2016 IEEE Tenth International Conference on Semantic Computing

Analysis of Cyber Aggression and Cyber-bullying in Social Networking

Tadashi Nakano[1], Tatsuya Suda[2], Yutaka Okaie[1], and Michael John Moore[3]
[1]*Graduate School of Frontier Biosciences, Osaka University, Japan*
[2]*University Netgroup Inc., Fallbrook, CA, USA*
[3]*University of California, San Diego, CA, USA*
Email: tadasi.nakano@fbs.osaka-u.ac.jp, tatsuyasuda@gmail.com,
yokaie@fbs.osaka-u.ac.jp, mjm022@ucsd.edu

***Abstract*—This paper considers Ask.fm, a social networking site where users create profiles and can send each other questions, and analyses aggressive user behavior that may potentially lead to cyber-bullying incidents. We hypothesize that anonymity is a primary cause of such aggressive user behavior and examine how anonymous and non-anonymous users behave in social networking. We collected data from Ask.fm and analyzed questions posted by anonymous and non-anonymous users and answers posted by non-anonymous users. Analysis of the collected data shows that anonymous users exhibit more aggressive behavior than non-anonymous users. Analysis also shows that users become more aggressive in answering aggressive anonymous questions than aggressive non-anonymous questions.**

***Keywords*-cyber-bullying; anonymity; social networking; user behavior**

I. Introduction

Cyber-bullying is a form of cyber-aggression that has become a major concern in today's information society [1]. Cyber-bullying is characterized as actions that abuse information and communication technology to attack victims intentionally, repeatedly and through an imbalance of power. Types of cyber-bullying include flaming (heated exchange between individuals), harassment (repetitive and aggressive intimidation of a victim), denigration (disseminating derogatory and untrue information about a victim), and impersonation (stealing victim's online identity and posing as a victim to communicate embarrassing information regarding the victim to others). These types of cyber-bullying occur through a variety of information and communication technologies such as instant messaging, e-mail, text messaging, social networking sites, blogs, and web sites.

Cyber-bullying incidents have been increasingly reported in online social networking sites, and research is being carried out to develop techniques that help automatically detect cyber-bullying [2]. In developing such techniques, many researchers apply machine-learning approaches to identify negative words and to learn language patterns used by cyber-bullies and victims in their messages with the ultimate goal of detecting possible occurrence of cyber-bullying [3], [4], [5], [6], [7]. Many researchers also apply a graph theoretical approach to obtain a graph theoretic measure (Hyperlink-Induced Topic Search) in order to help identify possible cyber-bullies and victims [8]. Further, research is also being carried out to develop software tools that help users of social networking sites intervene cyber-bullying incidents at early stages of cyber-bullying or and prevent cyber-bullying incidents before they may occur [5], [6].

This paper focuses on an online social networking site, Ask.fm, with the goal of understanding user behavior that may potentially lead to cyber-bullying incidents on social networking sites. Ask.fm is a social networking site where users interact with each other anonymously without identifying themselves or non-anonymously by identifying themselves. Based on findings in existing research that anonymity encourages aggressive user behavior in online forums [9], [10], we hypothesize that in Ask.fm, anonymity is a primary cause of the cyber-bullying incidents, and analyze how anonymity in Ask.fm is associated with aggressive user behavior that may potentially lead to cyber-bullying incidents.

The rest of the paper is organized as follows. Section II describes methods applied to collect data from Ask.fm, and Section III describes results obtained from the analysis of the collected data. Finally, Section IV provides a summary and future directions of the work presented in this paper.

II. Methods

A. Data Source

We chose Ask.fm to study the user behavior in social networking for the following reasons: Ask.fm has a large number of users (150 million users as of February 2015), there is an increasing number of cyber-bullying incidents reported on Ask.fm [11], and Ask.fm supports anonymous user interactions and serves as an ideal site to investigate the impact of anonymity on user behavior in social networking.

Ask.fm provides a question and answer-based social networking service that allows its users to interact with each other. In Ask.fm, each user has a profile that stores user's personal information such as a text description and a photograph of the user. A profile of Ask.fm user is publicly available to all Internet users including those who do not have an account with Ask.fm. In asking a question, an Ask.fm user (X) visits another Ask.fm user (Y)'s profile and leaves a question for Y to answer. If Y answers X's

978-1-5090-0662-5/16 $31.00 © 2016 IEEE
DOI 10.1109/ICSC.2016.111

question, X's question and Y's answer are both posted on Y's profile and become visible to the public (i.e., all Internet users including all Ask.fm users and non Ask.fm users). If Y does not answer the question, the question remains invisible to the public. X may leave a question on Y's page anonymously without revealing his/her Ask.fm user name, if Y's profile setting allows anonymous users to leave a question, or non-anonymously revealing his/her Ask.fm user name, unless Y's profile setting specifically blocks X from leaving a question. The default setting of a user profile in Ask.fm is to accept questions from anonymous users and not to block any specific non-anonymous users; however, a user may change the setting anytime to not accept questions from anonymous users and to block specific non-anonymous users from leaving a question.

B. Data Collection

We collected data from Ask.fm user profiles. We first selected a user profile and collected all question-answer pairs from the profile. We then examined all collected question-answer pairs from the selected user and visited all users who are found as "askers" (i.e., users who asked a question) in the question-answer pairs to collect all question-answer pairs from each of the askers. We repeated this process and visited user profiles in a breadth first search manner, collecting all recent question-answer pairs up to 10,000 question-answer pairs per profile.

After collecting data, we performed preprocessing of the collected data for analysis. The preprocessing includes (1) removing all question-answer pairs that are written in languages other than English (using the language detection tool [12]), (2) removing all question-answer pairs where questions are generated through the "get a random question" option of Ask.fm[1], and (3) removing images, videos, emoticons, and emoji in all question-answer pairs. The preprocessing also includes, for texts in each question and in each answer, counting the number of negative words that it contains. (The number of negative words in a question and in an answer is used in our analysis as a measure of aggressive user behavior.) A publicly available list of 349 negative words is used for this purpose [13]. Each question-answer pair is then converted into a data record containing (1) the record ID (for our internal reference purposes), (2) Ask.fm user name of an asker (i.e., a user who asked a question), (3) the number of negative words in the question, (4) Ask.fm user name of the answerer (i.e., a user who answered a question), (5) the number of negative words in the answer, and (6) the time (year, month, day, hour and minute) when the question-answer pair is posted. Note that as Ask.fm allows an asker to post a question anonymously, (2) is only available when the asker posts a message non-anonymously. Note also that, as the question-answer pair is posted on answerer's profile, the answerer is always non-anonymous.

[1] "Get a random question" is an option available to Ask.fm users. By using this option, an Ask.fm user may ask other Ask.fm users general questions such as "which do you prefer, fish or meat?" For this research, we used this option to generate approximately 1,400 distinct random questions and removed, from the collected data, all question-answer pairs where a question matches a random question.

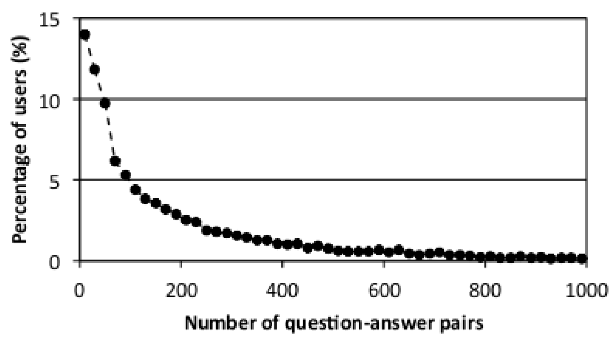

Figure 1. Distribution of the number of question-answer pairs per user profile

After the preprocessing, the data set contains 2,458,895 records from 9,778 user profiles; each user profile has an average of 251 question-answer pairs with an average of 28 unique askers. Fig. 1 shows the distribution of the number of question-answer pairs per user profile averaged over all user profiles we collected.

C. Data Collection Method Limitations

The following aspects of the data collection method employed in this paper may impact the data analysis and results presented in this paper.

- The dataset we collected only contains question-answer pairs where questions were answered. Questions that were not answered are not publicly available, and thus, we were not able to include them in the dataset.
- In this paper, anonymous users are defined as Ask.fm users who leave questions without revealing their Ask.fm user names, and non-anonymous users as those who answer questions or leave questions revealing their Ask.fm user names. Ask.fm user name may be user's real name, a pseudonym, or a user name being used for impersonation, and may or may not reveal the true identity of the user. In addition, an individual may use multiple Ask.fm user names. In this paper, however, each Ask.fm user name is considered to represent a distinct user.

III. Data Analysis

The goal of the data analysis descried in this section is to examine our hypothesis that anonymity is a primary cause of aggressive user behavior that may potentially lead to cyber-bullying incidents on Ask.fm. We first examine how frequently Ask.fm users answer anonymous questions. We

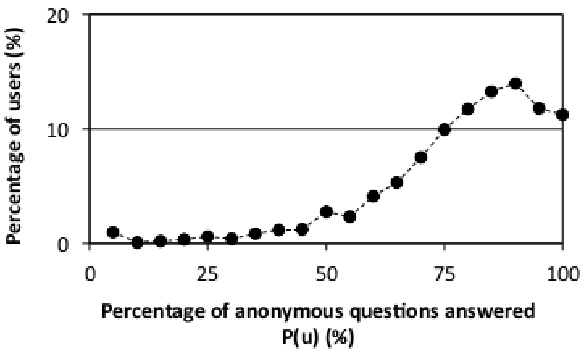

Figure 2. Percentage $P(u)$ of anonymous questions answered and the percentage of users

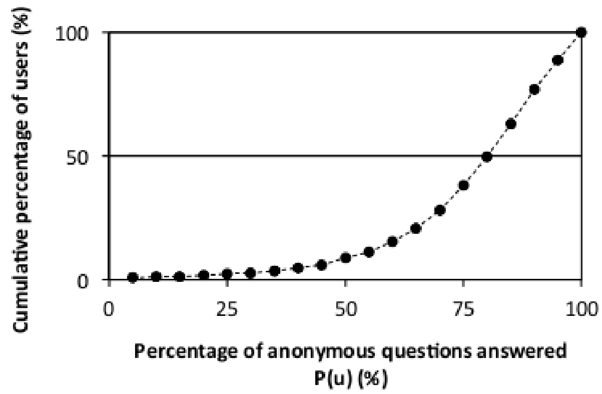

Figure 3. Percentage $P(u)$ of anonymous questions answered and the cumulative percentage of users

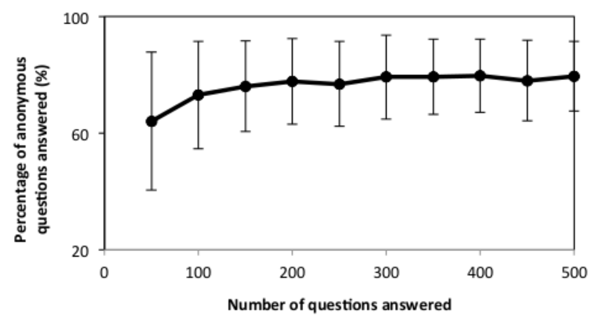

Figure 4. Changes in the percentage $P(u)$ of anonymous questions answered. Error bars show standard deviation.

then examine the degree of aggressiveness of anonymous and non-anonymous questions that an answerer answered through counting the number of negative words used in anonymous and non-anonymous questions.

A. Frequency of Answering Anonymous Questions

We first examine how often Ask.fm users answer anonymous questions. Based on the data we collected, we found that 96% of Ask.fm users use the default setting of accepting anonymous questions; only 4% of Ask.fm users change the profile setting to prohibit anonymous users from leaving a question on their profiles.

We next examine the percentage of anonymous questions answered by a given user (u). This is the percentage of anonymous questions that user u answered in the total number of (anonymous and non-anonymous) questions that user u answered and is defined as $P(u) = (N_A(u)/N_T(u)) \times 100$ (%), where $N_A(u)$ is the number of anonymous questions user u answered, and $N_T(u)$ is the total number of questions user u answered. For the data we collected, $P(u)$ has the average of 76.49% and standard deviation (SD) of 18.32.

Fig. 2 shows the percentage of users (answerers) where the horizontal axis is $P(u)$ and the vertical axis is the percentage of users who have the value of $P(u)$ on the horizontal axis. Fig. 3 shows the cumulative percentage of users where the horizontal axis is $P(u)$ and the vertical axis is the cumulative percentage of users who have the value of $P(u)$ smaller than the value on the horizontal axis. As shown in Figs. 2 and 3, many users are concentrated around the average and the higher values of $P(u)$ (in the range between 75% and 100% on the horizontal axis). These figures also show that most users (91.21% of users) have $P(u)$ larger than 50% and that a small fraction of users (only 8.79% of users) have $P(u)$ smaller than 50%. This means that most users (91.21% of users) answer more anonymous questions than non-anonymous questions.

Fig. 4 shows how $P(u)$ changes as the number of questions answered increases, or as users continue to use Ask.fm. This figure shows $P(u)$ (on the vertical axis) that is computed from the first (oldest) 50 (anonymous and non-anonymous) questions, second (next oldest) 50 (anonymous and non-anonymous) questions, and so on, as indicated by the horizontal axis. As shown in the figure, $P(u)$ increases from 64.10% (Standard Deviation (SD)=23.66%) (for the first 50 questions) to 73.04% (SD=18.38%) (the second 50 questions) and then stays relatively constant at around 79.55% (SD around 11.96%) (from 51st – 100th questions to 451st – 500th questions).

Fig. 4 shows that Ask.fm users (answerers) answer anonymous questions more frequently (64.10%–79.55%) than non-anonymous questions and that the level of user involvement with anonymous questions slightly increases with time as users use Ask.fm. Namely, Fig. 4 shows that answering anonymous questions is a major activity in Ask.fm (as evidenced by high percentage, 64.10%–79.55%, of answering anonymous questions) and are independent of time (as evidenced by relatively constant level of $P(u)$).

Table I
NEGATIVE WORDS IN ANONYMOUS AND NON-ANONYMOUS QUESTIONS

	Anonymous questions	Non-anonymous questions
# of questions	1,570,022	525,178
Pr [N(Q) = 0]	0.9419	0.9458
Pr [N(Q) $\geq$ 1]	0.0590	0.0542
Pr [N(Q) $\geq$ 2]	0.0085	0.0085

Table II
NEGATIVE WORDS IN ANSWERS TO ANONYMOUS QUESTIONS AND NON-ANONYMOUS QUESTIONS

	Anonymous questions	Non-anonymous questions
Pr [N(A) = 0 | N(Q) = 0]	0.942	0.965
Pr [N(A) $\geq$ 1 | N(Q) = 0]	0.058	0.035
Pr [N(A) = 0 | N(Q) $\geq$ 1]	0.799	0.847
Pr [N(A) $\geq$ 1 | N(Q) $\geq$ 1]	0.201	0.153

B. Use of Negative Words in Questions and Answers

In order to further understand the nature of user behavior in Ask.fm, we next examine the degree of aggressiveness of anonymous users through counting the number of negative words used in anonymous questions. We examined both anonymous questions and non-anonymous questions and obtained the probability that a question in each set of questions contains no negative words (Pr [N(Q)=0]), at least one negative word (Pr [N(Q)$\geq$1]), and more than two negative words (Pr [N(Q)$\geq$2]), where N(Q) represents the number of negative words in the question.

Table I summarizes the results. By comparing Pr [N(Q)=0] and Pr [N(Q)$\geq$1] for anonymous questions and those for non-anonymous questions in the table, we observe that there is probability difference of 0.0039 ($p < 0.001$) in how often negative words appear between anonymous and non-anonymous questions. This result shows that anonymous users use negative words more frequently than anonymous users.

To further investigate a possible role of anonymity in inducing aggressive user behavior, we examine how Ask.fm users (answerers) respond to anonymous questions that contain negative words. Table 2 shows whether Ask.fm users use negative words in answering anonymous questions and non-anonymous questions given that these questions contain zero or given that these questions contain at least one negative word. In this table, Pr [N(A)=0|N(Q)=0] is the probability that an answer contains no negative word for a question containing no negative word. Note that in this table, Pr [N(A)$\geq$1|N(Q)=0] = 1 - Pr [N(A)=0|N(Q)=0]. Pr [N(A)=0|N(Q)$\geq$1] is the probability that an answer contains no negative word for a question containing at least one negative word (and Pr [N(A)$\geq$1|N(Q)$\geq$1] = 1 - Pr [N(A)=0|N(Q)$\geq$1]).

As shown in Table II, Pr [N(A)$\geq$1|N(Q)=0] for anonymous questions is larger than that for non-anonymous questions ($p < 0.001$), indicating that Ask.fm users (answerers) exhibit more aggressive behavior in answering anonymous questions than in answering non-anonymous questions, when questions are not aggressive and contain no negative word. Table II also shows that Pr [N(A)$\geq$1|N(Q)$\geq$1] for anonymous questions is larger than that for non-anonymous questions ($p < 0.001$), indicating that Ask.fm users (answerers) exhibit more aggressive behavior in answering anonymous questions than in answering non-anonymous questions, when questions are aggressive and contain at least one negative word.

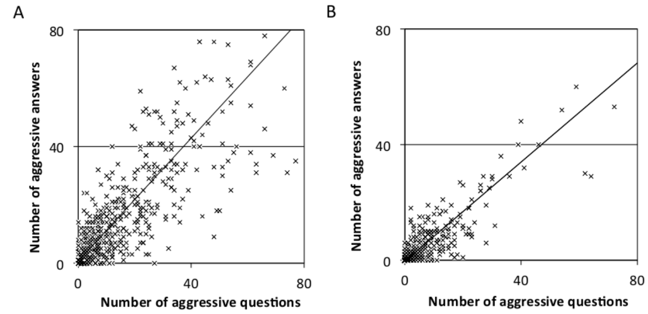

Figure 5. Number of aggressive answers (A) when questions are aggressive and anonymous or (B) when questions are aggressive and non-anonymous

In order to demonstrate the tendency that Ask.fm users (answerers) exhibit more aggressive behavior in answering aggressive anonymous questions than in answering aggressive non-anonymous questions, Fig. 5 plots the relationship between the number of aggressive questions that a given user received and the number of aggressive answers that the given user provided for each user. Fig. 5A is for anonymous questions, and Fig. 5B is for non-anonymous questions. In Figs. 5A and B, each point represents a particular user with a particular number of aggressive questions (shown on the horizontal axis) and a particular number of aggressive answers (shown on the vertical axis). Both figures show that the number of aggressive questions and that of aggressive answers are correlated, each with a high correlation coefficient of 0.87 (in Fig. 5A) and of 0.89 (in Fig. 5B). We further conducted linear fitting of these points in each figure and show the result as a solid line in each figure. Fig. 5A has a slope larger than 1, while Fig. 5B has a slope smaller than 1, indicating that Ask.fm users exhibit more aggressive behavior in answering anonymous questions than in answering non-anonymous questions, when questions are aggressive and contain at least one negative word.

IV. CONCLUSION

This paper examined how anonymity impacts user behavior in Ask.fm. The first finding is that, in asking a question, anonymous users exhibit the higher degree of aggressiveness than non-anonymous users. The second finding is that Ask.fm users become more aggressive in responding to aggressive anonymous questions than to aggressive non-anonymous questions.

In the data analysis, we used the number of negative words as an indication of aggressive behavior. Prior research on Ask.fm, however, points out that aggressive user behavior is not always associated with cyber-bullying and that non-anonymous users often use negative words to defend cyber-bullying victims and also in daily conversation with close acquaintances [11]. Future research is necessary to consider the context where negative words are used in order to determine the degree of aggressiveness of such words, to classify types of aggressive user behavior and identify negative words associated with each type of aggressive user behavior, and to accurately assess how likely different types of user behavior and negative words may lead to cyber bullying incidents.

Acknowledgment

The work presented in this paper was supported by the Kurata Memorial Hitachi Science and Technology Foundation.

References

[1] R. M. Kowalski, S. P. Limber, and P. W. Agatston, *Cyberbullying: Bullying in the Digital Age*. Wiley-Blackwell, 2012.

[2] M. J. Moore, T. Nakano, T. Suda, and A. Enomoto, *Social Network Engineering for Secure Web Data and Services*. IGI GLOBAL, 2013, ch. Social Interactions and Automated Detection Tools in Cyberbullying, pp. 67–87.

[3] K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning to detect cyberbullying,” in *10th International Conference on Machine Learning and Applications and Workshops (ICMLA)*, 2011, pp. 241–244.

[4] A. Kontostathis, K. Reynolds, A. Garron, and L. Edwards, “Detecting cyberbullying: query terms and techniques,” in *5th Annual ACM Web Science Conference*, 2013, pp. 195–204.

[5] H. Lieberman, K. Dinakar, and B. Jones, “Let’s gang up on cyberbullying,” *Computer*, vol. 9, no. 44, pp. 93–96, 2011.

[6] K. Dinakar, B. Jones, C. Havasi, H. Liberman, and R. Picard, “Common sense reasoning for detection, prevention, and mitigation of cyberbullying,” in *ACM Transactions on Interactive Intelligent Systems*, vol. 2, no. 3, 2012.

[7] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, Q. L. R. Han, and S. Mishra. Detection of cyberbullying incidents on the instagram social network. arXiv:1503.03909.

[8] V. Nahar and C. P. X. Li, “An effective approach for cyberbullying detection,” *Communications in Information Science and Management Engineering*, vol. 3, no. 5, pp. 238–247, 2013.

[9] M. J. Moore, T. Nakano, A. Enomoto, and T. Suda, “Anonymity and roles associated with aggressive posts in an online forum,” *Computers in Human Behavior*, vol. 28, no. 3, pp. 861–867, 2012.

[10] E. Omernick and S. O. Sood, “The impact of anonymity in online communities,” in *2013 International Conference on Social Computing*, 2013, pp. 526–536.

[11] H. Hosseinmardi, R. I. Rafiq, S. Li, Z. Yang, R. Han, S. Mishra, and Q. Lv, “A comparison of common users across instagram and ask.fm to better understand cyberbullying,” arXiv:1408.4882.

[12] S. Nakatani, “Language detection library for java,” http://code.google.com/p/language-detection/, 2010.

[13] Swear word list & curse filter. [Online]. Available: http://www.noswearing.com/dictionary