iAnon: Leveraging Social Network Big Data to Mitigate Behavioral Symptoms of Cyberbullying

Zahra Ashktorab[1], Srijan Kumar[1], Soham De[1], Jennifer Golbeck[1]
[1]University of Maryland, College Park

Abstract
Because of the widespread use of social networks, it is difficult for victims of cyberbullying to seek refuge from this torment of bullying. Some social networking sites, like ask.fm, enable abusive behavior by allowing users to send public messages to one another anonymously. Under the guise of anonymity, users of these services can send abusive messages to one another without being accountable for their actions. This abusive behavior has real world consequences.In this paper, we introduce iAnon, a tool that aims to support victims of cyberbullying by providing anonymous support through their social networks. iAnon automatically detects ask.fm users who are at risk and allows third party "do-gooders" to anonymously send friendly encouraging messages to victims. We hope that our tool, iAnon, can help mitigate feelings of depression and loneliness that are often felt by victims of cyberbullying by providing an online support system for cyberbullying victims.

Keywords:(social networks; interface design; cyberbullying mitigation)

Contact: parnia@umd.edu

1 Introduction

The advent of social media and social networks allow people to stay connected. Because social media has become so ubiquitous, it allows users to stay connected during all times of the day. While bullies used to only have limited access to their victims, the advent of social networking sites allows bullies to taunt their victims online at any time of day thus making it impossible for victims to seek respite from the torment of bullying. Anonymous question-asking social network services like ask.fm and formspring.me enable abusive behavior by allowing users to post anonymously (Kontostathis, Reynolds, Garron, & Edwards, 2013). Under the guise of anonymity, users of these services often send abusive messages to one another. This abusive behavior has recently resulted in teens committing suicide because of anxiety and loneliness. In this paper, we introduce iAnon, a tool that potentially may help prevent victims of cyberbullying feel anxious and depressed. iAnon detects ask.fm users who are at risk and allows people to anonymously post friendly encouraging messages on victim's pages, without publicly identifying them. By introducing iAnon, we hope that anonymous friendly messages can help reduce anxiety felt by victims of cyberbullying by creating an online support system that ultimately can prevent them from inflicting self-harm. We hope that our tool, iAnon, will help mitigate feelings of depression, isolation and loneliness that are often felt by victims of cyberbullying by creating an online support system.

2 Cyberbullying Data Collection

We collected data from the social networking site, ask.fm. Ask.fm allows users to ask questions from one another and provides the option to ask questions anonymously. We queried ask.fm through Google for variations of term "go kill yourself" and "go die". The search yielded nineteen public profiles, from which we collected user information. User profile information includes: username, user biography, user headline, 100 most recent questions with respective time posted, author information if applicable, and respective answer. Additionally, we collected the 25 "best" questions for each user, which also includes respective time posted, answer, author information (if question was non-anonymous), and likers of that particular question. Each user on ask.fm has a page in which the "best" questions are viewable. The "best" questions are questions that have received the most number of "likes" with respect to other questions in a user's network. In order to gain a better understanding of these nineteen user's networks, we collected user information from all public

profiles of those users, who had liked questions on the "best" question pages of the nineteen original profiles queried. This step of the data collection yielded over 40,000 user profiles, from which we again collected user profile information described above. We repeated this process for a second time, collecting the likes from the 19 user's "best" question page, ultimately yielding more than 40,000 user profiles. We chose this method of sampling because we wanted to collect data of users who were in the immediate neighborhood of individuals who were being cyberbullied.

3 Detection Methodology

We collected more than 8 million posts from over 40,000 users. In order to detect the users that were at risk for depression and anxiety, we first detected the individual posts that had traces of cyberbullying in them. If the ratio of detected bullied posts over the number of non-bullying posts in a user's profile is greater than a particular threshold that we defined, in this case, .4, we classified the user as being at risk and thus a candidate for the iAnon website. Our training data consisted of 13160 posts from spring.me, a question/asking with anonymous asking similar to ask.fm (Kontostathis et al., 2013). Kontostathis et al. used Mechanical Turk to label each post three times. A post was labeled as "Yes" or "No" depending on whether it contained traces of cyberbullying and rated on a scale of 0-10 for severity of bullying. We considered the following features for our classifier: whether the post includes a lexicon of pre-determined abuse words, whether the post includes a lexicon of pre-determined nice words, whether the question was asked by an anonymous person, whether the word "kill" or "die" was included in the question, whether the question or the answer was capitalized and length of question asked. Using a logistic regression classifier and the above features, we were able to accurately find traces of bullying. To conjure an accurate evaluation of our detection algorithm, we used 10-fold cross validation on the spring.me data which yielded an average precision, recall and f1-score of 0.914, 0.936 and 0.908 respectively. We used the algorithm on a sample of 15050 posts from the ask.fm and found that 2018 posts contained traces of cyberbullying. Based on our odds-ratio heuristics, from 324 profiles, 164 profiles were at risk for anxiety and depression.

4 Interface Design for Isolation Prevention

The iAnon interface design process consisted of several iterations amongst the team members before concluding the best design. Iterations of designs were judged on simplicity and usability. The details of our final interface are shown in Figures 1, 2, and 3. iAnon scrapes ask.fm daily for new profiles. Based on the cyberbullying detection method described above, the system identifies users who are at risk for depression and anxiety. In our final iteration of iAnon, we included a point system for iAnon users to keep track of the comments they have made.

5 Privacy Concerns

There are many ethical considerations when analyzing data that identifies users with behavioral ills (De Choudhury, 2013). As the name of this application implies, iAnon (i-anonymous), cares deeply about the anonymity of the victims of cyberbullying, which it is capable of detecting. Thus privacy of individuals is the utmost concern. Our application preserves the anonymity of users detected as being victims of cyberbullying. The name, user id, or any other identifiable information is never revealed through the interface (show in Figures 1, 2, and 3). We hope that the anonymity displayed in iAnon can inspire future designers who work with sensitive data to attempt to help people through interface design while still respecting the privacy of individuals.

6 Future Work and Evaluation of System

Based on Jakob Nielson's heuristics of user interface design, "visibility of system status" and feedback is an important component of any system (Nielsen & Molich, 1990). In future work, we want to continue scraping the profiles of people who are at risk to evaluate whether bullying on their profiles has increased or decreased from the time the iAnon message was sent. When iAnon users send positive message to someone

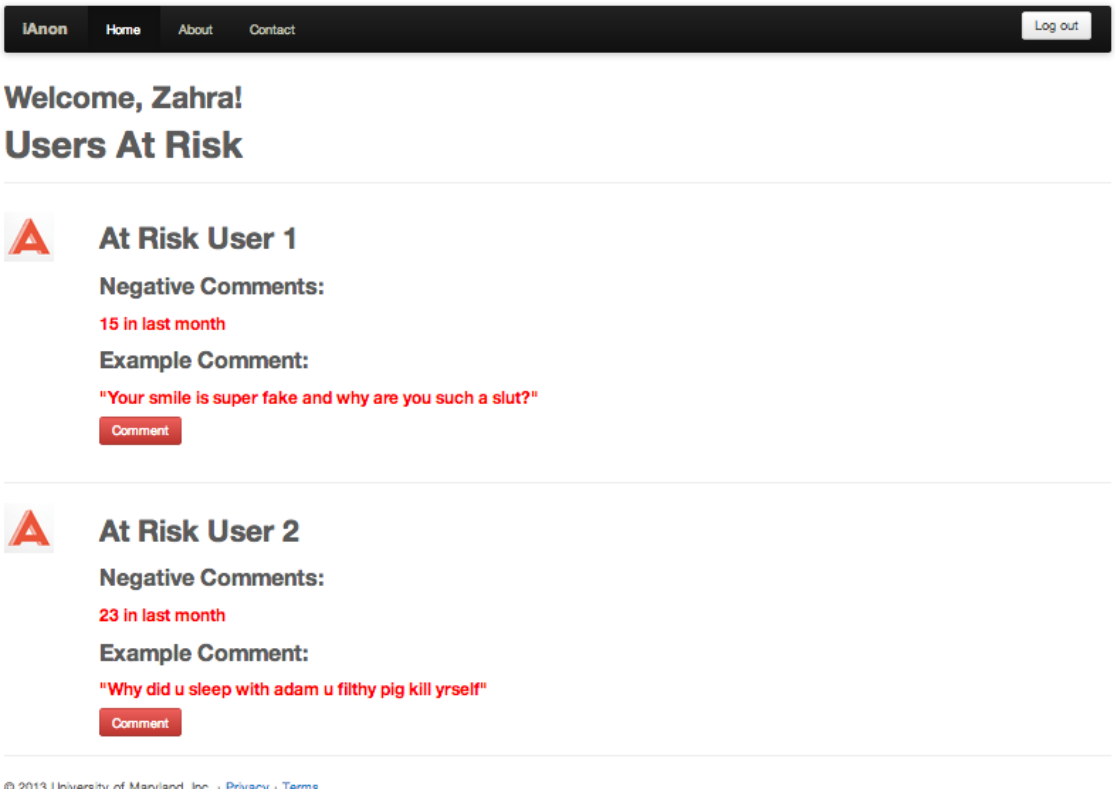

Figure 1: Interface that shows users who at risk of depression and anxiety based on the number of cyberbullying posts on their profile.

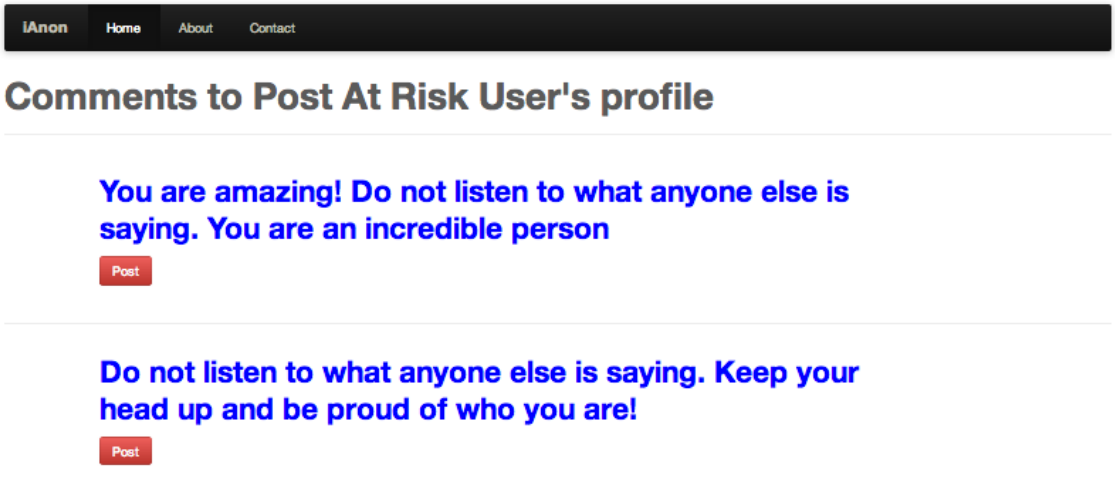

Figure 2: List of potential comments users can send to cyberbullying victim

who has been victimized by cyberbullying, they can accumulate points. In future work, users can compete for points and share their posts to social networks. The iAnon point system described above will then notify the iAnon user through the point dashboard in 3 that the cyberbullying victim's profile has been scraped and the number of negative posts on that user's wall have decreased. We also want to give users of iAnon the opportunity to formulate their own positive and encouraging messages. Once a certain threshold of points is reached, iAnon users will then be given the opportunity to formulate their own posts. To prevent iAnon from being abused, a third party administrator and a detection algorithm will evaluate the incoming comment to

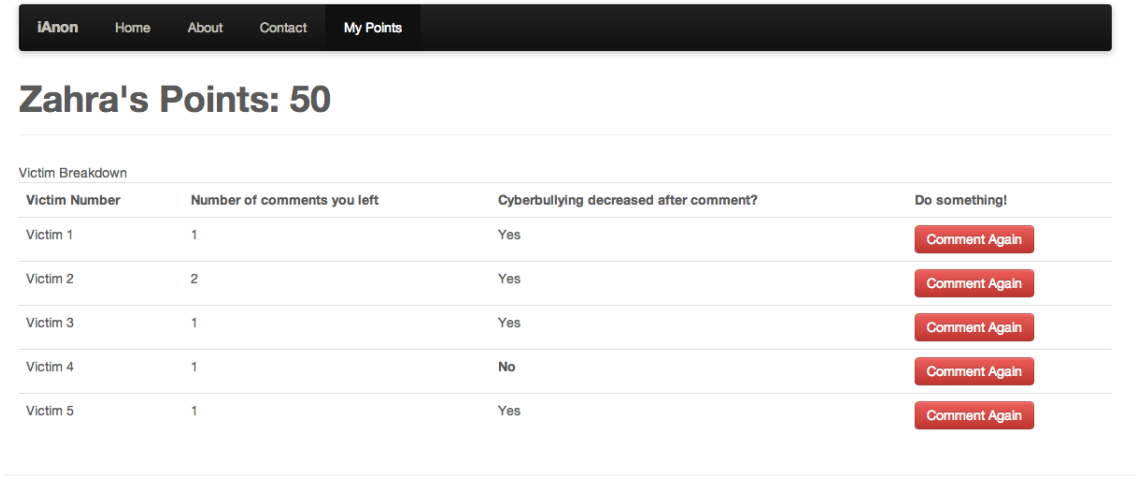

Figure 3: Dashboard of history of activity with cyberbullying victims and status of cyberbullying posts since last engagement with cyberbullying victim.

check that it is appropriate before it is posted.

7 Conclusion

Tools like iAnon can potentially help mitigate feelings of loneliness, depression, and anxiety that result from cyberbullying posts on sites that allow anonymous postings. In future work, we envision evaluating this tool among real users of tools like ask.fm and spring.me to see the emotional outcome of sending friendly anonymous posting. This project showed that it is possible to leverage existing technologies to build a tool that can detect cyberbullying posts with high accuracy, which enabled us to detect possible users at risk of depression and anxiety with high accuracy. The goal of our web application, iAnon, is to help mitigate feelings of depression and anxiety among victims of cyberbullying.

References

De Choudhury, M. (2013). Role of social media in tackling challenges in mental health. In *Proceedings of the 2nd international workshop on socially-aware multimedia* (pp. 49–52).

Kontostathis, A., Reynolds, K., Garron, A., & Edwards, L. (2013). Detecting cyberbullying: query terms and techniques. In *Proceedings of the 5th annual acm web science conference* (pp. 195–204).

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 249–256).

Table of Figures