Towards Understanding Cyberbullying Behavior in a Semi-Anonymous Social Network

Homa Hosseinmardi, Richard Han,
Qin Lv and Shivakant Mishra
Department of Computer Science
University of Colorado at Boulder
{homa.hosseinmardi,richard.han,qin.lv,shivakaht.mishra}@colorado.edu

Amir Ghasemianlangroodi
Department of Electrical, Computer and
Energy Engineering
University of Colorado at Boulder
amir.ghasemianlangroodi@colorado.edu

***Abstract*—Cyberbullying has emerged as an important and growing social problem, wherein people use online social networks and mobile phones to bully victims with offensive text, images, audio and video on a 24/7 basis. This paper studies negative user behavior in the Ask.fm social network, a popular new site that has led to many cases of cyberbullying, some leading to suicidal behavior. We examine the occurrence of negative words in Ask.fm's question+answer profiles along with the social network of "likes" of questions+answers. We also examine properties of users with "cutting" behavior in this social network.**

I. Introduction

One of the most pressing problems in high schools is bullying. However, with today's technology, bullying is moving beyond the schoolyards via cell phones, social networks, online video and images, etc. As bad as fighting and bullying were before the prevalence of personal technology, the recording and posting of hurtful content has magnified the harmful reach of bullying. On average, 24% of high school students have been the victim of cyberbullying [1]. Cyberbullying happens in many different ways, including: mean, negative and hurtful comments, pictures or videos posted online or on cell phones, or through the spread of rumors or threats via technology.

Although cyberbullying may not cause any physical damage initially, it has potentially devastating psychological effects like depression, low self-esteem, suicide ideation, and even suicide [2], [3]. For example, Phoebe Prince, a 15 year old high school girl, committed suicide after being cyberbullied by negative comments in the Facebook social network [4]. Hannah Smith, a 14 year old, hanged herself after negative comments were posted on her Ask.fm page, a popular social network among teenagers [5]. Cyberbullying is such a serious problem that nine suicides have been linked with cyberbullying on the Ask.fm Web site alone [6]. Cyberbullying was viewed as a contributing factor in the death of these teenagers [1]. Given the gravity of the problem and its rapid spread among middle and high school students, there is an immediate and pressing need for research to understand how cyberbullying occurs today, so that techniques can be rapidly developed to accurately detect, prevent, and mitigate cyberbullying.

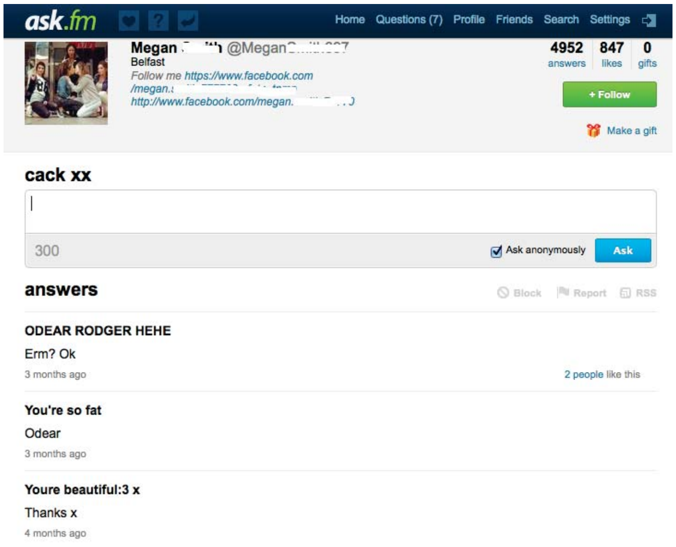
While most current studies have focused on the prevalence and impact of cyberbullying in education and psychology [7]–[9], our interest is in understanding how social networks are being used to enable cyberbullying. Prior work in cyberbullying analysis and detection in social networks has largely focused on such social networks as Youtube, Formspring, MySpace, and Twitter [10]–[12]. [10] investigated both explicit and implicit cyberbullying by analyzing negative text comments on Youtube and Formspring profiles. [11] investigates how integration of MySpace user profile information like gender in addition to text analysis can improve the accuracy of cyberbullying detection in social networks [12] tries to detect bullying in Twitter text data by looking for inappropriate words using a Naive Bayes classifier. They track potential bullies, their followers and the victims. All of these works focused on text-based analysis of negative words, and did not exploit social network relationships in their investigation of cyberbullying.

Previous work [13] has considered characterizing the language model of MySpace users, though the emphasis was not on negative word usage. Our work seeks to understand whether social network relationship information can be useful in supplementing purely text-based analysis in helping to identify negative user behavior in social networks. We are interested in building graphs from social networks and analyzing their properties to extract features such as in-degree or out-degree that may be useful in flagging such bad behavior.

In particular, this paper chooses to focus on analyzing the Ask.fm social network for the following key reasons. First, Ask.fm is a major source of cyberbullying on the Internet. In fact, it ranks as the fourth worst site in terms of percentage of young users bullied according to a recent survey [14], after Facebook, YouTube and Twitter. Second, very little research has been published to date concerning the Ask.fm social network. Third, Ask.fm is a highly popular and rapidly growing social network, with over 70 million registered users as of August 2013 [15]. Finally, Ask.fm provides publicly accessible data.

One of our big challenges in analyzing the Ask.fm social network is that it behaves as a semi-anonymous social network. User profiles are public, but postings to each profile by users other than the owner are by default anonymous. In addition, we cannot obtain from the public profiles which users are following which other users. As a result, it is not possible for us to construct a social graph based on friendships. However, we observe that another type of graph called an interaction graph [16] can be extracted from the "likes" of comments. We use this insight to build and analyze interaction graphs that embed social relationship information, to help identify negative user behavior in this semi-anonymous social network.

ASONAM 2014, August 17-20, 2014, Beijing, China
978-1-4799-5877-1/14/$31.00 ©2014 IEEE



Megan [redacted] @Megan[redacted]
Belfast
Follow me https://www.facebook.com
/megan.[redacted]
http://www.facebook.com/megan.[redacted]

4952 answers | 847 likes | 0 gifts

cack xx

☑ Ask anonymously

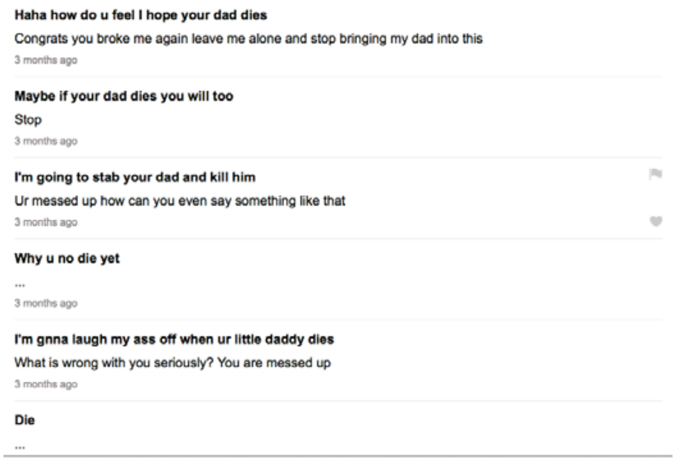
answers

ODEAR RODGER HEHE
Erm? Ok
3 months ago
2 people like this

You're so fat
Odear
3 months ago

Youre beautiful:3 x
Thanks x
4 months ago

Fig. 1. Typical public profile on Ask.fm with questions/comments (usually anonymous) to the profile owner and answers from the profile owner.



Haha how do u feel I hope your dad dies
Congrats you broke me again leave me alone and stop bringing my dad into this
3 months ago

Maybe if your dad dies you will too
Stop
3 months ago

I'm going to stab your dad and kill him
Ur messed up how can you even say something like that
3 months ago

Why u no die yet
...
3 months ago

I'm gnna laugh my ass off when ur little daddy dies
What is wrong with you seriously? You are messed up
3 months ago

Die
...

Fig. 2. An example of anonymous cyberbullying comments posted on a user's profile in Ask.fm.

This paper makes the following contributions. It is the first paper to to provide a detailed characterization of key properties of the important Ask.fm semi-anonymous social network. Second, it builds and analyzes interaction and word graphs and finds that properties of the interaction graph such as in-degree and out-degree are strongly related to the amount of negative user behavior expressed on a profile, i.e. highly positive profiles exhibit the highest degree of sociability in terms of liking others and being liked by others, whereas profiles with a high number of negative questions exhibit the lowest degree of sociability.

In the following, we describe our data collection efforts, build graphs of users' "likes" as well as negative word graphs, and use these to illustrate the relationship between negative words and user activity. We also analyze a particularly high risk set of users who state on their profile that they have "cut" themselves.

II. Data Collection

In this section, we discuss key aspects of the Ask.fm social network, what information was collected from Ask.fm's publicly accessible profiles, e.g. questions, answers, and likes, and our process of building interaction and word graphs.

A. Ask.fm

We observe that Ask.fm is generally an example of a semi-anonymous social network, in that the identity of users who post questions/comments to a profile is typically anonymous (though posters may choose to reveal their identity, we have seen this happen only rarely), whereas the identity of the target user is publicly known. People can search Ask.fm users via their name, id or email address. This is unlike non-anonymous social networks such as Facebook and Twitter, where the identity or ID of posters and target users are both publicly visible. Thus, in Ask.fm, any one (even people without an account) may post on another user's profile, and this posting of a question or comment is usually done anonymously (in fact, that is the default).

There are some policies specific to Ask.fm. First, only the target user may post an answer to a question/comment on this site. Only after answering a question will the question and answer appear on his/her profile. Further, a user may choose to "like" at the granularity of a question+answer pair, but cannot like the question nor the answer individually. Liking is non-anonymous, so that the identity of the likers is publicly known. Another feature of Ask.fm is that users may follow other users. However, this relationship data is not available publicly and only the profile owner knows who he/she is following. Even the user who is being followed can only know how many followers he/she has (which is not publicly available to other users), not who is following him/her.

Figure 1 shows a typical publicly accessible profile obtained from the Ask.fm social Web site. We see that other users may post questions/comments on a target user's profile, and that the target user may answer each question/comment. In this example, we see both a negative comment "You're so fat" mixed in with a positive comment "You're beautiful". A more serious example of cyberbullying from the Ask.fm Web site is shown in Figure 2, where we observe that anonymous negative comments have been repeatedly posted on the target's profile wishing or threatening death upon the target's father.

B. Description of Collected Data

The data we can extract from a common profile includes the following fields: userID, personal information (if any, as it is optional), total number of answers, total number of likes, content of answered questions posted on a user's page, and the userID of people who liked the questions+answers.

An interaction graph can be constructed from the "likes" of answered questions, i.e. each directed edge in the graph connects user i to neighbor j if user i has liked a question+answer pair in j's profile. Note that the edges are not bidirectional, so that i liking an question+answer on j's profile does not imply j liking one of i's question+answer. In order to extract this interaction graph from Ask.fm, we conducted a breadth-first search starting from a couple of random seed nodes. Seed nodes should have non-zero liked question+answer pairs. For each seed node we found all nodes that liked an answered question on its page (incoming edges are publicly known for each profile). However by only looking at a user's profile, we are not privy to any of the nodes that this profile owner liked (outgoing edges). In the second step, we collect the profile information of all neighbors of the seed node. These steps are then repeated. Ultimately, the outgoing edges from a profile can be reconstructed from the incoming edges of other profiles. Note that because we're crawling profiles using breadth-first search, we can only find that subset of the outgoing edges for each profile that happen to be incoming edges on other crawled profiles. The only way to find all outgoing edges for each profile is to crawl the entire Web site, which is impractical. Since the breadth-first search is terminated before crawling all profiles, then this is called snowball sampling, and results in an interaction graph wherein all the internal nodes have been fully crawled or sampled, but also a typically small fraction of nodes on the edge that have not yet been crawled. Our analysis below focuses only on fully sampled nodes in the interaction graph. Using snowball sampling, 30K profiles were crawled from October to December 2013.

To provide some general context, we first perform an analysis of the total number of question+answer pairs and likes per Ask.fm user. We observe they both have heavy tail distributions in Figure 3. The total number of likes has a heavier tail than the total number of question+answer pairs. Looking at the tail, we observe that as the total number of answered questions in a user profile increases, the total number of likes also grows, Figure 4. The red line is the best fit obtained through linear regression in the transformed (log-log) space. It seems when a user is more popular and active, they receive more questions, and more user visits and likes on his/her page. Also pushing a like button is easier than writing a question, which makes the number of likes larger than the number of questions.

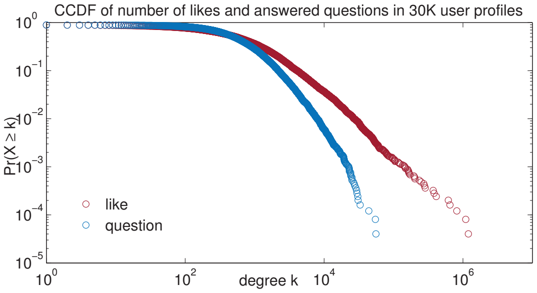


Fig. 3. Probability that the total number of likes and question+answer pairs for a user is greater than or equal to k

Figure 4 describes the correlation between the number of answered question+answer pairs and the number of likes. The correlation value when the number of answered questions is less than 50 is -0.05. This shows that when the number of answered questions in a user page is low, the number of people who like these users does not have any special relation with the number of answered questions. When the number of answered questions is larger than 50, the correlation between the number of likes and the number of answered questions increases to 0.33. This means that as the number of question+answer pairs increases beyond around 50, the number of likes becomes weakly related to the number of answered questions.

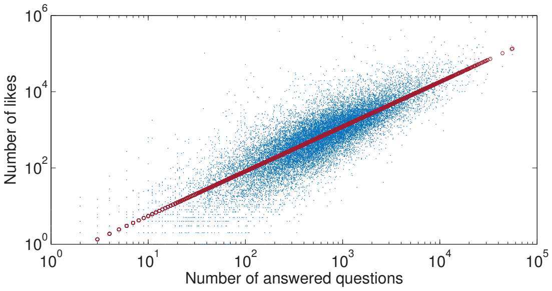


Fig. 4. Total number of likes versus total number of answered questions a user receives

C. Modeled Interaction Network

The preceding analysis provided a general overview of likes and answered questions in Ask.fm, but we are more interested in likes and answered questions in the context of cyberbullying. Our search for cyberbullying on the Ask.fm is based on the insight that repetitive negative words represent the core of the abusive text posted on profiles. Following the occurrence of negative words led us to many examples of cyberbullying. However, after a preliminary analysis of the answers by the profile owner, we found that many such examples seemed to have no effect on the targeted user, namely based on the answers the target seemed indifferent to the negative comments. In contrast, there were other more serious cases where the target seemed particularly vulnerable, making statements like "you broke me" (See Figure 2).

Looking at the profiles, we observed that the users who seemed most vulnerable to negative questions were often those who were most isolated, with few "likes" and also rarely liking others' comments. In contrast, users who were subject to but seemingly indifferent to negative questions appear to have a fair number of likes and also seemed somewhat active in liking others' comments.

Based on these two observations that (i) cyberbullying is the behavior of posting questions with negative words and (ii) vulnerable targets of cyberbullying (based on their answers) seem isolated, we sought to build and analyze social graphs and word graphs derived from our data that would capture the negativity and isolation of users.

In order to capture the greatest degree of interaction between the users, we collected the top 15 questions for each user that had the highest number of likes. For some users with very low activity, the number of highest liked questions was less than 15. Analyzing the most popular questions of profiles, we have built our network modeled as a directed bipartite network. In order to build our graph, if user i likes a question+answer in the page of user j, then there is a link from i to words on that question and a link from those words to node j.

From this bipartite graph, we seek to derive characteristics that can highlight the negativity and isolation associated with targeted users. In order to project a bipartite network with adjacency matrix B, to the network of words W, we have $W = BB^T$. Then we can similarly build the network of users with adjacency matrix U from our bipartite network. That is, based on this data, our idea is to build and use a like-based interaction graph between users (U) and examine the balance of in-degree vs out-degree of the users with high degree of negativity in their pages and small number of positive posts; such users have both low in-degree and low out-degree in graph terminology. In contrast, users who received on average the same amount of negativity but have positive questions at the same time, show healthier in-degree and out-degree.

D. Extracting Positive and Negative Words

We next consider which negative words are pertinent to our analysis of negative user behavior. The natural approach is to select those negative words that have the highest frequency in the sampled profiles. However, we observe that it is the collective effect of negative words that is exploited by cyberbulliers. Negative words that may be commonly used but more in isolation rather than in a collective fashion would be less likely to create a strong effect of cyberbullying. Therefore, we sought to create a word graph that measures the relationship between words, i.e. are they being used together on the same profile to bully a victim. By one mode projection from a bipartite graph comprised of 30K users and a dictionary of around 1500 negative words (obtained from [17]), we constructed a word graph wherein each word signifies a node in the graph and each edge indicates that the two words have been used in the same profile.

We found that there is a cluster of connected negative words in the center of the word graph, but there are also many negative words that were not connected to any word. Out of 1500 words, the eigenvector centrality of 968 words is zero, which means 968 negative words either have not been seen in any profile or have not been seen together in a profile. These words were eliminated from our analysis. The top row of table I shows the remaining negative words that had the highest frequency of appearance. The second row of Table I shows the negative words with the highest eigenvector centrality. Though there is a fair degree of overlap with the highest frequency words, we note that the sets do indeed differ. Since eigenvector centrality captures to some extent the collective negativity of cyberbullying, we focused our word graph analysis below on the 80 negative words with eigenvector centrality values larger than 0.5. A similar approach was followed for a collection of 1000 positive words. 80 positive words with highest eigenvector centrality were chosen for the following analysis.

III. Network Statistics

A. Building the Interaction Graph

We collected and analyzed about 30K profiles that had a complete list of likes, using the snowball sampling method, gathering the top 15 most liked questions. We found that there were on average 14.5 most liked questions per user, that is most users had close to 15 questions with likes. We also focused on a collection of 80 negative words and 80 positive words with the highest eigenvector centrality as explained previously.

Average number of answers per user	14.5
Average number of negative questions per user	0.778
Average number of positive questions per user	4.57
Average number of negative words per user	1.04
Average number of positive words per user	5.42

TABLE II. Average number of questions and words per user

We found that 46.0% have at least one question including one of the top negative words. We term such a question a "negative" question. Furthermore, 7.74% of the users have at least three questions with top negative words, and 96.0% users had at least one question with a top positive word in their profiles. We term such a question a "positive" question. Table II shows the average number of positive and negative questions and words per user.

By a one mode projection of the bipartite graph, we were able to obtain the like-based interaction graph U between these 30K users in the Ask.fm social network. In the adjacency matrix U, created from the users' interaction graph, edges are weighted. In fact the weight of each edge e_{ij} , from node i to node j is a vector $[n_1 \; n_2]$, where n_1 is the number of questions in node j including at least one negative word (from our selected dictionary) and n_2 is the number of questions with no negative word liked by node i. Since we are most interested in negative behavior, we group positive and neutral behavior together into a "non-negative" category.

For further analysis, we have divided matrix U into two matrices U_{neg} and $U_{non-neg}$, where U_{neg} (negative adjacency matrix) is the adjacency matrix with weights n_1 and $U_{non-neg}$ (non-negative adjacency matrix) is the adjacency matrix with weights n_2. Figure 5 shows the CCDF for the in-degree and out-degree distribution of matrices U_{neg} and $U_{non-neg}$. Authors in [18] show that in-degree and out-degree distributions in social networks are approximately the same. We can see from Figure 5 the in-degree and out-degree for non-negative degree distribution is approximately the same for non-negative degree distributions. *However, the negative in-degree and negative out-degree distributions of the interaction graphs clearly differ, indicating that negative behavior in the Ask.fm semi-anonymous social network has different properties than previously reported in other social networks.*

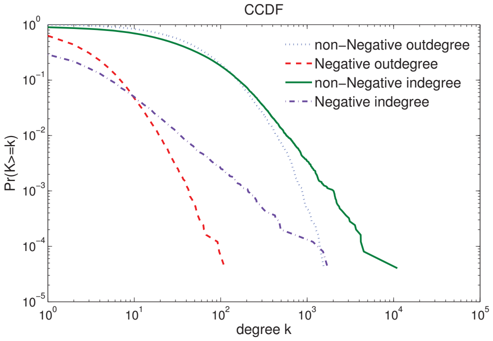


Fig. 5. Complementary cumulative distribution function for negative and non-negative interaction graphs/matrices.

Figure 6 shows the percentage of common users among x% highest in-degree and x% highest out-degree. In fact this figure shows whether the users with high out-degree are the same people with high in-degree. Authors in [18] show in

hoe	n**ga	stupid	slut	kill	p**sy	cut	suck	gay	fat	sex	bad	d**k	die	ugly	s**t
f**k	ass	bitch	s**t	hate	d**k	ugly	sex	bad	suck	p**sy	fat	gay	stupid	die	kill

TABLE I. THE FIRST ROW SHOWS THE NEGATIVE WORDS WITH THE HIGHEST FREQUENCY AND THE SECOND ROW SHOWS THE NEGATIVE WORDS WITH THE HIGHEST EIGENVECTOR CENTRALITY (VALUES DECREASE FROM LEFT TO RIGHT).

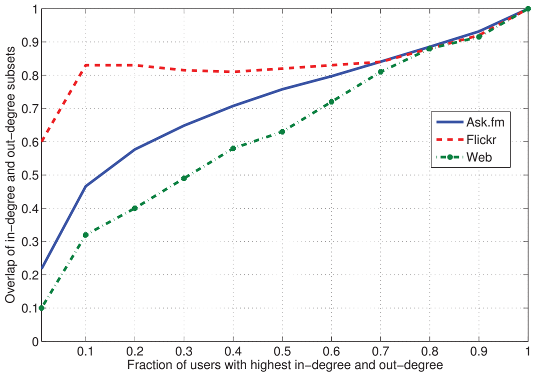


Fig. 6. Percentage of common users among two subsets of x% highest in-degree and x% highest out-degree. Measurements for Flicker and Web have been obtained from [18]

social networks for users with 1% highest in-degree and out-degree, more than 60% of users are common. This value is less than 20% for the Web. Here we observe that Ask.fm's semi-anonymous social network exhibits behavior that is between previously analyzed social networks and the Web. Overlap of the users with 1% highest in-degree and out-degree is more than 20% and less than 30%. It seems the correlation between in-degree and out-degree in an interaction graph built from like-based interactions is much less than the correlation in friendship-based social graphs. Friendship is usually a symmetric relation. However, in like-based interactions symmetry is less probable, which causes less correlation between in-degree and out-degree. In addition, there is the ambiguity in Ask.fm as to whether a like is for a question or answer. This decreases the correlation between in-degree and out-degree because we do not know whether the liker is liking the question as a support of the question's content or the answer to support the profile owner.

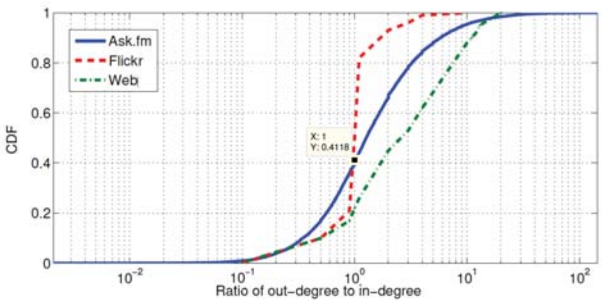


Fig. 7. Cumulative distribution function for ratio of out-degree to in-degree. Measurements for Flicker and Web have been obtained from [18]

Figure 7 shows the CDF for the ratio of out-degree to in-degree. It has been shown in [18] that for social networks more than 50% of users have in-degree within 20% of their out-degree. However in the Ask.fm this value is around 16%.

Again we observe that the relationship between in-degree and out-degree in Ask.fm is weaker than what was found for prior social networks and is stronger than the Web.

The mean reciprocity of the interaction graph/matrix U is 28.2%, which is a low number compared to other social graphs like Yahoo! 360 with reciprocity 84% and Flickr with reciprocity 68%, [19]. However Twitter has a more similar structure to Ask.fm (in the sense that there is no friendship concept between users) and has an even lower reciprocity equal to 22.1% [19]. Ask.fm's network negative reciprocity is a very low 3.61%, which shows how much users like each others' negative questions. This gives an insight that users who both have negative posts do not tend to like each others' negative questions. Reciprocity between non-negative questions is a far higher 27.9%.

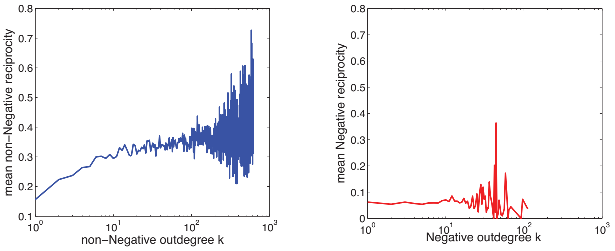


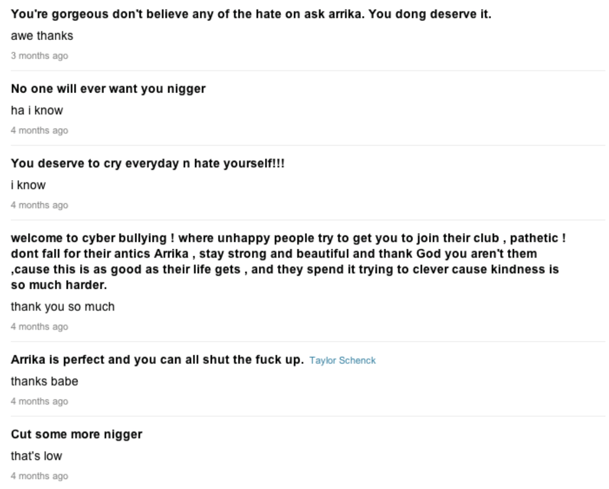
Fig. 8. Reciprocity for non-negative and negative interaction graphs/matrices

Figure 8 shows the mean reciprocity for the two graphs/matrices U_{neg} and $U_{non-neg}$ versus out-degree. It shows that in normal (non-negative) behavior the more active a user is, the more he/she will receive likes from other users or vice versa. On the other hand, the mean reciprocity does not increase with out-degree in the negative matrix. When the negative out-degree is high, this means that a user is active and likes others user's negative questions. However this type of active user receives a low negative in-degree in return. We consider two likely possibilities that explain this result. First, the user is supportive and popular, liking answered questions often to show support for the answerer of a negative question. Due to the lack of liking granularity in Ask.fm, this like is recorded as a like of a negative question, even though the intent of the user was to like the answer. Indeed, we found cases on Ask.fm where a profile owner asked a liker why they liked a negative question, and the liker responded that in fact they were liking the profile owner's response instead. Such a supportive user would be unlikely to receive many negative questions on their profile. Alternatively, the user is a bully and unpopular, frequently liking other's negative questions, and not receiving many likes of negative questions in return. Again, we see that negative behavior follows a different pattern than non-negative behavior.

Next, we explore the relationship between the degree of negativity (and positivity) on a profile and the profile's graph properties. We compute the average number of negative questions in a set of profiles with negative questions that the profile owner answers to show his/her unhappiness about the questions. For example, as we saw in Figure 2, the profile owner says "you broke me again" in response to repeated negative questions. The average number of negative questions in 150 profiles of this type was 2.87 and therefore we chose a threshold of 3 negative questions to define negative groups. In fact, we segmented the user base into 4 different groups:

1) Highly Negative: users with at least 3 negative posts and no positive posts (HN)
2) Highly Positive: users with more than 10 positive posts (HP)
3) Positive-Negative: users with at least 3 negative posts and more than 4 positive posts (PN)
4) Others (OTR)

The definition of these groups helps us identify the properties of the most negative profiles, namely the HN users who have no positive support, while also allowing us to contrast them with the graph properties of other users who have some or a lot of positive words in support. They were chosen based on our observations that targets of negative behavior could be roughly divided into a group that receives support from bystanders and a group that has been left alone and doesn't receive any support with positive posts. Figure 9 shows an example of a mixed profile with both positive and negative comments, representative of the PN group.



You're gorgeous don't believe any of the hate on ask arrika. You dong deserve it.
awe thanks
3 months ago

No one will ever want you nigger
ha i know
4 months ago

You deserve to cry everyday n hate yourself!!!
i know
4 months ago

welcome to cyber bullying ! where unhappy people try to get you to join their club , pathetic ! dont fall for their antics Arrika , stay strong and beautiful and thank God you aren't them ,cause this is as good as their life gets , and they spend it trying to clever cause kindness is so much harder.
thank you so much
4 months ago

Arrika is perfect and you can all shut the fuck up. Taylor Schenck
thanks babe
4 months ago

Cut some more nigger
that's low
4 months ago

Fig. 9. An example of a profile with both positive and negative comments.

	HN	HP	PN	OTHR
Negative reciprocity	0.131	0	0.106	0.066
Non-Negative reciprocity	0.210	0.379	0.246	0.339
Negative in-degree	6.39	0	14.0	2.16
non-Negative in-degree	21.1	112	49.4	70.1
Negative out-degree	3.48	2.73	3.73	3.59
non-Negative out-degree	32.8	117	46.0	70.1
Total number of likes	1027	5995	1765	2906
Ratio of likes per answer	1.59	2.58	1.94	2.73

TABLE III. AVERAGE RECIPROCITY AND DEGREE FOR DIFFERENT GROUPS

Table III summarizes some of the key results of our graph analysis. We measure the average reciprocity of the four different groups, as well as the in-degree and out-degree, based on the interaction graphs U_{neg} and $U_{non-neg}$ calculated earlier. Note the in-degree has two subcategories, pertaining to negative in-degree and non-negative in-degree, i.e. a user X will have a negative in-degree if another user likes one of the questions posted on X's profile that has a negative word in it, while another user liking a question without a negative word will count towards the non-negative in-degree of X. Similarly, out-degree has negative and non-negative subcategories.

From Table III, a key finding is that HN users are distinguished by having the smallest total (negative plus non-negative) in-degree and the smallest total out-degree. It shows they are either not popular in terms of being liked or do not tend to have activity in this social network in terms of liking others. That is, the high degree of negativity that these users are subject to is related to less sociable behavior on this social network. Our results indicate that we could leverage low total in-degree and low total out-degree of Ask.fm's interaction graph to suggest a greater likelihood of determining highly negative profiles of cyberbullying victims on this semi-anonymous social network.

Another major finding is that as the amount of positive support increases, we find a greater in-degree and greater out-degree of the users, that is users become more social and actively like and are liked more often. This is demonstrated first by the PN group, which like the HN group has at least 3 negative posts, but the PN group also has positive support. We observe that the PN group has increased activity in terms of about twice the amount of total in-degree and total out-degree as the HN group. Highly positive HP profiles exhibit the highest sociability in terms of actively liking other profiles and being liked by other profiles. We see that the total in-degree and the total out-degree are both over 100, which is much higher than the second nearest group OTHR, which has a total in-degree and out-degree around 70 each. This confirms the trend that higher positivity is strongly related to higher sociability on Ask.fm's interaction graph.

An interesting result of this analysis is that we can distinguish the HN group from the PN group not merely by the total in-degree and total out-degree, but also by just the negative in-degree. We see that the negative in-degree of HN is 6.39, while PN's negative in-degree is 14.0, over twice as large. While the mean number of negative questions is the same in both HN and PN, i.e. about 3, why would there be so many more likes for the same number of negative questions for the PN group? The explanation for this lies again we believe with the limited granularity of liking in Ask.fm. In the case of liking a question+answer pair, where the question/comment is negative and the answer is positive, then it would not be clear whether a user is liking a bullier's negative comment or supporting a positive answer from the target of bullying. The higher volume of likes of negative questions we believe is actually users liking the target's positive response. This agrees with an examination of a variety of examples, where we saw that the likes were mostly representative of support and a positive sign in this case.

We also observe that HP users have little interaction with negative posts. That is, their negative out-degree is clearly lower than any of the other three groups. That means that not only do HP users not have negative questions posted on their profiles, but *HP users spend very little effort liking negative questions on other users' profiles, focusing the vast majority of* *their effort on liking positive questions.* Looking at negative reciprocity, it is clear that HP has negative reciprocity 0 as it does not have any negative in-degree. In the group OTHR it seemed that there existed either a set of users with 1 or 2 negative questions without any limitation on the positive posts, or a set of users with more than 2 negative questions and less than 10 positive questions. However, the average number of negative questions is 0.774, which shows we have mostly users belonging to the first set. We see that the reciprocity on negative questions in this group is lower than the HN and PN groups. The reason is either because they have few negative questions compared to the HN and HP groups (which in average have less), or they have less support (they have in average 4.76 positive questions compared to 6.53 positive questions for the HP group).

In order to calculate the local clustering coefficient for each node we first turned our network into a simple graph. It means either node i has liked node j's question, or node j has liked node i's question, regardless of the number of likes. We set $U_{ij} = 1$ and $U_{ji} = 1$, otherwise $U_{ij} = U_{ji} = 0$. We make this assumption that either user i receives a like from user j or posts a like on his/her profile, they are in the general category of having some familiarity or "connectedness". The expected clustering coefficient of the network is 0.11 and the averaged local clustering coefficient 0.356, as defined in [20]. Comparing with numbers reported by [21], shown in Table IV, the clustering coefficient of Ask.fm is pretty small. In Figure 10 the local clustering as a function of degree has been depicted. As we expect, the local clustering coefficient decreases when the degree increases in social networks. Looking at the clustering coefficient of each group in Table V, we observe that among 4 defined groups, HN has the highest mean local clustering coefficient despite having the lowest degree, while the group HP with highest degree has the lowest mean local clustering. The implication is that users of the HN group probably have a few people that know each other. However, users of the HP group are very social and know many people, and therefore the proportion of their friends that know each other is small. In general, we can see each group that has higher degree has lower mean local clustering.

HN	HP	PN	OTHR
0.499	0.237	0.380	0.357

TABLE V. LOCAL CLUSTERING COEFFICIENT FOR DEFINED GROUP

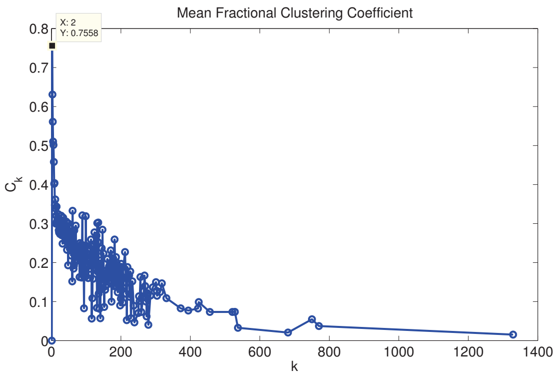


Fig. 10. Mean local clustering coefficient versus degree

IV. USERS WITH CUTTING BEHAVIOR

One of the most disturbing behaviors that we encountered was the problem of "cutting" (slicing one's wrists). Looking at the profiles of these cases and studying their answers, it seemed that the profile owner exhibited weak confidence and depression problems, sometimes admitting to earlier failed attempts at suicide. In this section we first found 150 profiles for which their owners have explicitly expressed the experience of "cutting" behavior and label them. A human labeled the profile as "cutting" behavior if and only if the profile owner has expressed explicitly in his/her answers that he/she has had such an experience. We observe in Figure 11 that among the words that the word "cut" has been connected to are the words "depress", "stressful", "sad", and "suicide" . This association could be used to detect these type of users.

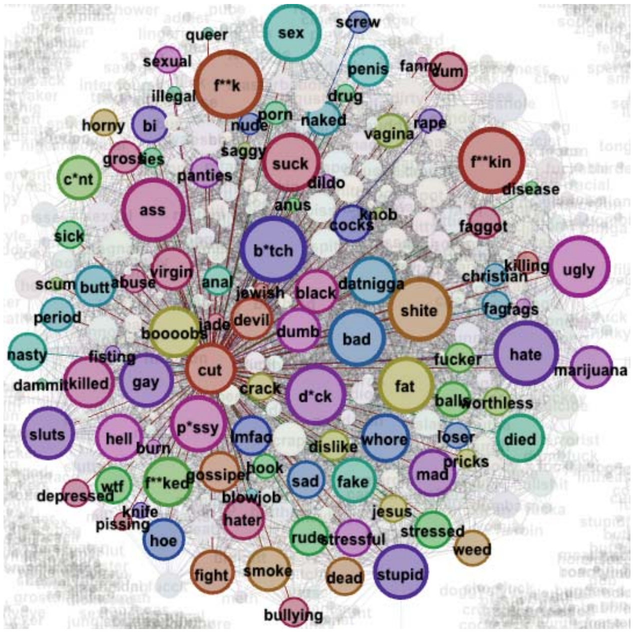


Fig. 11. Word usage with the word "cut" in Ask.fm. The size of the circle indicates Eigenvector Centrality score

The frequency of negative words used with "cut" has been shown in Figure 12. We can see these user profiles have two peaks at words "ugly" and "f**k".

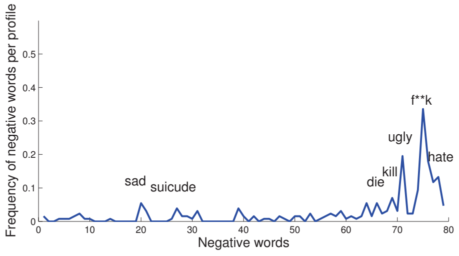


Fig. 12. Frequency of negative words seen with word "cut".

Table VI illustrates the statistics of the labeled users for "cutting" behavior. We found that such profiles have positive questions close to PN profiles however the average number of negative posts is less by a factor 0.65 compared to the PN and HN groups. This suggests surprisingly but encouragingly that

social media	Ask.fm	Facebook	Twitter	Gplus	Flickr	Orkut	You Tube
Clustering Coefficient	0.356	0.606	0.565	0.490	0.313	0.171	0.136

TABLE IV. LOCAL CLUSTERING COEFFICIENT FOR SOME SOCIAL NETWORKS. CLUSTERING COEFFICIENT VALUES FOR FLICKR, ORKUT AND YOU TUBE ARE FROM PAPER [18] AND VALUES FOR FACEBOOK, TWITTER, GPLUS ARE FROM THE WEBSITE [21].

Average number of questions per user	15
Average number of negative questions per user	2.31
Average number of positive questions per user	5.29
Average number of negative words per user	3.38
Average number of positive words per user	6.65

TABLE VI. AVERAGE NUMBER OF QUESTIONS AND WORDS PER USER WITH "CUTTING" LABEL

these profiles do receive more support and less negative posts compared to general HN users.

Table VII shows the statistics of the average reciprocity, in-degree and out-degree for cutting victims. *Though we originally expected this group to exhibit behavior similar to the HN group, we found instead that this cutting group appears to exhibit collectively a different behavior than the HN, HP, PN, and OTHR groups measured in Table III.* The in-degree is 1.5 times more than PN's in-degree and the out-degree is also more by a factor of 1.4. In fact total in-degree and out-degree is more similar to the group OTHR, though there is a marked difference in negative in-degree compared to OTHR. This group has an average number of negative and positive questions most closely related to PN profiles. However, it receives more likes (higher in-degree), and exhibits more activity (higher out-degree) compared to PN groups. This deserves further investigation to explain the reasons behind the differences.

	users with "cutting" label
Negative reciprocity	0.146
Non-Negative reciprocity	0.248
Negative in-degree	17.9
non-Negative in-degree	78.7
Negative out-degree	4.42
non-Negative out-degree	65.0

TABLE VII. AVERAGE RECIPROCITY AND DEGREE FOR USERS WITH "CUTTING" LABEL

V. CONCLUSIONS AND FUTURE WORK

As far as we are aware, this paper is the first to present a detailed analysis of user behavior in the Ask.fm social network. We analyzed nearly 30K profiles of Ask.fm users using interaction graphs, word graphs, and frequency distributions, and characterized key properties such as reciprocity, clustering coefficient, and the influence of negativity on in-degree and out-degree. Some of the key findings of the work are that (1) When people have highly negative profiles without any positive support, they also have the lowest activities in terms of both in-degree and out-degree, that is they are the least sociable. (2) As the amount of positive support increases, we find a greater in-degree and greater out-degree of the users, that is users become more social and actively like and are liked more often. For example, users with negative profiles that also receive positive support have higher in-degree and out-degree than owners of highly negative profiles that lack positive support, that is they are more sociable. When people have highly positive profiles, then they also have the highest in-degree and out-degree, which shows that they socialize the most on this social network. This suggests that we may be able to use the interaction graph's in-degree and out-degree as an indicator of the extent of negativity or positivity on a given profile. (3) The negative in-degree and negative out-degree do not exhibit similar behavior, unlike the similarity of in-degree and out-degree found in other social networks. (4) Finally, our analysis of cutting behavior on Ask.fm reveals that such profiles have surprisingly high positive support, and exhibit a different signature than the other group segments studied.

Our initial analysis of cyberbullying on Ask.fm has uncovered a plethora of future research directions. We hope to design classifiers to detect cyberbullying and evaluate their accuracy and false positives/negatives over the general user population on Ask.fm. Another way to improve detecting victims is looking at the answers in addition to the question. Here by not inspecting the content of the answers, we are potentially missing useful information. Answers give us further insight on when a behavior received from other users starts to disturb the profile owner. Also, we have only looked at the top 15 liked questions. We can extend this to all questions, and also investigate the role of the most recently posted questions on a user's page. We further intend to conduct a more extensive sampling of Ask.fm, obtaining a larger set of profile data from this social network. We would like to investigate in more detail the characteristics of high risk cutting victims to ease their identification. In order to determine the effect of anonymity on the degree of negativity in user behaviors, we intend to compare the semi-anonymous social network Ask.fm with non-anonymous social networks.

ACKNOWLEDGMENT

We wish to acknowledge financial support for this research from the US National Science Foundation (NSF) through grant CNS 1162614.

REFERENCES

[1] J. W. Patchin, "Cyberbullying Research Center:Research summary (2004-2013)," 2013, [accessed 14-January-2014]. [Online]. Available: http://cyberbullying.us/data-posts/

[2] S. Hinduja and J. W. Patchin, "Cyberbullying research summary, cyberbullying and suicide," 2010.

[3] A. N. E. Menesini, "Cyberbullying definition and measurement. some critical considerations," *Journal of Psychology*, vol. 217, no. 4, pp. 320–323, 2009.

[4] R. Goldman, "Teens indicted after allegedly taunting girl who hanged herself, bbc news," 2010, [accessed 14-January-2014]. [Online]. Available: http://abcnews.go.com/Technology/TheLaw/teens-charged-bullying-mass-girl-kill/story?id=10231357

[5] L. Smith-Spark, "Hanna smith suicide fuels calls for action on ask.fm cyberbullying, cnn," 2013, [accessed 14-January-2014]. [Online]. Available: http://www.cnn.com/2013/08/07/world/europe/uk-social-media-bullying/

[6] R. Broderick, "9 teenage suicides in the last year were linked to cyber-bullying on social network ask.fm," 2013, [accessed 14-January-2014]. [Online]. Available: http://www.buzzfeed.com/ryanhatesthis/a-ninth-teenager-since-last-september-has- committed-suicide

[7] R. K. P. W. Agatston and S. Limber, "Students' perspectives on cyber bullying," Journal of Adolescent Health, 2007.

[8] A. Schrock and D. Boyd, “Problematic youth interaction online: Solicitation, harassment, and cyberbullying,” Computer-Mediated Communication in Personal Relationships, 2011.

[9] H. Vandebosch and K. V. Cleemput, “Defining cyberbullying: A qualitative research into the perceptions of youngsters,” CyberPsychology and Behavior, 2008.

[10] C. H. H. L. K. Dinakar, B. Jones and R. Picard, “Common sense reasoning for detection, prevention, and mitigation of cyberbullying,” *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 3, pp. 18:1–18:30, Sep. 2012.

[11] M. Dadvar, F. M. G. de Jong, R. J. F. Ordelman, and R. B. Trieschnigg, “Improved cyberbullying detection using gender information,” in *Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012), Ghent, Belgium*. Ghent: University of Ghent, February 2012, pp. 23–25.

[12] S. K. H. Sanchez, “Twitter bullying detection,” ser. NSDI’12. Berkeley, CA, USA: USENIX Association, 2012, pp. 15–15.

[13] S. W. J. Caverlee, “A large-scale study of myspace: Observations and implications for online social networks,” ser. In Proceedings from the 2nd International Conference on Weblogs and Social Media (AAAI), 2008.

[14] D. the Label-Anti-Bullying Charity, “The annual cyberbullying survey 2013.” [Online]. Available: http://www.ditchthelabel.org/annual-cyber-bullying-survey-cyber-bullying-statistics/

[15] “Ask.fm wikipedia page,” 2014, [accessed 14-January-2014]. [Online]. Available: http://en.wikipedia.org/wiki/Ask.fm

[16] C. Wilson, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, “Beyond social graphs: User interactions in online social networks and their implications,” *ACM Trans. Web*, vol. 6, no. 4, pp. 17:1–17:31, Nov. 2012.

[17] N. words list, “Negative words list form, luis von ahn’s research group,” 2014. [Online]. Available: http://www.cs.cmu.edu/ biglou/resources/

[18] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC ’07. New York, NY, USA: ACM, 2007, pp. 29–42.

[19] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?” in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW ’10. New York, NY, USA: ACM, 2010, pp. 591–600.

[20] M. E. J. Newman, “Networks, an introduction,” published by Oxford.

[21] J. McAuley and J. Leskovec, “Stanford network analysis project,” http://snap.stanford.edu/data/egonets-Facebook.html, 2012, [accessed 14-January-2014].