

Hsu, Wenling; Jacobsen, Guy; Jin, Yu; Skudlark, Ann

Conference Paper

Using social media data to understand mobile customer experience and behavior

22nd European Regional Conference of the International Telecommunications Society (ITS): "Innovative ICT Applications - Emerging Regulatory, Economic and Policy Issues", Budapest, Hungary, 18th-21st September, 2011

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Hsu, Wenling; Jacobsen, Guy; Jin, Yu; Skudlark, Ann (2011) : Using social media data to understand mobile customer experience and behavior, 22nd European Regional Conference of the International Telecommunications Society (ITS): "Innovative ICT Applications - Emerging Regulatory, Economic and Policy Issues", Budapest, Hungary, 18th-21st September, 2011, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/52180>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Wenling Hsu, Guy Jacobson, Yu Jin, Ann Skudlark
Using Social Media Data to Understand Mobile Customer Experience and Behavior

ABSTRACT

Understanding mobile customer experience and behavior is an important task for cellular service providers to improve the satisfaction of their customers. To that end, cellular service providers regularly measure the properties of their mobile network, such as signal strength, dropped calls, call blockage, and radio interface failures (RIFs). In addition to these passive measurements collected within the network, understanding customer sentiment from direct customer feedback is also an important means of evaluating user experience. Customers have varied perceptions of mobile network quality, and also react differently to advertising, news articles, and the introduction of new equipment and services. Traditional methods used to assess customer sentiment include direct surveys and mining the transcripts of calls made to customer care centers.

Along with this feedback provided directly to the service providers, the rise in social media potentially presents new opportunities to gain further insight into customers by mining public social media data as well. According to a note from one of the largest online social network (OSN) sites in the US [7], as of September 2010 there are 175 million registered users, and 95 million text messages communicated among users per day. Additionally, many OSNs provide APIs to retrieve publically available message data, which can be used to collect this data for analysis and interpretation.

Our plan is to correlate different sources of measurements and user feedback to understand the social media usage patterns from mobile data users in a large nationwide cellular network. In particular, we are interested in quantifying the traffic volume, the growing trend of social media usage and how it interacts with traditional communication channels, such as voice calls, text messaging, etc. In addition, we are interested in detecting interesting network events from users' communication on OSN sites and studying the temporal aspects—how the various types of user feedback behave with respect to timing. We develop a novel approach which combines burst detection and text mining to detect emerging issues from online messages on a large OSN network. Through a case study, our method shows promising results in identifying a burst of activities using the OSN feedback, whereas customer care notes exhibit noticeable delays in detecting such an event which may lead to unnecessary operational expenses.

JEL codes: C89, L11

Keywords: Mobile customer experience, social media, text data mining, customer feedback

Wenling Hsu, hsu@att.com

Yu Jin, yjin@research.att.com

Guy Jacobson, guy.jacobson@att.com

Ann Skudlark, skudlark@att.com (corresponding author)

AT&T Labs, 180 Park Avenue, Florham Park, NJ 07932

1. INTRODUCTION

Social media, due to its low cost and ease of access, has been experiencing tremendous growth. There are hundreds of different types of online social network (OSN) sites which attract a large number of users who share common interests or backgrounds, such as business users, active gamers, music fans and classmates, etc., and enable them to communicate with each other through online messages, real-time chat, blogs, and micro blogs. The largest OSN site claims to have 750 million registered users worldwide [4].

There have been numerous general studies and models of behavior on online social networks [1,3,5,6,7]. However, little attention has been paid to how customers use their mobile devices in a cellular network for their online social network activities. Due to the rapid growth of mobile data user population and the increasing popularity of smartphone devices coupled with established mobile applications (*apps*), accessing OSN sites through mobile devices can be as convenient or even more convenient, than from personal computers. From the perspective of a cellular service provider, we are interested in understanding the importance and growing trend of social media usage by mobile data users. In particular, *what are the most popular social media activities in a nationwide cellular network? How much traffic is contributed by these mobile OSN users?* Since social media activities represent a new facet of mobile communication behavior in comparison to the more traditional communication modes such as voice, text messaging (SMS), and email, *how different are they from these traditional communication channels?* Moreover, as a subset of mobile data users, *what are the unique characteristics of these mobile social media users? Can social media activities provide unique insights in detecting emerging events that have an impact on cellular network services?*

To answer these questions, we study the social media usage in one of the largest cellular networks in the US over a 14 - month time span. By correlating data collected from various sources, such as network measurements, customer care notes, messages from one of the largest OSN sites, and so forth, we isolate the four most popular OSN sites in the network and determine that mobile users associated with these sites grew rapidly in the past year. Even though these mobile users only account for about 6% of the overall user population, they access network services more intensively and contribute to more than 14% of the data traffic. Our study also indicates that mobile users are spending more time on these OSN sites and are relying on social media as a complementary or even a replacement for voice, SMS and other traditional communication channels. Therefore, understanding and meeting the demands of social media usage is important to enhance the satisfaction of a growing mobile social media user population.

Communication of users on OSN sites (represented as messages or MSGs in short) can be an informative source for understanding the emerging user demands. In the second part of the paper, we propose a novel approach for extracting network events from MSGs collected from a large OSN site. Our approach consists of two steps: (1) burst detection identifies MSGs with strong temporal concentration (2) text-mining reveals the root cause of these bursts and provides useful and actionable insights to the cellular content providers regarding the emerging user demands. We illustrate through a case study that a set of messages posted to an OSN site can lead to increases in call volume to customer care centers for weeks. By detecting such a burst of scam MSGs on the OSN site, cellular service providers can take appropriate action to reduce the associated operational cost of increased calls to the care centers.

2. BACKGROUND AND DATASETS

Cellular Network Overview: The cellular network under study uses UMTS (Universal Mobile Telecommunication System), a popular 3G mobile communication technology supporting both voice and data services. Fig. 1 depicts the key components in a typical UMTS network. The UMTS network has a hierarchical structure: where each Radio Network Controller (RNC) controls multiple node-Bs, and one Serving GPRS Support Node (SGSN) serves multiple RNCs (see [2] for detailed description of UMTS networks).

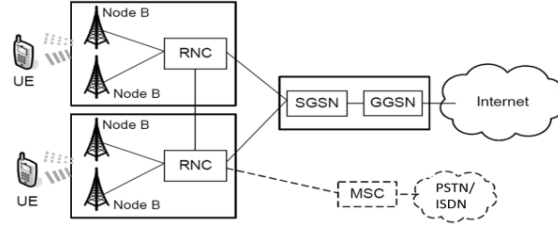


Figure 1: UMTS network architecture

Social Media Channels: In this paper, we focus primarily on four of the most popular social media channels.

Datasets: Our study uses data collected in the UMTS network for three dates during a 14 month period. No customer private information was used in the analysis and all customer identities were anonymized before any analysis was conducted. Moreover, to adhere to the company confidentiality standards, at times we present normalized views of our results while retaining the scientifically relevant magnitudes in this paper.

3. OVERVIEW OF SOCIAL MEDIA USAGE FROM MOBILE USERS

In this section, we present an overview of social media usage in the mobility network under study. We first introduce the four most popular OSN sites for the analysis. By studying users associated with these sites, we highlight the significance of mobile social media users within the entire mobile user population. By investigating usage on these four sites from a large cellular data network over a 14 - month time span, we pinpoint the fast growth of mobile OSN user population and the overall OSN usage using mobile devices.

3.1 Evolution of Mobile Social Media Usage

Given the large number of OSN sites, collecting traffic from all OSN sites is impractical. Instead, we focus our analysis primarily on four popular OSN sites in the network. Both the number of users and the amount of data associated with these four sites are a magnitude higher than any other OSN site. We summarize these four sites below. For proprietary reasons, we do not use the actual names of these sites.

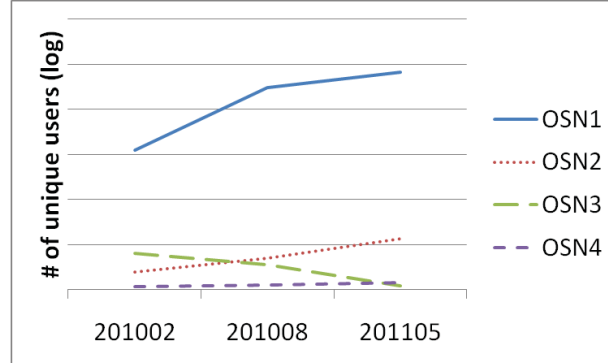


Figure 2 Evolution of overall mobile social media usage

OSN1 is a social networking site which claims to have acquired 750 million users. Initially it was restricted to college students – but that requirement was lifted in 2006. *OSN2* provides social networking and messaging services like Short Message Service (SMS), allowing users to send messages (referred to as MSGs) to one or many other users (followers). *OSN3* was the most popular US networking site in 2006, and focused on social networking among musicians and their audiences. Its popularity has declined as customers substitute other OSN sties. In contrast to the three social websites above, *OSN4* is a business-oriented social networking site, which focuses on business contacts potential business opportunities.

To assess the evolution of the OSN usage among mobile subscribers, we compare the total number of unique users accessing each OSN at least once during each of our observation dates (one day in Feb 2010, August 2010 and May 2011, all three of which are Tuesdays) using a cellular data network. The results can be seen in Figure 2. Note that the y axis is in log scale. The trend line of OSN1, the most popular OSN of the four sites studied, indicates fast growth rate during the first 6 month of our observation, followed by a slower growth rate for the

next 8 months. OSN2, although it has only one eighth of the popularity of OSN1, displays a steady growth throughout our observation. The trend line for OSN3 exhibits an opposite phenomenon -- decreasing in popularity in the first 6 month and then a faster decrease rate in the next 8 months. OSN4 has only 2% of OSN1's popularity, and displays a small but steady growth.

These social media users constitute only a small subset of mobile users. Using a two-week data set from August 2010, we found that users of these four OSN sites account for around 6% of all the mobile data users. However, these 6% of all users contributed to more than 14% of the overall data traffic in the network. This implies that these OSN users access the mobility data service more intensively than most of other data users. Given the fast growth trend of mobile OSN users, we expect to see a more significant amount of traffic from mobile OSN users in future. As a consequence, understanding social media user behavior is important for service providers to meet the potential usage demands from these users and, as we shall see in later sections of this paper, to utilize these social media users as valuable information source for detecting emerging network issues and allocating appropriate resources to resolve these issues to help control operational costs.

3.2 Intensity of Mobile Social Media Usage

Accompanying the growth of the social media user population, we also witness an increase of the intensity of social media usage. Figure 3 (a) shows the average number of bytes per user for the four OSN sites under study. The usage of OSN1 and OSN2 grew much faster in the second 8-month time period than in the first 6 months. It is likely due to both the increasing popularity of multimedia content on these sites, such as large images and video clips, and the fact that mobile users spend more time on these sites.

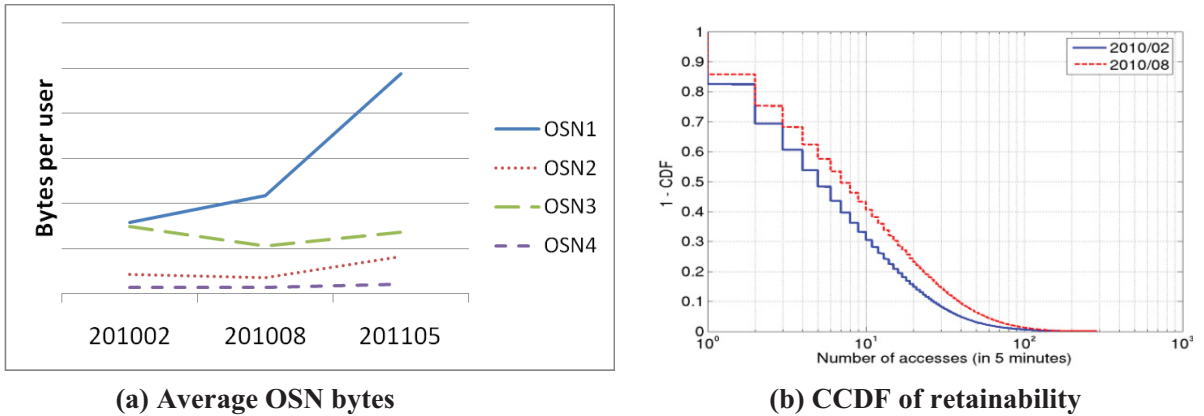


Figure 3 Evolution of overall mobile social media usage

To justify the second point, we measure the retainability of users on these sites, which measures the overall time that a user spends on a particular OSN site. Given the fact that our network measurements are collected at a 5-minute granularity, we define retainability as the number of 5 minute intervals that a user is observed to have accessed that site. This implies that the maximum retainability that each user can have for a particular OSN per day is 288 $((24 \times 60) / 5)$. We demonstrate in Figure 3 (b) the complementary cumulative distribution function (CCDF) of the retainability for OSN1 for two different days in February 2010 and August 2010, respectively. We observe a clear increase of retainability for the OSN1 users, indicating that users are observed to spend more time accessing OSN1 after 6 months. For example, around 30% users have a retainability of 10 in February 2010, while 40% of the users are observed to have a retainability of 10 in August 2010.

3.3 Temporal Properties of Social Media Usage

We now compare mobile social media users and other mobility users in terms of their temporal properties. In particular, we are interested in understanding at what time of day people tend to access social media sites and how such usage is affected by users' daily activities and the nature of the social media sites.

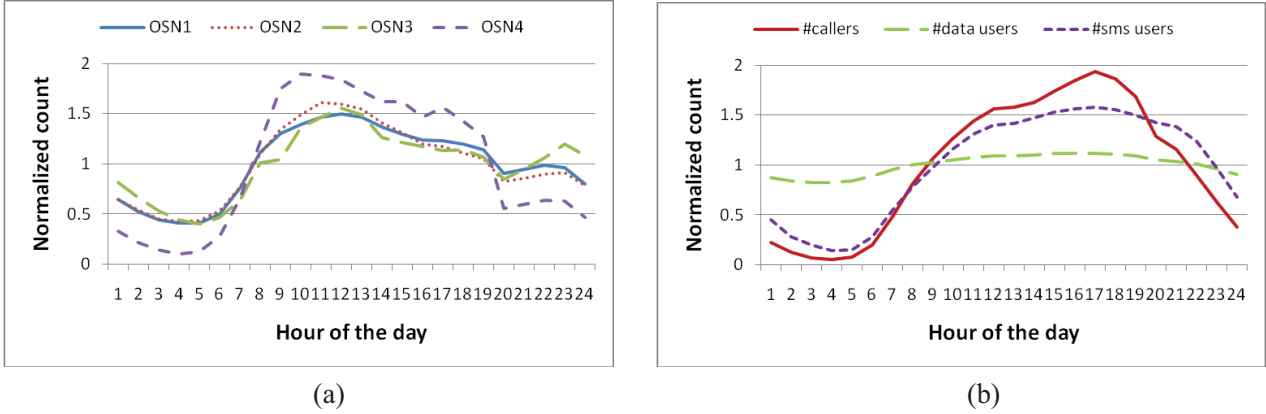


Figure 4 Temporal properties of social media usage vs. mobile service usage.

We show in Figure 4 (a) the number of unique users who access each of the four popular social media sites in each hour of a particular day. Note that each curve has been normalized by the corresponding mean value in order to highlight the variability of each type of social media usage. We observe that the usage on all four sites exhibits a strong time-of-the-day effect, where more active users are observed during the day time than at night. Among these sites, OSN4 usage has the strongest variability throughout the 24-hour day. This is not surprising, since OSN4 is a business oriented social media site, fewer business activities are expected during the night time. In contrast, OSN3 usage shows the least variability, and we conjecture this is because fans are more likely to access OSN3 during off-business hours for news regarding their favorite musicians and albums. The usage curves of OSN1 and OSN2 largely overlap. Even at night, there are still many users communicating through these sites.

For the purpose of comparison, we illustrate in Figure 4 (b) the number of unique users (again normalized by the mean value of each curve) who access voice, data and SMS services every hour in the entire day. A similar time-of-the-day effect is observed on these three mobile service usages. However, even though social media usage is a subcategory of data usage, its usage pattern is strikingly different from the overall data usage, which actually shows little difference between day and night. In fact, the social media usage patterns resemble most the patterns of voice and SMS. This implies that, unlike other data services, access to social media sites is affected greatly by daily user activities and OSNs serve as a complement or even an alternative to traditional communication channels among mobile users. OSNs are closer in spirit to interpersonal communications such as traditional voice and text messages, whereas many other forms of data usage on cellular network are impersonal such as web browsing, video streaming and file downloading. As a consequence, it is possible that interactions within OSNs encourage mobile users to consume more data and may have higher performance demands. Social media usage can affect mobile users' experience in mobile data services. We shall discuss this in detail in the next section.

4. CHARACTERISTICS OF SOCIAL MEDIA USERS

In this section, we study the unique characteristics of social media users by correlating various data sources, such as their voice usage, SMS usage, overall byte usage, and complaint rates.

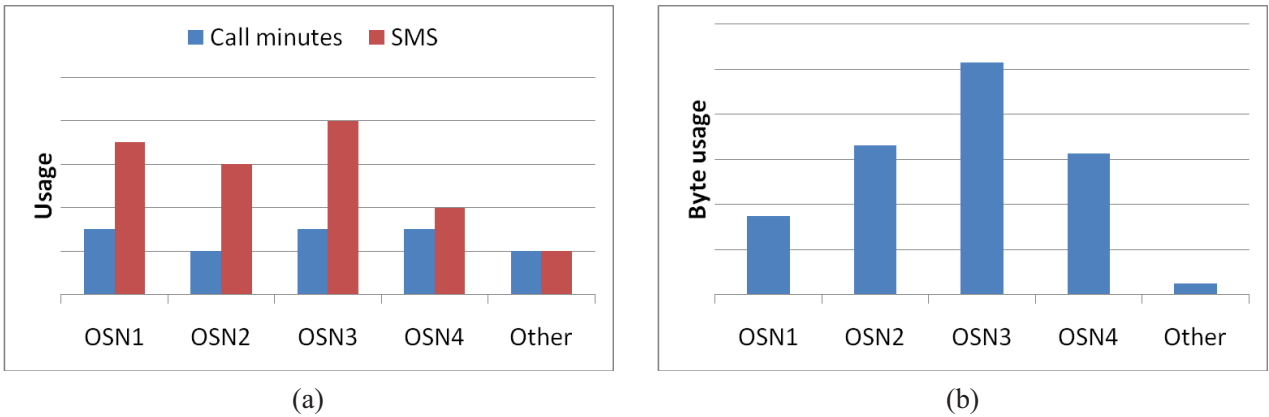


Figure 5 Voice, SMS and data usages from social media users and other data users.

We compare social media users with other data users in terms of *total call minutes per user*, *SMS messages per user* and *bytes per user* in a particular day. Data users are defined as the set of users who have observed data traffic during the observation time period. In general, social media users display more intensive usage in all three services. Studies show that these social media users are more socially active people and are high communicators – which may contribute to their more intensive usage among all three services. In addition, most of the users of OSN1, OSN2 and OSN3 are younger users [9], who tend to communicate with their friends more through SMS messages and via social media sites, however, their voice usage is not significantly higher compared to other data users. One exception is the OSN4 users. Since most of them are business users, they show a much lower SMS usage but a slightly higher voice usage in comparison to other social media users. An interesting observation is that OSN2 users send a relatively small number of SMS messages. Investigation shows that there is a statistically significant negative correlation between OSN2 usage (i.e., the amount of OSN2 traffic) and the SMS usage. This is likely because OSN2 share similar functionality as traditional SMS service, and can propagate information towards a large number of other users (followers) with little cost. Therefore, this can potentially cause migration of SMS usage or OSN2 usage and part of our ongoing work is to study and predict the trend of such migration in the near future.

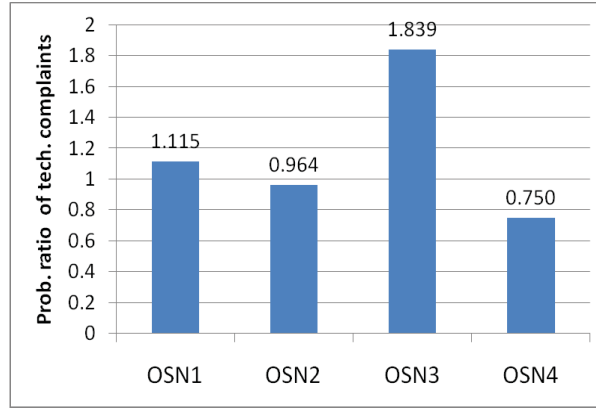


Figure 6 Technical complaint rates from different social media users w.r.t. other data users.

We next investigate how social media usage affect the users' experience of network service. We note that since our samples of mobile users are uniform across the entire US and little geographical concentration is observed for social media users, we assume that, at the same network location, different social media users are experiencing similar network performance. Under this assumption, the technical complaint rates, i.e., the number of users who called the customer support line to report technical problems (e.g., a dropped call, a connectivity issue or a network coverage issue, etc.), are expected to be similar across all social media user populations. We show the technical complaint rates in Figure 6, where they are normalized by the complaint rate of other data users. OSN3 is observed to be associated with a 80% higher complaint rate, which is likely because OSN3 users tend to access many video/audio clips and hence are more sensitive to network performance changes. In contrast, OSN4 users show a much lower complaint rate, possibly due to the greater focus on text rather than multi-media. Additionally, in other work we conducted on complaint rates by device we observed that business users with the identical phone models as consumers had lower complaint rates. We conjecture that these business users often have technical support teams in their firms which can solve many technical problems without the need to contact the cellular service provider and complain. The OSN2 users also complain less. As observed by [6], many complaints are sent via MSGs on OSN2 instead of going through the customer care channels.

5. DETECTING EMERGING NETWORK ISSUES IN SOCIAL MEDIA DATA

In [6], Qiu et. al. demonstrated the potential for detecting network-level problems based on emerging popular topics on social media sites. In particular, they showed that by tracking the bursts of keywords in MSGs related to performance issues (e.g., connectivity, slow, outage, etc.), a number of network problems can be detected. However, their technique is limited because it requires domain knowledge on the keywords to be tracked.

Therefore, the detection results are also limited to a few specific performance issues. In fact, social media sites provide us with far more information that can help service providers understand the demands of mobile users and allocate appropriate resources to meet such demands. To address such limitation, in this section, we propose a statistical approach to extract interesting social events from MSGs and we walk through this method with a case study of detecting a social media propagated phenomenon in MSGs on OSN2.

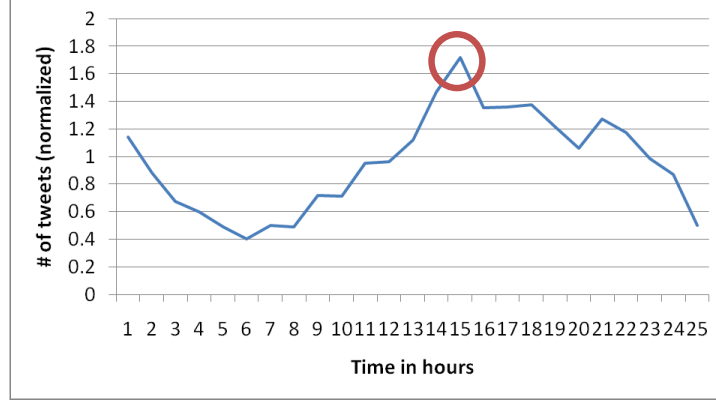


Figure 7 Number of MSGs per hour in a particular day, the red circle indicates the detected burst.

The proposed method consists of two steps: burst detection and text mining. In the burst detection step, we focus on detecting the time periods (hours) when a significant number of MSGs are observed. We note that any advanced burst detection method can be applied for this step. As a proof of concept, we adopt a simple approach as follows. We assume the number of MSGs received in each hourly time bin follows a normal distribution $N(\mu, \sigma^2)$ and we calculate the mean and standard deviation for each bin based on the historical data. A burst is detected if the number MSGs observed in a time bin exceeds $\mu + 1.96\sigma$. We note that due to the high volume of MSGs, we only look at the MSGs containing the keyword X (company name). Figure 8 shows the number of MSGs received in each hour of a particular observation day and we illustrate the detected burst around 3PM EST with a red circle.

After detecting the bursts, we next apply text mining techniques to determine the root cause of each burst, i.e., the emerging popular topic that drives each burst. A simple way of ranking topics based on the word frequencies (or a similar function of the frequency, e.g., tf-idf) does not work well here since there are quite a number of redundant but less informative keywords in MSGs and it is difficult to enumerate and filter them with a stop list. Instead, we explore the correlation across keywords. In particular, we measure the importance of co-occurring keyword pairs based on the tf-idf score. Let $p_i(a, b)$ be a pair of keywords a and b , and n_{ij} stands for the frequency of p_i in MSG d_j . The term frequency (tf) is defined as $tf_{ij} = n_{i,j} / \sum_k n_{kj}$, note that $n_{i,j}$ often equals 1 given a MSG can only contain a maximum of 140 characters. For a MSG d_j with k words, tf_{ij} is often the reciprocal of the number of keyword pairs, i.e., $tf_{ij} = 1/(k(k-1))$. Let D denote the corpus of MSGs in a particular time bin. The inverse document frequency is defined as $idf_i = \log(|D| / |\{j : p_i \in d_j\}|)$, where $|\{j : p_i \in d_j\}|$ represents the number of MSGs that contain the keyword pair p_i . Therefore, $tfidf_{ij} = tf_{ij} \times idf_i$.

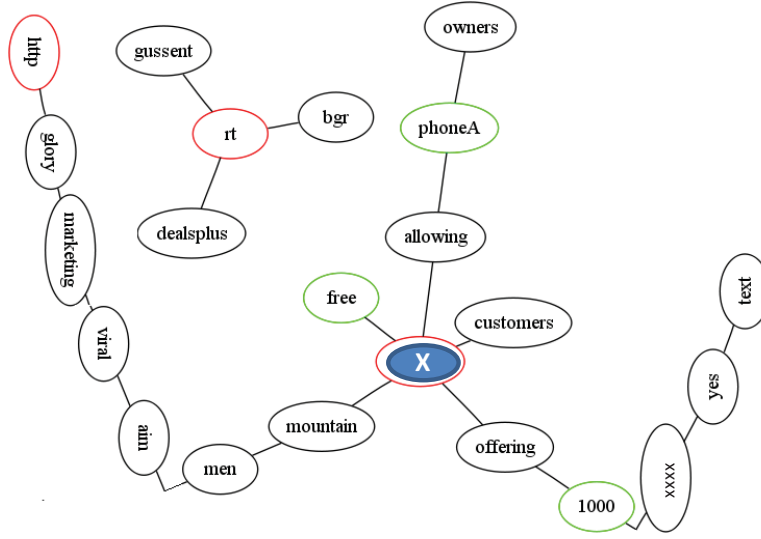


Figure 8 Visualizing the most dominant keyword pairs in the MSG burst at 3PM EST.

After ranking all the co-occurring keyword pairs, we visualize the top 20 dominant pairs using a keyword graph. In the graph, nodes represent keywords and an edge between two keywords indicates that the corresponding keyword pair is dominant in the data. We use red and green to highlight the keywords with high frequencies (red keywords are more frequent than green ones and both are more frequency than black keywords). The keyword graph is illustrated in Figure 8, which is drawn using the Graphviz tool [10]. We find interesting keywords like *free*, *1000*, *offering*, etc. Investigation on the raw data shows that many of the MSGs are reporting that Company-X offers 1000 free voice minutes after texting to Company-X with the short code xxxx. The results from mining customer care notations also show an increase in calls from customers asking why they did not receive the offer from Company-X after texting Company-X. (The actual reason is that only certain customers were eligible, but the MSGs reporting this offer on OSN2 did not mention this condition). In fact, the increased load on customer care to handle calls related to this issue lasted for several days, and over ten thousand such calls were received each day. This case study implies that OSN2 can have a wide influence on mobility users and enables fast propagation of information (compliments, complaints, rumors, information, and misinformation). Detecting emerging trends from MSGs using the proposed method can help service providers allocate appropriate resources in time to meet new user demands. For example, in this case, service providers can use social media or email, or update the interactive voice response (IVR) system to advise customers and to clarify the situation. Using these methods could eliminate calls to customer care thereby helping to control operational costs.

6. CONCLUSIONS AND FUTURE WORK

In this paper we correlated different sources of measurements and user feedback to understand the social media usage patterns from mobile data users in a large nationwide cellular network. Our analysis showed that the population of social media users has been growing rapidly in the past year, and mobile users are accessing these OSN sites more regularly. We also found that mobile social media users accessed different cellular network services more intensively and hence were more sensitive to network performance changes. Many of them relied on OSN sites as a complement to or even a replacement for traditional communication. They shared network usage experiences, expressed their complaints, and spread information (and misinformation) on OSN sites. We proposed a novel approach to gain insight by mining information from messages exchanged on a large OSN site, and we successfully identified an event on that site which resulted in an increase in call volume to customer care centers.

7. REFERENCES

- [1] **Cha , Meeyoung, H. Haddadi, F. Benevenuto, K. Gummadi** Measuring User Influence in Twitter: The Million Follower Fallacy. [Conference] // 4th International AAAI Conference on Weblogs and Social Media. - Washington, DC : [s.n.], 2010.
- [2] **Jin , Yu, N. Duffield, A. Gerber, P. Haffner, W. Hsu, G. Jacobson, S. Subhabrata, S. Venkataraman, Z. Zhang** Making Sense of Customer Tickets in Cellular Networks [Conference] // The 30th IEEE International Conference on Computer Communications. - Shanghai, China : IEEE INFOCOM, 2011.
- [3] **Krishnamurthy, Balachander, P. Gill, M. Arlitt** A Few Chirps About Twitter [Conference] // First Workshop on Online Social Networks. - NY, NY : ACM, 2008. - pp. 19-24.
- [4] **Kopytoff, Verne.** (July 7, 2011). Facebook offers video chat in arrangement with Skype. *The New York Times*, p. B3.
- [5] **Mislov, Alan, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee** Measurement and Analysis of Online Social Networks [Conference] // 7th ACM SIGCOMM Conference on Internet Measurement. - NY, NY : ACM, 2007.
- [6] **Qiu, Tongqing, J. Feng, Z. Ge, J. Wang, J. Zu, J. Yates** Listen to Me if You can: Tracking user experience of mobile network on social media [Conference] // ACM/USENIX Internet Measurement Conference. - Melbourne, Australia : ACM, 2010.
- [7] **Schneider, Fabian, A. Feldmann, B. Krishnamurthy, W. Willinger** Understanding Online Social Network Usage from a Network Perspective [Conference] // Internet Measurement Conference. - [s.l.] : ACM, 2009.
- [8] Twitter Facts [Online] // Twitter. - September 14, 2010. - February 3, 2011. - www.twitter.com/about.
- [9] comScore 2010 U.S: U.S. Demographic Profiles at Facebook and Twitter
<http://www.comscoredata.com/2011/02/us-demographic-profiles-at-facebook-and-twitter/>
- [10] Graphviz - Graph Visualization Software, <http://www.graphviz.org/>.