Rank	Observation Time Frame		
	1 day	3 days	7 days
1	Interact-F9 (0.15)	Post-F5 (0.27)	Post-F5 (0.46)
2	Interact-F11 (0.12)	Trend-F19 (0.18)	Post-F6 (0.31)
3	Interact-F10 (0.11)	Post-F6 (0.18)	Trend-F19 (0.28)
4	Interact-F12 (0.11)	Interact-F9 (0.16)	Post-F1 (0.27)
5	Trend-F18 (0.05)	Post-F1 (0.16)	Post-F7 (0.23)
6	Interact-F15 (0.04)	Post-F7 (0.13)	Trend-F20 (0.21)
7	Post-F1 (0.04)	Interact-F15 (0.12)	Interact-F15 (0.21)
8	Interact-F8 (0.04)	Interact-F11 (0.12)	Post-F2 (0.19)

Table 3: The top 8 feature and its categories ranked by information gain (values shown in parentheses).

cally, we rank features based on *Information Gain* [18], which measures feature's distinguishing power over the two classes of data. We list the top 8 features in Table 3. As expected, prediction power varies significantly, and information gain drops off quickly (particularly for 1 day) after the top 4 features. To validate their prediction power, we repeat each experiment with only their top 4 features. The results in Figure 18 show that the top 4 features achieve most of the accuracy of the entire classifier, but with much less complexity.

Then we take a closer look at the top features. First, we note that the 1-day classifier relies on different set of features compared with 3- and 7-day classifiers. The 1-day models rely heavily on *interaction features*. Intuitively, the model predicts whether a user will stay engaged based on how actively the user participates in social interactions. If a user received many replies or actively replied to others on her first day, there's a high chance for this user to stay longer. For 3- and 7-day models, we find that the key features shift to user's *content posting* and *activity trend* features. This means once we monitor the users for a longer period, the user's intention to stay or leave can be more accurately reflected in her posting frequency and volume, and whether that activity is declining over time.

Engaging Users with Notifications. Stimulating user engagement is a key goal for any new service. One tool Whisper has already deployed is push notifications that deliver the "whisper of the day" to users' mobile device every evening between 7 and 9pm. The exact notification time varies each day and between Android and iOS devices. To examine the impact of these notifications, we conduct a small experiment. We monitor the notification time on 5 different phones every day for 6 days. We look at user activity in the Whisper stream for 5 minute and 10 minute intervals following the notifications, and find no statistically significant increase in new replies or whispers compared to other 5 or 10 minute windows between 7 and 9pm. This means that while these notifications may serve to engage users to read popular whispers, there is no significant increase in new whispers or replies as a result.

6. CONTENT MODERATION IN WHISPER

Anonymity facilitates free speech, but also inevitably fosters abusive content and behavior [21,35]. Like other anonymous communities, Whisper faces the same challenge of dealing with abusive content (*e.g.*, nudity, pornography or obscenity) in their network. In addition to a crowdsourcing-based user reporting mechanism, Whisper also has dedicated employees to moderate whispers [16]. Our basic measurements (§3.2) also suggest this has a significant impact on the system, as we observed a large volume of whispers (>1.7 million) has been deleted during the 3 months of our study. The ratio of Whisper's deleted content (18%) is much higher than traditional social networks like Twitter (<4%) [1,30].

Topic	**Top 50 Keywords Most Related to Deleted Whispers**
Sexting (36)	sext, wood, naughty, kinky, sexting, bj, threesome, dirty, role, fwb, panties, vibrator, bi, inches, lesbians, hookup, hairy, nipples, freaky, boobs, fantasy, fantasies, dare, trade, oral, takers, sugar, strings, experiment, curious, daddy, eaten, tease, entertain, athletic
Selfie (7)	rate, selfie, selfies, send, inbox, sends, pic
Chat (7)	f, dm, pm, chat, ladys, message, m
Topic	**Top 50 Keywords Least Related to Deleted Whispers**
Emotion (17)	panic, emotions, argument, meds, hardest, fear, tears, sober, frozen, argue, failure, unfortunately, understands, anxiety, understood, aware, strength
Religion (10)	beliefs, path, faith, christians, atheist, bible, create, religion, praying, helped
Entertain. (8)	episode, series, season, anime, books, knowledge, restaurant, character
Life story (6)	memories, moments, escape, raised, thank, thanks
Work (5)	interview, ability, genius, research, process
Politics (1)	government
Others (3)	exactly, beginning, example

Table 4: Topics of top and bottom 50 keywords related to whisper deletion.

In this section, we take a closer look at content deletions in Whisper. First, we analyze the content of deleted whispers to infer the reasons behind deletions. Second, we analyze the lifetime of deleted whispers to understand how fast do whispers get deleted. Third, we focus on authors of deleted whispers and compare their behavior to the norm.

Before we begin, we note that while users can delete their own whispers, we believe server-side content moderation is responsible for the large majority of missing whispers in our data. Intuitively, users who reconsider and later delete their own whispers are likely to do so within a relatively short time frame. In contrast, our "deleted" dataset comes from our followup crawl for replies, which runs once a week. In fact, since our main crawler on the latest stream runs every 30 minutes, we expect most self-deleted whispers will not even show up in our core dataset.

Content Analysis of Deleted Whispers. To explore the reasons behind deletion, we analyze the content of deleted whispers. Since whispers are usually very short, Natural Language Processing (NLP) tools do not work well (we confirmed via experiments). Thus we take a keyword-based approach: we extract keywords from all whispers and examine which keywords correlate with deleted whispers. First, before processing, we exclude common stopwords[4] from our keyword list. Also to avoid statistical outliers, we exclude low frequency words that appear in less than 0.05% of whispers. Then for each keyword, we compute a *deletion ratio* as the number of deleted whispers with this keyword over all whispers with this keyword. We rank keywords by deletion ratio, and examine the top and bottom keywords.

We run this analysis on all 9 million original (not including replies) whispers in our dataset, 1.7M of which are later deleted. This produces 2324 keywords ranked by deletion ratio. We list the top and bottom 50 keywords in Table 4 and classify them manually into topic categories. Not surprisingly, many deleted whispers violate Whisper's stated user policies on sexually explicit messages and nudity. In contrast, topics related to personal expression, religion, and politics are least likely to be deleted.

[4] `http://norm.al/2009/04/14/list-of-english-stop-words`

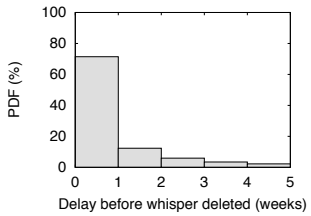

Figure 19: Deletion speed (coarse-grained).

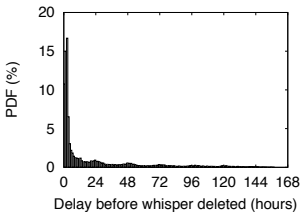

Figure 20: Deletion speed (fine-grained).

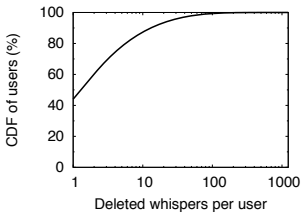

Figure 21: # of Deleted whispers per user.

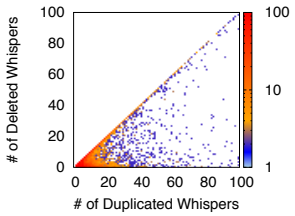

Figure 22: Duplicated vs. deleted whispers.

Deletion Delay. Next we analyze the deletion delay of whispers, *i.e.* how long do whispers stay in the system before they are deleted? Recall that our reply crawler works once a week, and thus detects deleted whispers on the granularity of once a week. As shown in Figure 19, the majority (70%) of deleted whispers are "deleted" within one week after posting. A small portion (2%) of whispers have stayed for more than a month before deletion. Since most whispers lose user attention after one week (Figure 5), we believe these deletions are not the results of crowdsourcing flagging, but deleted by Whisper moderators.

To get a more fine grain view of whisper deletions, we perform a period of frequent crawls on a small set of whispers. On April 14, 2014, we select 200K new whispers from our crawl of the latest whisper stream, and check on (recrawl) these whispers every 3 hours over a period of 7 days. Of the 200K whispers, 32,153 whispers are deleted during our monitoring period (a week). The more fine-grained distribution of the lifetime (hourly) of these whispers is shown in Figure 20. We find the peak of whisper deletion to be between 3 and 9 hours after posting, and the vast majority of deletions happen within 24 hours of posting. This suggests that the moderation system in Whisper works quickly to flag and remove offensive whispers. However, it is unclear whether this level of responsiveness is sufficient, since user page views focus on the most recent whispers, and moderation after 3 hours is possibly too late to impact the content most users see.

Characterizing Authors of Deleted Whispers. Finally, we take a closer look at the authors of deleted whispers to check for signs of suspicious behavior. In total, 263K users (25.4%) out of all users in our dataset have at least one deleted whisper. The distribution of deleted whispers is highly skewed across these users: 24% of users are responsible for 80% of all deleted whispers. The worst offender is a user who had 1230 whisper deleted during the time period of our study, while roughly half of the users only have a single deletion (Figure 21).

We observed anecdotal evidence of duplicate whispers in the set of deleted whispers. We find that frequently reposted duplicate whispers are highly likely to be deleted. Among our 263K users with at least 1 deleted whisper, we find 25K users have posted duplicate whispers. In Figure 22, we plot each user's number of duplicated whispers versus the number of deleted whispers. We observe a clear clustering of users around the straight line of $y = x$. This indicates that when users post many duplicated whispers, there's a higher chance that most or all duplicated whispers are deleted.

We also observe that authors of deleted whispers change their nicknames more often than the average user. Figure 23 shows the distribution of total number of nicknames used by each user. We categorize users based on how many deletions they have, and also include a baseline of users with 0 deletions. We find users with no deletion rarely change their nicknames, if ever, but nickname changes occur far more frequently for users with many deleted whispers. We speculate that perhaps users change their nickname to avoid being flagged or blacklisted. Since users cannot see their own GUID when using the app, they may assume the system identifies them using only their nickname.

7. TRACKING WHISPER USERS

In the final component of our Whisper study, we take a close look at a vulnerability that exposes detailed location of Whisper authors to the system. In practical terms, this attack allows a Whisper user to accurately track (or potential stalk) another Whisper user through whispers they've written, by writing simple scripts that query Whisper servers. This attack demonstrates the inherent risks to user privacy in mobile applications, even for apps that target user anonymity as a core goal. Note that we met the Whisper team in person and informed them of this attack. They are supportive of this work, and have already taken steps to remove this vulnerability.

In this section, we describe details of this location tracking attack. The attack makes use of Whisper's "nearby" function, which returns a list of whispers posted nearby, attaching a "distance" field to each whisper. The attack generates numerous "nearby" queries from different vantage points, and uses statistical analysis to reverse engineer the whisper author's location. We validate the efficacy of this attack through real-world experiments.

7.1 Pinpointing User Locations

We start by describing the high-levels of the attack: when a user (*i.e.* the victim) posts a new whisper, he exposes his location to the Whisper server. An attacker in an nearby area can query the nearby list to get their "distance" to the whisper author. The methodology is simple: the attacker can move to different (nearby) locations and query the nearby list for the distance to the victim. Using multiple distance measurements, the attacker can *triangulate* the whisper author's location. The fact that Whisper does not authenticate location in its queries makes this easier, an attacker can issue numerous distance queries from different locations all while sitting in the comfort of her living room.

With a bit more effort, an attacker can even track the victim's movement over time, by triangulating his location every time he posts a whisper. In practice, this means the attacker can physically go and stalk the victim. While the effective error is roughly 0.2 miles (details below), it is more than sufficient to infer the victim's movement to specific points of interest. Considering most Whisper users are young adults or teenagers [4], this attack can lead to severe consequences.

Distance Granularity and Errors. Implementing this attack is nontrivial. Whisper's design team has always been aware of location tracking risks to its users, and built in basic defense mechanisms into the current system. First, they apply a distance offset to every whisper, so the location stored on their servers is always

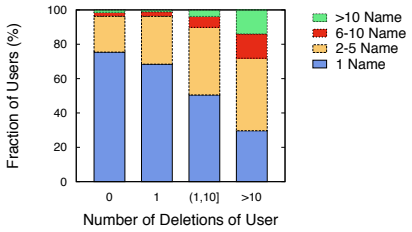

Figure 23: User's number of deletions vs. number of nicknames.

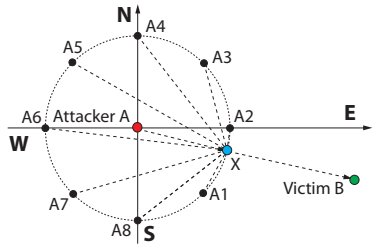

Figure 24: Estimating the distance and direction to the victim.

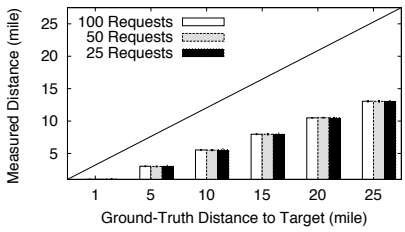

Figure 25: True distance vs. measured average distance (>1 mile).

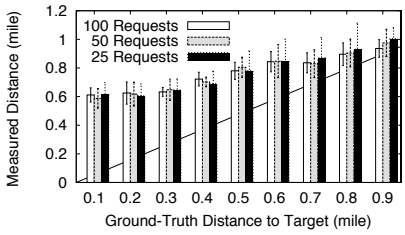

Figure 26: True distance vs. measured average distance (within 1 mile).

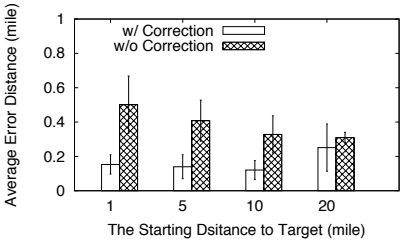

Figure 27: The final error distance of the attack.

Figure 28: Number of hops to approach the victim.

off by some distance to the actual author location. Second, the distance field returned by the nearby function is a coarse-grained integer value (in miles). This was a recent change made by Whisper in February 2014, before which the nearby function returned distances with decimal values. Third, Whisper server adds a random error to the answer to each query, *i.e.* when we query the nearby list repetitively from the same location, each query returns a different distance for the same whisper. The specific error function is unknown.

Attack Details. To accurately pinpoint a user location, our approach is to extensively measure the "distance" from different vantage points, and use large-scale statistics to infer user's location. Specifically, our attack exploits a key property of Whisper: servers allow anyone to query the nearby list with arbitrarily self-reported GPS values as input, and impose no rate limits on such queries. This effectively helps us to overcome the limitations (*i.e.* random error, coarse granularity) on the returned distance. First, we can reduce or eliminate per-query noise by taking the average distance across numerous queries from the same observation location. Second, even though the absolute distance is still not accurate, we can estimate the *direction* to the victim based on the measurements from different locations. Then with distance and direction, an attacker can repeat the measurement from a location closer to the victim, thus iteratively deducing the victim's real location.

We use a simple example to illustrate how this works. Suppose user A (attacker) finds user B (victim)'s whisper in the nearby list, and A wants to pinpoint B's location:

1. A queries the nearby list to get its current distance (d) to victim B (averaged across multiple queries).
2. To estimate the direction, A needs additional observation points. We pick 8 points $\{A_1, A_2, ...A_8\}$ evenly distributed on a circle centered at A with radius d (Figure 24). From each point, A queries the nearby list to measure its distance to victim $\{d_1, d_2, ..., d_8\}$. Suppose X is a dot on the circle, then objective function $Obj = \sqrt{\frac{\sum_{i=1}^{8}(|\overrightarrow{A_iX}|-d_i)^2}{8}}$ reaches the minimum if $\overrightarrow{AX}$ is the right direction to the victim.
3. Then the attacker moves to the next location using $\overrightarrow{AX}$ and d, and repeats step 1 and 2. The algorithm terminates if $d < Thre_1$, or the distance d from two consecutive rounds differs $< Thre_2$.

In practice, the attacker can script all queries with forged GPS values and does not need to physically move.

Distance Error Correction. Finally, we introduce a final step that uses physical measurements to calibrate and add an additional "correction" factor to location data.

We first post a target whisper at a predefined physical location L (on UCSB campus). Then we measure distances to L using the nearby list from a set of observation points, each with known ground-truth distances to L. The ground-truth distance ranges cover from 1 to 25 miles (in 5 mile increments) and again from 0.1 to 0.9 miles (in 0.1-mile increments). At each increment, we use 8 observation points (as specified above) and use each to query the nearby list 100 times. Figure 25 and Figure 26 plot the ground-truth distance versus the measured distance (for 25, 50 and 100 requests per location). For distances greater than 1 mile, we find that our estimates underestimate true physical distance to the victim. Within 1 mile, it clearly overestimates. This mapping between true and measured distance serves as a guide for generating our "correction factor," which is applied to the final estimate.

7.2 Experimental Validation of the Attack

A Single-target Experiment. We first post a whisper at a pre-defined location on UCSB campus as the target (victim). Then we run the attack algorithm starting from distances of 1, 5, 10 and 20 miles away from the victim. Our algorithm takes the average distance over 50 queries per location, and terminates when the estimated distance from consecutive rounds differ < 0.1 mile or when estimated distance < 0.5 mile (based on Figure 26). We repeat each experiment 10 times and test the performance with and without our distance *error correction factor*. Results are shown in Figure 27 and Figure 28.

We make two key observations. First, the algorithm is very accurate. The final error distance, *i.e.* distance from the estimated victim location to the ground-truth location, is only 0.1 to 0.2 miles. With a radius of 0.2 miles, attackers can already effectively identify user's significant points of interest (*e.g.*, home, work, shopping mall) and reconstruct a victim's daily routine using mobility traces [3]. Second, the results show that distance error correction improves algorithm accuracy significantly and reduces the number of iterations needed to determine the victim's location.

Geographically Diverse Targets. To make sure our results are not biased and specific to a single location, we apply the correction factor computed from local measurements (Figure 25 and Figure 26) to carry out attacks in different cities. More specifically, we post target whispers in Santa Barbara and Seattle Washington, Denver Colorado, New York City, New York and Edinburgh Scotland. All whispers are posted via an Android phone with forged GPS coordinates. Then we run the algorithm with distance error correction. We find the final error distances are consistently less than 0.2 miles, and that our correction factor can be generalized to improve estimation accuracy regardless of geographic region.

7.3 Countermeasures

This type of statistical attack cannot be mitigated simply by adding more noise into the system. Attackers can always apply increasingly sophisticated statistical and data mining tools to eliminate noise and determine the true location of a whisper. Instead, the key is to restrict user access to extensive distance measurements. This means putting more constraints (*e.g.*, rate limits) on queries to the nearby list. For instance, one approach is to enforce per-device rate limits. Another is detect fake GPS values, either by relying on client hardware (difficult) or by detecting "unrealistic" movement patterns by potential attackers. Finally, the ultimate defense is to simply remove the "distance" field altogether. While the Whisper engineering team has already addressed this issue, we are not aware of the specific steps they took to do so.

8. RELATED WORK

Online Social Networks. Over the last few years, researchers have performed measurement studies on online social networks (OSNs) including Facebook [36,39], Twitter [8,25], Pinterest [12], and Tumblr [9]. Today's OSNs have stored large volumes of sensitive data about users (*e.g.*, personal profile, friending information, activity traces), all of which pose potential privacy risks. Various techniques have been proposed to compromise user anonymity and infer users' sensitive information from social network data [5,26, 27,44]. Our study focuses on anonymous social networks, which prioritize user privacy at the cost of eliminating persistent identities as well as social links.

Anonymous Online Communities. Anonymous online services allow users to post content and communicate without revealing their real identity. Researchers have studied various anonymous platforms including anonymous forums [32], discussion boards [6, 23] and Q&A sites [21]. Most earlier works study user communities focusing on content and sentiment analysis. More recently, anonymous social networks have emerged, particularly on mobile platforms. A recent work [31] conducted a user survey on SnapChat to understand how they used the anonymous social app. In comparison, our study is the first to quantitively study user interaction, user engagement, and security implications in the anonymous Whisper network.

Device Localization. Our attack algorithm to localize Whisper users is inspired by existing techniques used for device localization in wireless (mobile) networks [15,20,43]. We differ from existing techniques in our approach to deal with the random errors injected by Whisper server. Also, our contribution is more on identifying and validating the security vulnerability instead of the localization algorithm itself.

9. CONCLUSION AND FUTURE WORK

Anonymous, mobile-only messaging apps such as Whisper mark a clear shift away from traditional social networks and towards privacy-conscious communication tools. To the best of our knowledge, our study is the first large data-driven study of social interactions, user engagement, content moderation and privacy risks on the Whisper network. We show that without strong user identities or persistent social links, users interact with random strangers instead of a defined set of friends, leading to weak ties and challenges in long-term user engagement. We show that even in anonymous messaging apps, significant attacks against user privacy are very feasible. We believe that this shift towards privacy in communication tools is here to stay, and insights from our study on Whisper provides value for developers working on next generation systems in this space.

Whisper is not only a social communication tool, but also a network for sharing anonymous content. Analysis and modeling of topics and sentiments in Whisper would be interesting topics for future work. For example, whether and how do users establish communities around "topics" or "themes"? How can anonymous posts and conversations impact user sentiment and emotions? How does user behavior on Whisper compare to those of existing content networks such as Digg and Quora?

Acknowledgments

We would like to thank our shepherd Alan Mislove and the anonymous reviewers for their comments. This project was supported in part by NSF grants IIS-1321083, CNS-1224100, IIS-0916307, by the DARPA GRAPHS program (BAA-12-01), and by the Department of State. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

10. REFERENCES

[1] ALMUHIMEDI, H., WILSON, S., LIU, B., SADEH, N., AND ACQUISTI, A. Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proc. of CSCW* (2013).

[2] ANDREESEN, M. Public tweets. Twitter, March 2014.

[3] ASHBROOK, D., AND STARNER, T. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput. 7*, 5 (2003), 275–286.

[4] ASSOCIATED PRESS. Whispers, secrets and lies? anonymity apps rise. USA Today, March 2014.

[5] BACKSTROM, L., DWORK, C., AND KLEINBERG, J. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proc. of WWW* (2007).

[6] BERNSTEIN, M. S., MONROY-HERNÁNDEZ, A., HARRY, D., ANDRÉ, P., PANOVICH, K., AND VARGAS, G. G. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *Proc. of ICWSM* (2011).

[7] BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R., AND LEFEBVRE, E. Fast unfolding of communities in large networks. *JSTAT 2008*, 10 (2008).

[8] CHA, M., HADDADI, H., BENVENUTO, F., AND GUMMADI, K. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proc. of ICWSM* (2010).

[9] CHANG, Y., TANG, L., INAGAKI, Y., AND LIU, Y. What is tumblr: A statistical overview and comparison. *CoRR abs/1403.5206* (2014).

[10] CLAUSET, A., SHALIZI, C. R., AND NEWMAN, M. E. Power-law distributions in empirical data. *SIAM review 51*, 4 (2009), 661–703.

[11] GARCIA, D., MAVRODIEV, P., AND SCHWEITZER, F. Social resilience in online communities: The autopsy of friendster. In *Proc. of COSN* (2013).

[12] GILBERT, E., BAKHSHI, S., CHANG, S., AND TERVEEN, L. "i need to try this!": A statistical overview of pinterest. In *Proc. of CHI* (2013).

[13] GILBERT, E., AND KARAHALIOS, K. Predicting tie strength with social media. In *Proc. of CHI* (2009).

[14] GONG, N. Z., XU, W., HUANG, L., MITTAL, P., STEFANOV, E., SEKAR, V., AND SONG, D. Evolution of social-attribute networks: measurements, modeling, and implications using google+. In *Proc. of IMC* (2012).

[15] GONZALEZ, M. A., GOMEZ, J., LOPEZ-GUERRERO, M., RANGEL, V., AND OCA, M. M. GUIDE-gradient: A guiding algorithm for mobile nodes in wlan and ad-hoc networks. *Wirel. Pers. Commun. 57*, 4 (2011).

[16] GROVE, J. V. Secrets and lies: Whisper and the return of the anonymous app. CNet News, January 2014.

[17] GUO, L., TAN, E., CHEN, S., ZHANG, X., AND ZHAO, Y. E. Analyzing patterns of user content generation in online social networks. In *Proc. of KDD* (2009).

[18] GUYON, I., AND ELISSEEFF, A. An introduction to variable and feature selection. *JMLR 3* (2003), 1157–1182.

[19] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The weka data mining software: an update. *SIGKDD Explor. Newsl. 11*, 1 (2009).

[20] HAN, D., ANDERSEN, D. G., KAMINSKY, M., PAPAGIANNAKI, K., AND SESHAN, S. Access point localization using local signal strength gradient. In *Proc. of PAM* (2009).

[21] HOSSEINMARDI, H., HAN, R., LV, Q., MISHRA, S., AND GHASEMIANLANGROODI, A. Analyzing negative user behavior in a semi-anonymous social network. *CoRR abs/1404.3839* (2014).

[22] JONES, J. J., SETTLE, J. E., BOND, R. M., FARISS, C. J., MARLOW, C., AND FOWLER, J. H. Inferring tie strength from online directed behavior. *PLoS ONE 8*, 1 (2013), e52168.

[23] KNUTTILA, L. User unknown: 4chan, anonymity and contingency. *First Monday 16*, 10 (2011).

[24] KWAK, H., CHOI, Y., EOM, Y.-H., JEONG, H., AND MOON, S. Mining communities in networks: a solution for consistency and its evaluation. In *Proc. of IMC* (2009).

[25] KWAK, H., LEE, C., PARK, H., AND MOON, S. What is Twitter, a social network or a news media? In *Proc. of WWW* (2010).

[26] MISLOVE, A., VISWANATH, B., GUMMADI, K. P., AND DRUSCHEL, P. You are who you know: inferring user profiles in online social networks. In *Proc. of WSDM* (2010).

[27] NARAYANAN, A., AND SHMATIKOV, V. Robust de-anonymization of large sparse datasets. In *Proc. of IEEE S&P* (2008).

[28] NEWMAN, M. E. Modularity and community structure in networks. *PNAS 103*, 23 (2006), 8577–8582.

[29] NEWMAN, M. E. J. Assortative mixing in networks. *Physical Review Letters 89*, 20 (2002), 208701.

[30] PETROVIC, S., OSBORNE, M., AND LAVRENKO, V. I wish i didn't say that! analyzing and predicting deleted messages in twitter. *CoRR abs/1305.3107* (2013).

[31] ROESNER, F., GILL, B. T., AND KOHNO, T. Sex, lies, or kittens? investigating the use of snapchat's self-destructing messages. In *Proc. of FC* (2014).

[32] SCHOENEBECK, S. Y. The secret life of online moms: Anonymity and disinhibition on youbemom.com. In *Proc. of ICWSM* (2013).

[33] STRAPPARAVA, C., AND VALITUTTI, A. Wordnet affect: an affective extension of wordnet. In *Proc. of LREC* (2004).

[34] STUTZMAN, F., GROSS, R., AND ACQUISTI, A. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality 4*, 2 (2013).

[35] SULER, J., AND PHILLIPS, W. L. The bad boys of cyberspace: Deviant behavior in a multimedia chat community. *Cyberpsy., Behavior, and Soc. Networking 1*, 3 (1998), 275–294.

[36] UGANDER, J., KARRER, B., BACKSTROM, L., AND MARLOW, C. The anatomy of the facebook social graph. *CoRR abs/1111.4503* (2011).

[37] WAKITA, K., AND TSURUMI, T. Finding community structure in mega-scale social networks: [extended abstract]. In *Proc. of WWW* (2007).

[38] WATTS, D. J., AND STROGATZ, S. Collective dynamics of 'small-world' networks. *Nature*, 393 (1998), 440–442.

[39] WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K., AND ZHAO, B. User Interactions in Social Networks and Their Implications. In *Proc. of EuroSys* (2009).

[40] WORTHAM, J. New social app has juicy posts, all anonymous. NY Times, March 2014.

[41] WORTHAM, J. Whatsapp deal bets on a few fewer 'friends'. NY Times, February 2014.

[42] XU, T., CHEN, Y., JIAO, L., ZHAO, B. Y., HUI, P., AND FU, X. Scaling microblogging services with divergent traffic demands. In *Proc. of Middleware* (2011).

[43] ZHANG, Z., ZHOU, X., ZHANG, W., ZHANG, Y., WANG, G., ZHAO, B. Y., AND ZHENG, H. I am the antenna: Accurate outdoor AP location using smartphones. In *Proc. of MobiCom* (2011).

[44] ZHELEVA, E., AND GETOOR, L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proc. of WWW* (2009).