Listen to Me if You can: Tracking user experience of mobile network on social media

Tongqing Qiu[†] Junlan Feng[‡] Zihui Ge[‡] Jia Wang[‡] Jun (Jim) Xu[†] Jennifer Yates[‡]

[†] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
[‡] AT&T Labs – Research, Florham Park, NJ 07901,USA

ABSTRACT

Social media sites like 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 giant 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 traditional source of user experience, i.e. customer care tickets, and correlate both of them with network incident reports. 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, users on twitter are more accurate and faster to report network problems that impact user experiences. Third, tweets can show some short term performance impairments, which are not recorded in incidents reports. These observations prove that twitter a complimentary source for monitoring network performance and their impact on user experiences.

1. INTRODUCTION

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

Online social network (OSN) has gained significant popularity during recent years [18]. Microblogging, like twitter, is one of the popular ways users share information and experience on the web. It's a breakthrough in social networking because it takes communication into another level. Comparing to 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 from 2008 [1].

In this paper, we analyze tweets related to one of the largest mobile service providers in United States. We first identify network performance issues reported on twitter, and compare them with customer care trouble tickets. Second, we correlate these two sources of customer feedback with network incident reports. To the best of our knowledge, this is the first paper to exploit the microblogging content for network performance monitoring. Our findings are threefold: (1) Issues reported on twitter are complimentary to customer care calls. (2) Twitter is faster to report network performance issues compared to customer complaints to the customer care center. (3) Tweets report some short term problems, which are not recorded in incidents reports.

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

2. DATASETS

In this section, we discuss the methodology to collect data and general information of collected data. Section 2.1 presents Twitter data, Section 2.2 presents customer care call data and Section 2.3 shows the 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 are deemed as good queries for retrieving tweets relevant to the mobile service provider we consider. Second, we followed twitter search API to obtain tweets and archived the retrieved tweets along with the associated meta data to our local disk. The meta data includes time that the tweet was submitted, the user who wrote the tweets, etc. Third, we fetched user information through twitter REST APIs for those who authored tweets that we archived. User information consists of user profile such as user location and the user's social connections such as the number of followers.

After data archiving, our next step is to identify tweets related to the 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; (3) Tweets should not contain advertising indicating words like Ads and price symbols $. To verify the effectiveness of these rules , we randomly sampled 100 tweets, and manually annotated whether they are related or not to the mobile performance issues. Comparing rule-based prediction with human annotation, we observed 87% agreement. This is decent accuracy[1] for our study. There are potential methods via natural language processing and machine learning to improve the performance of this step. We will leave it as future work.

2.2 Customer Care Calls

We obtained customer care tickets based on customer calls from the cellular service provider. Note that these tickets are anonymized; no customer identification information is used in this analysis. Moreover, they do not have details of what customer reports to custom care. These tickets are tagged with 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 mainly focus on the customer calls regarding service impairments and outages, each of which results in a trouble ticket associated with the type of service, the time that the call is received by the customer care team (the trouble ticket is often issued at the same time), the location, and description of the performance issues that the customer experiences, etc. 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 uses when the performance impairment occurred.

In our data, the common customer trouble tickets include performance 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 service and report the performance impairments that they experienced. In addition, we will show later that customers may not call the customer service immediately after they experience the performance impairments.

2.3 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 *Network Incident Review* in a collaborative editing system. The network incident review keeps track of information regarding important network or service incidents as they are reported, diagnosed, resolved and concluded. And the review report serves as a channel to communicate summary-level information about such 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 crowd (e.g., highway accident causing traffic congestion and unexpected high cellular network congestion), etc. Reading through the incident 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 upgrade. We have also excluded very long incidents (e.g., greater than three hours) in our 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 incident review report contains crucial information about the event. All entries contain temporal information (i.e. the start and end of time of the incidents), and coarse spatial information, i.e. the primary market region where the incidents occur (e.g northeast, west, etc.). Moreover, all entries have *facilitator/incident manager* contact information, for further investigation purpose, and the expected or estimated scale of customer impact. Most of the entries have *incident summaries* which describe the nature of the incidents in detail [2]. Some of the incidents contain root cause descriptions which explain the cause of the incidents in detail. In this paper, we only rely on the temporal and spatial information of the incidents.

3. RESULTS

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

3.1 Incidents Report Results

We first analyze temporal distribution of incident report. Figure 1 shows the cumulative distribution function (CDF) of incidents duration. Note that the incidents here mean the ones from incident report, rather than all incidents happen in the network. From this figure, we observe a diversity of incident durations given the log scale of X-axis. As we mentioned in Section 2.3, in the later part of our analysis we

[1] In statistical scene, the estimation result 87% is within 95% confidence interval, with a margin error less than 0.1, given that we randomly selection 100 samples.

[2] 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.

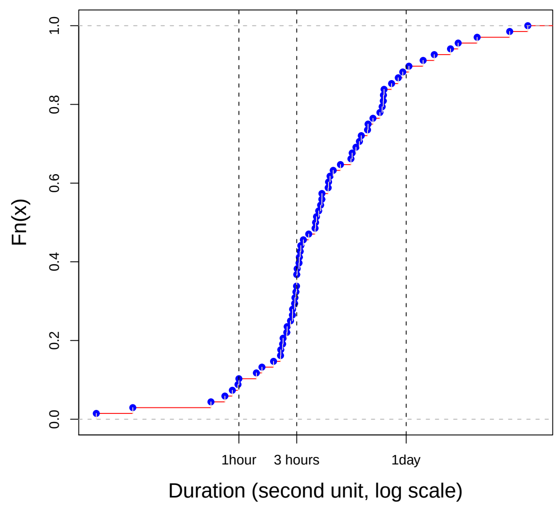

Figure 1: The CDF of incidents duration

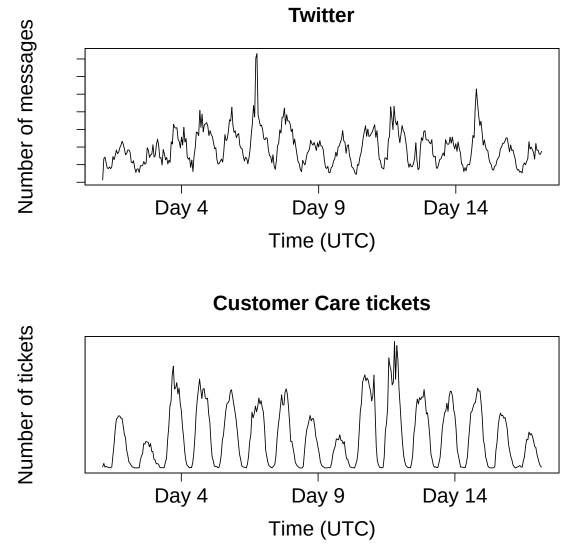

Figure 2: The number of tweets and tickets per hour. Due to privacy issues, the concrete number of tickets is not reported.

focus on the incidents less than three hours.

3.2 Tweets vs. Customer Care Tickets

In this section, we compare two sources of users experience: tweets and tickets based on customer care calls. First, we conduct the comparison using the *raw* data, which include not only performance related ones but all tweets that we have archived and considered as relevant to the mobile service provider. Then we classify the data into different categories and drill down to the performance-related ones. 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 volume of tweets and volume of customer call tickets over the same period of time. Figure 2 shows the number of tweets and tickets per hour based on *raw* data. From Figure 2, we can 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 the service provider in United States. We will later (in Section 3.3) see that such a diurnal pattern will not show up when we only focus on performance related issues from twitter. Another interesting observation of Figure 2 is the spike for twitter data on Day 7. It is because of the discussions on 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 ones.

By manually investigating the message content of twitter, we can observe the following three major types. (1) Comments and news regarding the product and customer service; (2) Advertisement (e.g. the sale of 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, type (1) and (2) contribute to the majority of the tweets. On the other hand, since the tickets have fixed structure and categorization, it is relatively easy to classify them. The majority (over 97 percent) of tickets are related to plan, bill or device questions or problems. In the rest 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, about only 1% messages are related to performance issues. Similarly, only 1% to 2% of tickets are related to network performance. Figure 3 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., difficult to send messages or to post something on websites by phone). Different from tweets, more tickets are related to no service or coverage (e.g. no coverage in some buildings). Very few of them are related to slow connection. Probably it is because different habits of different customers. The customers who prefer to report the issues by call may care less about the Internet connection via mobile device. Moreover, we also find that many tickets are about the voice quality,

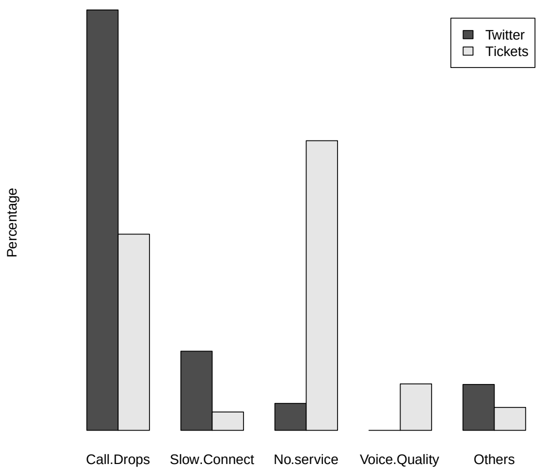

Figure 3: Classify performance related tweets/tickets based on the type of the issues. Due to privacy issues, the concrete percentages are not reported.

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%

which are not seen in twitter messages.

We also compare the tweets and tickets by locations. There are two ways to retrieve the location information of tweets: user profile and message content. The message itself is a more valuable source because it 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 or incomplete. There are three general categories of location information based on our analysis: city + state, state only, others (e.g. city name only, street). Table 1 shows the percentage of each category. We observe that a large proportion (37.2% of profile and 69.5% of twitter messages) of tweets have no location information at all. It reduces the number of valid tweets for later correlation purpose. In contrast, the location information in tickets is well organized based on the *market* regions. Each market could be a couple of states (e.g. Georgia + South Carolina) or a part of one state (e.g. northern California). Different data sources have different granularity of location information. Therefore, we use the most detailed common location for correlation purpose in Section 3.3 .

3.2.3 Timeliness

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

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

There is a delay between the time when the customers experience the incidents and the time when the customers report the incidents. The reporting time is the timestamp to post the twitter message or to make the call to customer care. It can be directly retrieved from the data. On the other hand, the time when customers experience the incidents, if exists, is embedded in the messages of twitter or the description of tickets. For example, in twitter, the message is like “a couple of call drops *today*”; in tickets, the description can be “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 can find that over 90.1% of twitter messages do have such timing information. In comparison, over 76.3% of tickets descriptions have no explicit timing information. Then we compute differences between timestamp to post the message or to report the ticket and the extracted time in the messages/description, illustrated in Table 2. We observe that most tweets are posted on the current date on which the customers experience the performance issues. It means that the response time of twitter is much faster than that of tickets. Note that timing information is very important for verifying or detecting incidents. In this sense, twitter is better than customer care tickets.

3.3 Correlation Results

In this part, we correlate both performance related twitter and tickets data to incidents report. Figure 4 shows the time series of incidents, tweets and tickets. Different from Figure 2, performance related tweets have no obvious diurnal pattern shown in Figure 4. It implies that these twitter messages are incident-driven. But customer tickets still have the diurnal pattern, because of the customers’ calling behavior and aforementioned delay in incident reporting.

To understand if there is strong correlation between incident reports and customer feedback, we conduct the statistic correlation [12] between incidents and twitter/tickets series, for each location. We compute Pearsons coefficient of correlation and conduct significance test. 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 problem. Correction of this time lag by some time shifting should be applied for correlation test. 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 as the future work.

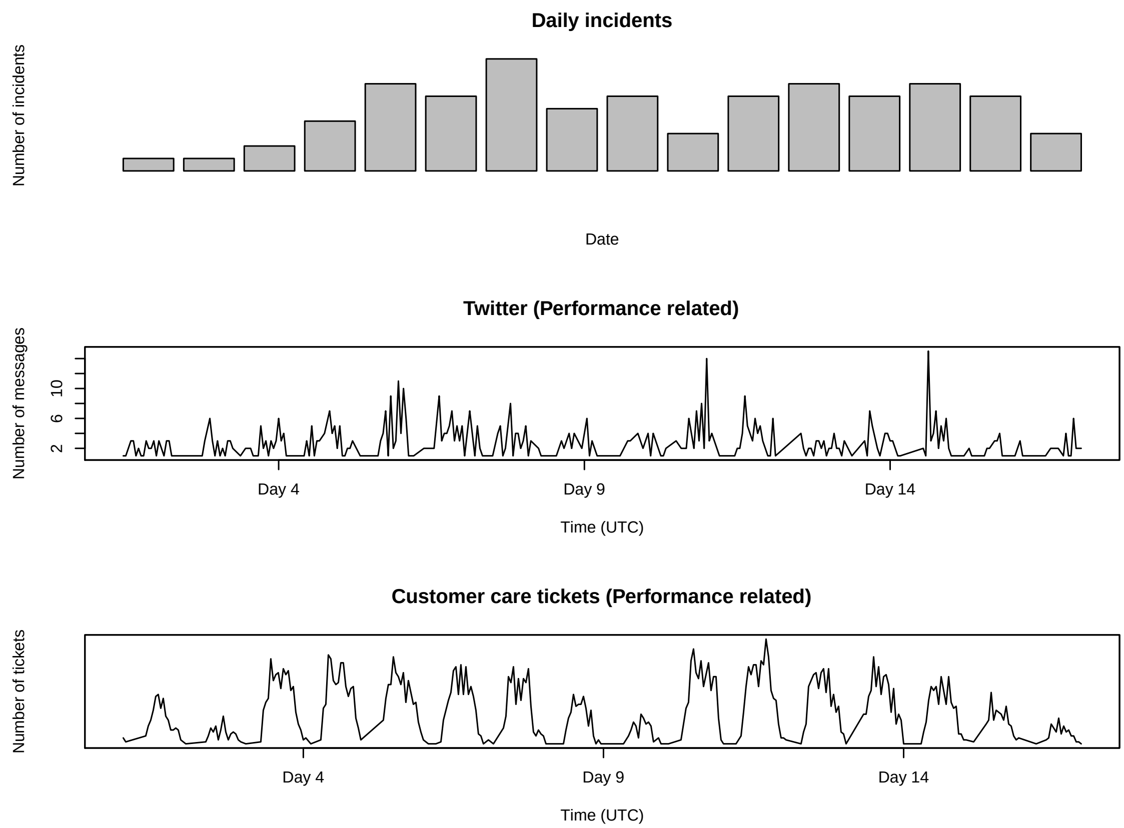

Figure 4: The time series of incidents, twitter and tickets. Due to privacy issues, the concrete numbers of incidents and tickets are not reported.

Instead of using statistical correlation, we use the incidents in the report as the ground truth and investigate whether these incidents are reported by twitter/tickets data. More specifically, if during the period of incident, there are messages or tickets regarding the performance issues at the same region (e.g. east, west, central, etc.), we deem these tweets or tickets *associated* with the incidents. We verify the incidents which last less than 3 hours. The reason to verify relative 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. We find that 55.6% of incidents can be found in twitter, and only 37.0% of them found in customer care tickets.[3]. 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 ratio c_1/c_2 is 8.3 for tweets and 6.8 for tickets. It suggests that chance of having user feedback under incidents is significantly higher than that under no ongoing incidents.

In Section 3.2.3 we quantitatively study 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 problem. However, the time when users experience the incidents is not an accurate timestamp, as we 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 problem. Figure 5 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. We find that tweets respond approximately 10 minutes faster than tickets in general. The fastest twitter response is in several minutes. The implication of this observation is that it may be possible to utilize twitter feedback to observe the impact of network per-

[3]Due to limited coarse location information, there are some potential mismatch cases or false positive.

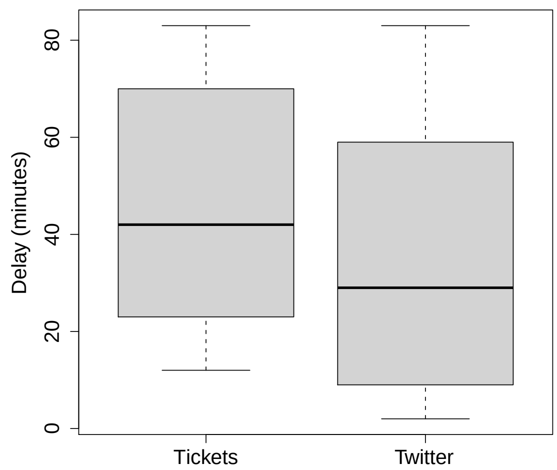

Figure 5: The CDF of delay distribution.

formance issues more timely than through customer tickets.

Finally, let us revisit Figure 4. There are obvious spikes on Day 11 for both twitter and customer care data, even though there 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 to incident reports covering the area.

4. RELATED WORK

There have been a number of studies regarding the social network. But most of them focus on the social network itself, e.g., users behavior [18, 13, 19, 2], the impact on network performance [20], community evolution [7, 14], information propagation [22, 5, 4], privacy issues [8, 6]. Very few studies focus on showing the value of social network content. One interesting example is [21]. In that paper, Vieweg *et al.* shows the microblogging like Twitter can contribute to the situation awareness during natural hazards events like flood and fire.

Correlating different type of 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 one new source to understand user experience of mobile network.

5. CONCLUSION AND FUTURE WORK

We have presented the first study of exploiting the social network content for network performance monitoring. Our data shows that users' feedback from twitter can respond to network incidents in a timely fashion. Therefore, it is a complimentary source for understanding network performance issues and their impact on user experience.

As future work, we plan to apply advanced techniques of natural language process 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 issue 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`.

[2] Fabrício 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. Whats 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] 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.

[20] 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.

[21] 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.

[22] 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.