very long incidents (e.g., greater than three hours) in the later part of the analysis so that the chance of falsely joining a network or service event to independent customer care tickets and tweets is low.

Each entry in the major network incidents report contains crucial information about an event. All entries contain temporal information (i.e., the start and end time of the incidents), and coarse spatial information, i.e., the primary market region where the incidents occurred (e.g, northeast, west, etc.). Moreover, all entries have facilitator/ incident manager contact details for follow up investigations, and the estimated scale of the customer impact. Most of the entries have *incident summaries* which describe the nature of the incidents in detail<sup>2</sup>. Some of the incidents contain root cause descriptions which explain the cause of the incidents in detail. In this paper, we only utilize the temporal and spatial information recorded for the incidents.

## 3. RESULTS

In this section, we first present the results on the major network incidents report, tweets and tickets data. Then we correlate both tweets and tickets with the incidents report. The data we analyze are based on a large cellular network provider in United States over the period of 16 days.

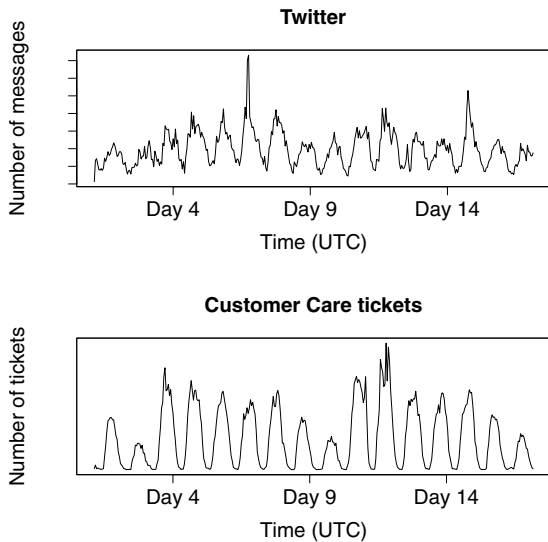
### 3.1 Major Network Incidents Report Results

We first analyze the temporal distribution of the major network incidents reported. Note that not all network incidents are included in the report - only those deemed most significant by the operations teams investigating. The report that we use contains only major network incidents. The vast majority of network incidents are fairly minor and not severe enough to be included in this report. Among these major incidents, we observe a very diverse duration ranging from seconds to over a day. As we mentioned in Section 2.3, in the later part of our analysis we focus on the incidents with duration of less than three hours.

### 3.2 Tweets vs. Customer Care Tickets

In this section, we compare two sources of user experience: tweets and tickets based on customer care calls. First, we conduct the comparison using the *raw* data, which include all tweets

<sup>2</sup>Although some descriptions contain certain kind of detail location information (for example, city, device, highway information), it is challenging to automatically and precisely extract them due to the loose structure of the description text.



**Figure 1: The number of tweets and tickets per hour. Due to privacy issues, the concrete number of tickets is not reported.**

that we have archived and considered as relevant to the mobile service provider (i.e., some tweets are not performance related). Then we classify the data into different categories and drill down to the performance-related tweets. Finally, we examine the delay of user feedback, namely, how long it takes for a user to report a problem after he/she experienced it.

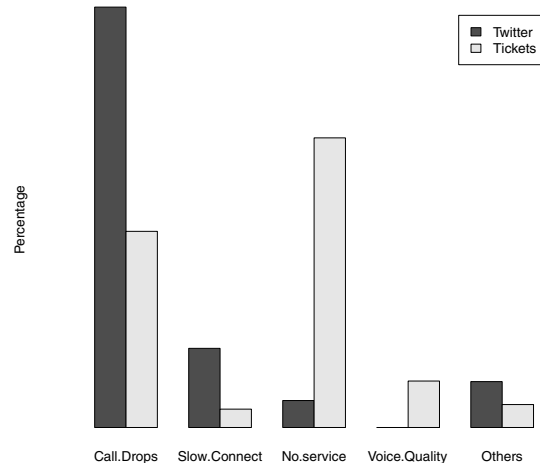
### 3.2.1 Time Series

We compare the volume of tweets and the volume of customer call tickets over a common period of time. Figure 1 shows the number of tweets and tickets per hour based on the *raw* data. From Figure 1, we observe the obvious daily pattern in both cases. It is easy to understand that customer calls have such pattern (common user behavior). It is also not surprising for tweets because of the daily pattern of social media access [2] and the fact that we focus on a US service provider. We will later (in Section 3.3) see that the diurnal pattern is weak when we focus on only performance related tweets. Another interesting observation from Figure 1 is the spike for Twitter data on Day 7. This spike is caused by discussions relating to new technology being made available to the consumer market.

### 3.2.2 Classification

We have investigated the *raw* data in the previous section. In this section, we move on to classify the data into different categories, and then focus on the performance related logs.

By manually inspecting the message content of the collected tweets, we observed the following three major types: (1) Comments and news regarding the product and customer service; (2) Advertisements (e.g. the promotion of a mobile phone); (3) Comments or complaints regarding performance related issues (which we are particularly interested in). Due to the loose organization of messages, it is difficult to precisely classify tweets automatically. But in general, the majority of the tweets are of type (1) or (2). On the other hand, since the tickets have a fixed structure and categorization, it is relatively easy to classify them. The majority (over 97%) of tickets are related to plan, bill or device issues. In the rest



**Figure 2: Classify performance related tweets/tickets based on the type of the issues. Due to privacy issues, the concrete percentages are not reported.**

**Table 1: Location information of tweets.**

Location Info.	No	Yes		
		City+State	State only	Others
Twitter Profile	37.2%	29.6%	20.5%	12.7%
Twitter Message	69.5%	20.5%	9.9%	1.0%

of the paper, we mainly focus on type (3), i.e., performance related tweets / tickets.

Among tweets extracted using the methodology described in Section 2.1, only about 1% of messages are related to performance issues. Similarly, only 1% to 2% of tickets are related to network performance. Figure 2 further breaks down the performance related issues into several categories. From this figure, we observe that issues reported by Twitter users and by customers who call customer care are very different. For Twitter users, call dropping is the most frequent complaint, followed by slow connection, no service and others (e.g., difficulty in sending messages or in posting something on websites). In contrast with the tweets, more tickets are related to lack of service or coverage issues (e.g. no coverage in some buildings). Very few of them are related to slow connection. Moreover, we also observe that many tickets are relate to voice quality, issues which are not typically seen in Twitter messages. This result illustrates that the types of performance issues reported by Twitter users are different from those reported by customer calls: the former tends to report short term and/or minor performance impairments, while the later tends to report more severe performance issues.

We also compare the tweets and tickets by locations. There are two ways to retrieve the location information of tweets: namely from the user profiles and from the message content itself. The message itself is a more valuable source because it presumably provides the location directly related to performance issues. Therefore, we first rely on the location information in messages. If there is no such information, then we consider the location in the user profile. Because both profile and messages are human language inputs without fixed structure, the location information could be missing

**Table 2: The delay between the customers experience the incidents and the customers report the incidents.**

Delay	0 day	1 day	> 1 day
Tweets	98.3%	1.2%	0.5%
Tickets	38.4%	21.2%	40.4%

or incomplete. There are three general categories of location information identified in our analysis: city + state, state only, others (e.g. city name only, street). Table 1 depicts the breakdown across these categories of the location information obtained from the user profiles and tweets analyzed. We observe that a large proportion of tweets have no location information at all - 37% of profiles have no location information and 69% of Twitter messages do not contain location specific information. It reduces the number of valid tweets for later correlation purposes. In contrast, the location information in tickets is well organized based on the *market* regions. A single market could include multiple states (e.g. Georgia + South Carolina) or cover just a part of one populous state (e.g. Northern California). Different data sources have different granularity of location information. Therefore, we use the most detailed common location for correlation purposes in Section 3.3 .

### 3.2.3 Timeliness

There is inherently a delay between the time when a customer experiences an incident and the time when the customer reports that incident. The reporting time is defined here to be the timestamp of when the Twitter message was posted or when the customer called customer care. It can be directly retrieved from the data. On the other hand, the time associated with when a customer experiences an incident is often hard to determine - if it is available, it will be embedded within a Twitter message or in a ticket’s description. For example, in Twitter, the message may reference “a couple of call drops *today*”; in tickets, the description may mention “no service since *yesterday*”. We extract such information from the tweets and tickets in a simple way: we identify the timing patterns like “now, today, yesterday, 3 days, 2010-04-03, this morning”. We observed that over 90.1% of our collected tweets do have such timing information. In comparison, fewer than 23.7% of ticket descriptions contain explicit timing information. We then computed differences between the times when the issues were reported and the estimated incident times as extracted from the tweet and ticket contents. The resulting distribution is illustrated in Table 2. We observe that most tweets are posted on the same day as the customers experience the performance issues. In contrast, most tickets are opened one or more days after the incident being reported. The result indicates that Twitter users respond much faster than those reporting issues via customer care. However, we should notice that these two groups of users may care about different kinds of performance issues. Therefore we cannot conclude that Tweets are faster than tickets for any particular incident.

## 3.3 Correlation Results

In this section, we correlate both performance related tweets and tickets with events reported in the major network incidents report. Figure 3 shows the time series of the incidents, tweets and tickets. Different from Figure 1, performance related tweets demonstrate only a weak diurnal pattern as shown in Figure 3. This implies that most of these Twitter messages are incident-driven. But customer tickets still have the strong diurnal pattern, because of the customers’ calling behavior and aforementioned delay in incident reporting.

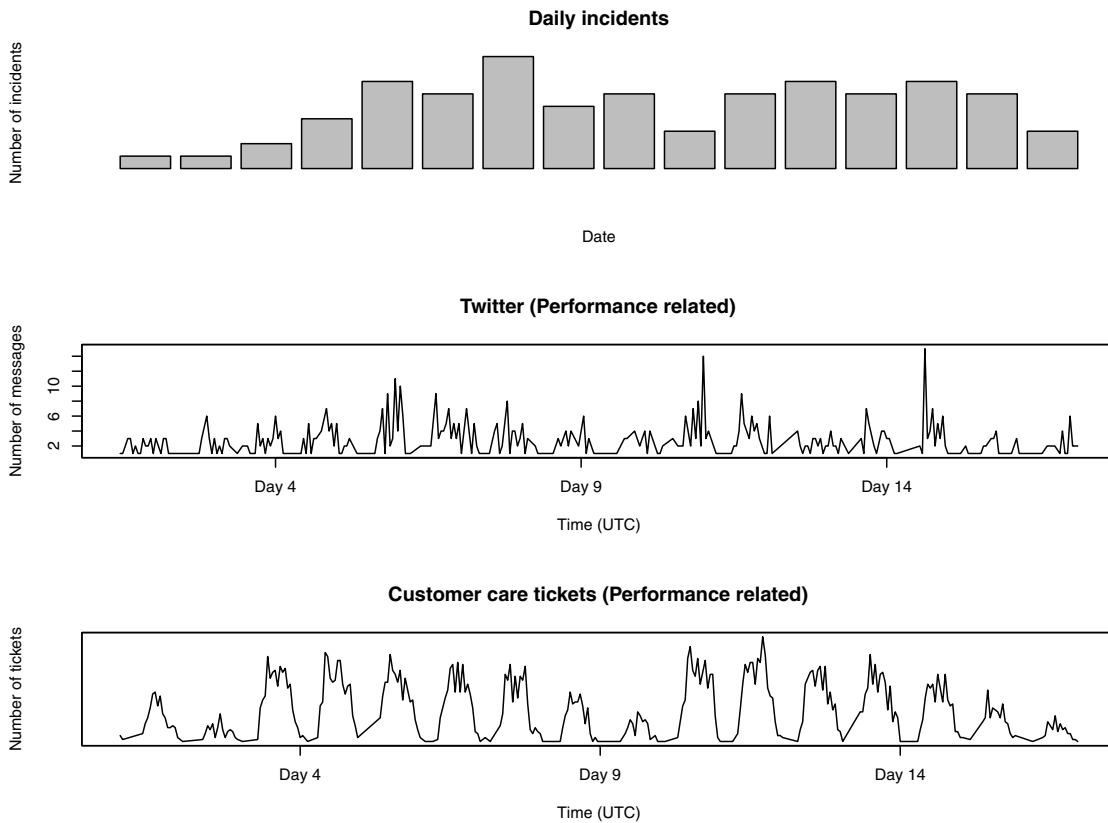
To understand if there is a strong correlation between incident reports and customer feedback, we conduct a statistic correlation [12] between the incidents and tweets/tickets series, for each location. We compute Pearson’s coefficient of correlation and conduct significance tests. Unfortunately, the test result shows no strong correlation between incidents and customers’ experiences. One possible reason is the time lag from the time when the incidents start to the time when users report performance problems. Correction of this time lag by some time shifting should be applied for the correlation tests. The challenge is that the time lags vary case by case and cannot be compensated systematically. We will leave the design of new statistical significance test methodology under this particular situation for the future work.

Instead of using statistical correlation, we use the incidents in the report as the ground truth and investigate whether these incidents are reported within the collected tweets/tickets. More specifically, if we observe tweets / tickers during the incident and occurring within the same region (e.g. east, west, central, etc.), we deem these tweets or tickets to be *associated* with the incidents. We verify the incidents which last less than 3 hours. The reason to verify relatively short-term incidents is that we have more confidence with the joins when the time window is limited. In other words, the chance of falsely joining an incident to independent tickets/tweets is low. Because of this filtering process, the results we observe during the correlation are valid for relatively short term incidents. We find that 55.6% of incidents can be found in tweets, and only 37.0% of them found in customer care tickets.<sup>3</sup> One interesting observation is that the matches found in tickets can also be found in tweets. Moreover, we use the number of associated tweets/tickets divided by the accumulated duration of incidents to describe the chance of observing tweets/tickets when incidents happen, denoted as  $c_1$ . Similarly, the chance of observing tweets/tickets when no incident happens, say  $c_2$ , can be measured as the number of un-associated tweets/tickets divided by the accumulated duration of non-incidents during the measurement period. We find that the ratio  $c_1/c_2$  is 8.3 for tweets and 6.8 for tickets. This suggests that the chance of having user feedback during an incident is significantly higher than that when there are no ongoing incidents.

In Section 3.2.3 we quantitatively studied the delay from the time when customers experience the service impact till the time when customers report the issue. In fact, there is another delay between the time when incidents take place and the time when users experience problems. However, the time when users experience the incidents is not an accurate timestamp, as we have discussed in Section 3.2.3. Therefore, we now quantitatively analyze the *total* delay from the time when the incidents start till the time when customers report the problem. Figure 4 shows the box statistics of two delay distributions. The bottom and top of the box are the 25th and 75th percentile, and the band near the middle of the box is the 50th percentile (the median). The ends of the whiskers represent the minimum and maximum value. The maximum value in both cases is approximately 80 minutes. Note that we filtered out all incidents exceeding 3 hours, which thus serves as the upper bound of the delay. We find that Twitter users respond approximately 10 minutes faster than customer who call the customer care team in general. The fastest Twitter users response is in several minutes. The implication of this observation is that it may be possible to utilize Twitter users’ feedback to observe the impact of network performance issues more timely than using customer tickets.

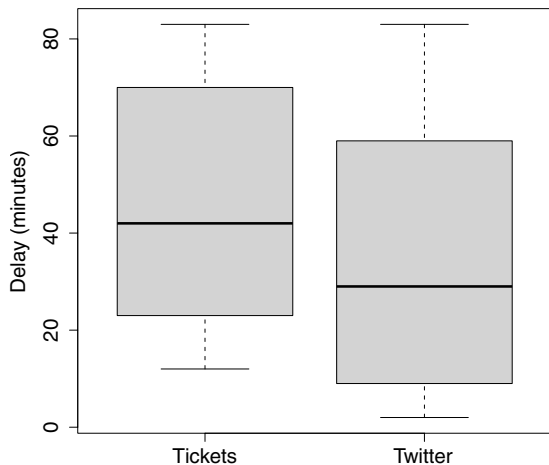
Finally, let us revisit Figure 3. There are obvious spikes on Day 11 for both Twitter and customer care data, even though there

<sup>3</sup>Due to limited coarse location information, there are some potential mismatch cases or false positive.



**Figure 3: The time series of incidents, tweets and tickets. Due to privacy issues, the concrete numbers of incidents and tickets are not reported.**

are very few incidents recorded in the same period of time. More specifically, there are many complaints from both Twitter and customer care regarding call drops during 8 PM to 10 PM in the central area. This may be an indication of certain short term network problems at that time, but yet none were reported in the major incident reports covering the area.



**Figure 4: The CDF of delay distribution.**

#### 4. RELATED WORK

There have been a number of studies regarding the social network. However, most of them focus on the social network itself, e.g., users behavior [18, 13, 20, 2], the impact on network performance [21], community evolution [7, 14], information propagation [23, 5, 4], privacy issues [8, 6]. Very few studies focus on showing the value of social network content. Vieweg *et al.* shows the microblogging such as Twitter can contribute to the situation awareness during natural hazards events like floods and fire [22]. Motoyama *et al.* use Twitter data to infer on-line Internet service availability [19].

Correlating across data sources is a common methodology used in anomaly detection [15, 9, 3] and network problem diagnosis [12, 10, 17, 16, 11]. Most of these papers focus on the derivation of statistics methods. Our study in this paper takes the first step to reveal the possibility of utilizing social media content as a new source to understand user experience of mobile networks.

## 5. CONCLUSION AND FUTURE WORK

We have presented the primary study of exploiting the social network content for network performance monitoring. Our data suggest that users' feedback regarding network incidents is often observed in twitter messages in a timely fashion. Therefore, it is a complementary source for understanding network performance issues and their impact on user experience.

As future work, we plan to apply advanced natural language processing techniques to better understand tweets, as well as tickets and incident reports. For example, it is important to advance the technique for intelligently extracting performance related issues from tweets. It would also be interesting to quantify the severity level of performance related tweets by the scale of the responses and the sentiment in the messages (e.g. to detect urgent issues based on users' attitude).

## 6. REFERENCES

- [1] comScore 2009 US Digital Year in Review, 2009. [http://www.comscore.com/Press\\_Events/Presentations\\_Whitepapers/2010/The\\_2009\\_U.S.\\_Digital\\_Year\\_in\\_Review](http://www.comscore.com/Press_Events/Presentations_Whitepapers/2010/The_2009_U.S._Digital_Year_in_Review).
- [2] Fabricio Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio A. F. Almeida. Characterizing user behavior in online social networks. In *Internet Measurement Conference*, pages 49–62, 2009.
- [3] Daniela Brauckhoff, Xenofontas Dimitropoulos, Arno Wagner, and Kavè Salamatian. Anomaly extraction in backbone networks using association rules. In *Proc. ACM IMC*, 2009.
- [4] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [5] Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 721–730, New York, NY, USA, 2009. ACM.
- [6] Catherine Dwyer, Starr Roxanne Hiltz, and Katia Passerini. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. In *Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS 2007)*, 2007. Paper 339.
- [7] Sanchit Garg, Trinabh Gupta, Niklas Carlsson, and Anirban Mahanti. Evolution of an online social aggregation network: an empirical study. In *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 315–321, New York, NY, USA, 2009. ACM.
- [8] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, New York, NY, USA, 2005. ACM.
- [9] Yiyi Huang, Nick Feamster, Anukool Lakhina, and Jim (Jun) Xu. Diagnosing network disruptions with network-wide analysis. *SIGMETRICS Perform. Eval. Rev.*, 35(1):61–72, 2007.
- [10] Srikanth Kandula, Ranveer Chandra, and Dina Katabi. What is going on? learning communication rules in edge networks. In *Proc. ACM SIGCOMM*, 2008.
- [11] Srikanth Kandula, Ratul Mahajan, Patrick Verkaik, Sharad Agarwal, Jitendra Padhye, and Paramvir Bahl. Detailed diagnosis in enterprise networks. In *Proc. ACM SIGCOMM*, 2009.
- [12] R. R. Kompella, J. Yates, A. Greenberg, and A. C. Snoeren. Detection and localization of network blackholes. In *Proc. INFOCOM*, 2007.
- [13] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *WOSP '08: Proceedings of the first workshop on Online social networks*, pages 19–24, New York, NY, USA, 2008. ACM.
- [14] Haewoon Kwak, Yoonchan Choi, Young-Ho Eom, Hawoong Jeong, and Sue Moon. Mining communities in networks: a solution for consistency and its evaluation. In *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 301–314, New York, NY, USA, 2009. ACM.
- [15] Anukool Lakhina, Mark Crovella, and Christophe Diot. Mining anomalies using traffic feature distributions. In *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 217–228, New York, NY, USA, 2005. ACM.
- [16] A. Mahimkar, Z. Ge, , A. Shaikh, J. Wang J. Yates, Y. Zhang, , and Q. Zhao. Towards Automated Performance Diagnosis in a Large IPTV Network. In *Proc. ACM SIGCOMM*, 2009.
- [17] A. Mahimkar, J. Yates, Y. Zhang, A. Shaikh, J. Wang, Z. Ge, and C. T. Ee. Troubleshooting chronic conditions in large ip networks. In *Proc. ACM CoNEXT*, 2008.
- [18] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, New York, NY, USA, 2007. ACM.
- [19] Marti Motoyama, Brendan Meeder, Kirill Levchenko, Geoffrey M. Voelker, and Stefan Savage. Measuring online service availability using twitter. In *USENIX 3rd Workshop on Online Social Networks (WOSN)*, 2010.
- [20] Atif Nazir, Saqib Raza, Dhruv Gupta, Chen-Nee Chuah, and Balachander Krishnamurthy. Network level footprints of facebook applications. In *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 63–75, New York, NY, USA, 2009. ACM.
- [21] Fabian Schneider, Anja Feldmann, Balachander Krishnamurthy, and Walter Willinger. Understanding online social network usage from a network perspective. In Anja Feldmann and Laurent Mathy, editors, *Internet Measurement Conference*, pages 35–48. ACM, 2009.
- [22] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1079–1088, New York, NY, USA, 2010. ACM.
- [23] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. On the evolution of user interaction in facebook. In *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42, New York, NY, USA, 2009. ACM.