

Understanding Types of Cyberbullying in an Anonymous Messaging Application

Arpita Chakraborty, Yue Zhang, and Arti Ramesh
SUNY Binghamton
{achakra4, yzhan202, artir}@binghamton.edu

ABSTRACT

The possibility of anonymity and lack of effective ways to identify inappropriate messages have resulted in a significant amount of online interaction data that attempt to harass, bully, or offend the recipient. In this work, we perform a preliminary linguistic study on messages exchanged using one such popular web/smartphone application—*Sarahah*, that allows friends to exchange messages anonymously. Since messages exchanged via *Sarahah* are private, we collect them when the recipient shares it on Twitter. We then perform an analysis of the different kinds of messages exchanged through this application. Our linguistic analysis reveals that a significant number of these messages (~ 20%) include inappropriate, hurtful, or profane language intended to embarrass, offend, or bully the recipient. Our analysis helps in understanding the different ways in which anonymous message exchange platforms are used and the different types of bullying present in such exchanges.

KEYWORDS

cyberbullying types, topic models, social media analysis, anonymous message exchanges

ACM Reference Format:

Arpita Chakraborty, Yue Zhang, and Arti Ramesh SUNY Binghamton {achakra4, yzhan202, artir}@binghamton.edu. 2018. Understanding Types of Cyberbullying in an Anonymous Messaging Application. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3184558.3191530>

1 INTRODUCTION

Recent times have witnessed a stupendous growth in online interactions. A significant portion of online interaction data is in the form of text such as posts on social networks, messages, and comments. With the proliferation of online interactions, there is also a corresponding increase in concern surrounding the nature of this content. Textual interactions that signify disturbing and negative phenomena such as online harassment, cyberbullying, cyber threats, stalking, and hatred are on the rise [2]. This detrimental online behavior can have significant traumatic effects on the individual experiencing this and can lead to severe psychological problems [25]. Furthermore, online data tends to be digitally preserved for a considerable length of time, aggravating the effect.

Anonymity has been shown to be a contributing factor in cyber harassment and bullying. Previous work on *Ask.fm* and *Yik-Yak* have shown that the possibility of anonymity significantly propels the number of cyberbullying messages [5, 12]. In this work, we focus on one such anonymous message exchange application that topped the download charts in the period between July–September, 2017 on App Store: *Sarahah*. The *Sarahah* mobile/web application can be added to other social networks such as Twitter and Facebook, hence allowing users to send anonymous messages to their friends in the network. Though the app was originally designed as a platform for exchanging anonymous messages, it soon transformed into a breeding ground for hate [3]. Most previous work on cyber bullying has been in settings such as *Ask.fm* and *Youtube*, where the people exchanging bullying or hateful comments need not necessarily know each other at a personal level. Our analysis especially brings forth the amount of negative content in messages exchanged between “friends”, making it more personal than other instances of bullying.

In this work, we perform an initial study on the types of messages exchanged through the *Sarahah* application on the Twitter social network. Since we don’t have direct access to this data through *Sarahah*, we collect it via Twitter, when the recipient shares the *Sarahah* message on their Twitter feed, sometimes along with a brief response to the message. Our dataset contains messages exchanged between August–October, 2017, which is when the application’s popularity peaked. While the sender remains unknown in this setting, the recipient is known, as he/she shares this message on Twitter. As only a portion of messages exchanged via the *Sarahah* platform are potentially shared by the recipients on Twitter, this data may not represent all the messages exchanged through *Sarahah* and can potentially be biased because of its collection using the Twitter search API [18]. Despite these possible biases, the unique characteristics of this data (anonymous sender, exchanged between friends on a social network, and presence of recipients’ responses) make this an important source of information to understand the different kinds of bullying present in online interactions.

We leverage topic modeling (also known as Latent Dirichlet Allocation (LDA)) to perform a linguistic analysis on the different types of conversations and bullying categories present in this data. We first identify the LDA topics that are related to bullying using the top words in each of these topics. By filtering the *Sarahah* messages in the LDA topics that are related to bullying, we observe that 20% of these messages fall in the bullying topic categories. We also observe that most bullying events happen when an anonymous sender shares specific opinions/confesses about their true feelings toward the recipient or asks embarrassing personal questions. Our analysis paves way for understanding the different types of conversations and bullying topic categories in anonymous message exchanges.

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191530>

2 RELATED WORK

Detecting and understanding bullying on social media has received considerable interest in the recent years. Corcoran et al. [10] and Hosseinmardi et al. [13] draw attention to the broader issue of cyber aggression rather than cyberbullying. Hosseinmardi et al. [13] identify media sessions on Instagram that have at least one profane word in their comments by users other than the profile owner.

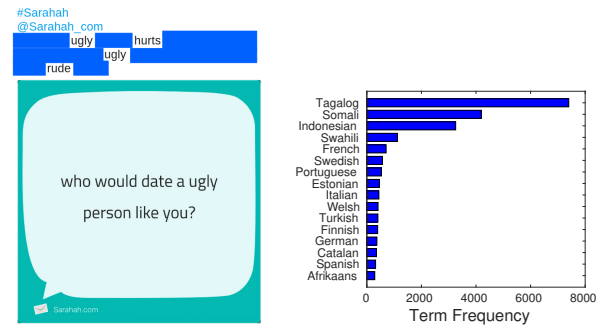
Raisi et al. [23, 24] propose a participant-vocabulary consistency model for identifying the instigators and victims of bullying in a social network and simultaneously building a bullying vocabulary by starting with a corpus of social interactions and a seed dictionary of bullying indicators. They evaluate the model on data from Twitter and Ask.fm and show that the proposed method can detect new bullying vocabulary as well as victims and bullies. Several work consider social interaction features along with textual features to detect cyberbullying [12, 14, 19]. Bigelow et al. [8] use latent semantic indexing to detect cyberbullying. There is also previous work on detecting abusive and hateful speech targeting specific groups including ethnicity, origin, religion, gender, sexual orientation and physical appearance [11, 28]. Dinakar et al. [11] show that individual classifiers perform better than multi-class classifiers for this problem on a Youtube comments dataset. Li et al. [15] analyze the negative and positive sense of the words on Instagram and Ask.fm networks. Margono et al. [17] analyze bullying patterns in Indonesia on Twitter. Whittaker et al. [29] examine the prevalence of cyberbullying in college students. Bellmore et al. [6] and Tokunaga et al. [27] study the socio-psychological issues of bullying in social media data.

There is also previous work that do not especially focus on cyberbullying or abusive language, but on similar problems. Nguyen et al. [20] study the k -suspector problem aiming to identify the k most suspected users in online social network from the set of victims who are already influenced by the misinformation. Mahendiran et al. propose a novel unsupervised learning algorithm which builds dynamic vocabularies using probabilistic soft logic in order to understand the true membership within the social group and capturing the dynamic trends and forecasting specific to election campaigns from eight different countries [16]. Bifet et al. [7] propose a sliding window Kappa statistic for mining opinions and analyzing sentiment in evolving Twitter data streams.

3 DATA

In this work, we focus on data from a recent anonymous mobile application, *Sarahah*. It entered the US Apple Store in June 13, 2017 and gradually spread out to Canada, India, and a few other countries. The popularity of the application spiked after an update that allowed people to share Sarahah messages was launched by Snapchat on July 5, 2017. Gradually, it became the top rated application in App Store leaving behind all popular social media applications such as Twitter, Facebook, and Snapchat. While this application was originally created to exchange anonymous messages, it soon became a breeding ground for hate and a platform for cyberbullying as users started posting threatening, hurtful, profane, and pornographic messages [3].

In this paper, we present analysis on data collected from 30th August, 2017 to 15th October, 2017, which coincides with the peak popularity period of Sarahah. We collect messages exchanged using Sarahah on Twitter by searching for images with the hashtag #Sarahah using the Twitter search API [1]. Since Twitter allows you to only extract tweets posted in the last week, we collect at one-week intervals during the specified period. We also extract Sarahah messages by crawling specific users' Twitter accounts. Figure 1(a) shows an example of a message exchanged through *Sarahah*. Since the messages exchanged using Sarahah are in the form of images, we use Google's optical character recognition software to extract text from the images [26]. The extracted data has three components: i) the textual message exchanged using Sarahah, ii) user's reaction to the message when the user shares this message on Twitter, and iii) other user-related information extracted from user's profile.



(a) Example Sarahah message and corresponding user reaction (b) Distribution of different languages in our Sarahah dataset

Figure 1: Figures showing an example message from Sarahah (left) and statistics of different languages in our data (right).

Message	Reaction
How can we s*xchat - can't wait to make love to you.	It's not free dear.
Can we meet and have random s*x?	Yes, only if you are the last man on earth.
Why you so ugly?	Cause b*tch your momma had me.

Table 1: Example messages exchanged via Sarahah and corresponding user reactions

Since Sarahah messages are generally from friends, they tend to be also in languages other than English. In our extracted dataset, we found the presence of several languages including German, French, Afrikaans, Swahili, Interlingua, Finnish, Somali, Czech, Tagalog, English, Romanian, Moldavian, Moldovan, Croatian, Panish, Norwegian, Latvian, Welsh, Portuguese, Catalan, Danish, Swedish, Estonian, Dutch, Slovenian, Italian, Albanian, Hungarian, Polish, Slovak, Turkish, Lithuanian, Vietnamese, Hindi, and Tamil. English remains the most popular language in our dataset. We report the distribution of messages across the top 50% languages other than English in Figure 1(b). We use Google's language detection library *langdetect*

to detect the language and convert the message and users' response to English for our analysis [21]. Since we follow a bag-of-words approach, translation errors have little effect on our analysis as they tend to occur mostly in the sentence construction.

In all, we collect 82,193 Sarahah messages and corresponding user reactions. Removing duplicates and empty messages, we have 76,278 messages and corresponding user responses in our dataset. We perform standard NLP preprocessing techniques of stop-word removal, tokenization and stemming using porter stemmer on the messages and responses.

Table 1 and Figure 1(a) give some examples of messages exchanged using Sarahah and corresponding user reactions. The first message is a sexually abusive message, the second is a casual message containing some sexually offensive words, and the third message expresses hate towards the recipient. The user reactions corresponding to these messages give some of the different ways in which people respond to offensive messages. From the first and second responses, it is unclear how much the message affected the recipient. While the user receiving the third bullying message responds to it by using offensive words, such as *b*tch*, in an attempt to hurt the sender. We observe a similar behavior in Figure 1(a), where the response contains words such as *hurt* and *rude* indicating that the user is affected by this message. The recipient also uses words such as *ugly* in the response in an attempt to hurt the sender. These user responses to Sarahah messages are a very unique aspect of our dataset.

4 LINGUISTIC ANALYSIS OF SARAHAH MESSAGES

We first perform linguistic analysis of the messages exchanged using Sarahah. Our analysis paves way for understanding the nature of these messages and identifying the different bullying categories.

4.1 Topic Analysis using Latent Dirichlet Allocation (LDA)

Topic modeling, also known as latent Dirichlet allocation (LDA) is a popular means to analyze document corpora [9]. We first start by using LDA to understand the presence of different bullying related words in our data. We consider each message as a document and run LDA for 10,000 iterations. We use standard values of $\alpha = 0.01$ and $\beta = 0.01$ for the hyperparameters and 30 topics. Using the LDA topics, we identify the different types of conversations present in our data and the types that are more likely to contain bullying messages. We then consider the top 50 words from topic-word distributions of these 30 topics. We identify the *bullying* words from this set of words and group them into different categories based on the type of bullying.

4.2 Types of Conversation using Sarahah

Analyzing the LDA topics, we identify the different ways in which users use Sarahah. We map the different types of messages to standard conversation types [22]. Table 2 gives us the different types of conversation and sub-categories in them, and some example messages in each category. We identify *four* types of conversations on Sarahah; we describe them briefly below:

Confession. Messages that are meant to convey feelings anonymously that the sender does not wish to disclose under non-anonymous circumstances. The messages in this category range from secret admiration, flirting, and love proposals to bullying messages such as hatred/sexually offensive messages. We break down messages in this category into three sub-categories: i) *positive*, ii) *romantic*, and iii) *negative* messages. Positive confessions mention positive feelings about a person without the explicit mention of romantic feelings. The second sub-category captures confessions that convey romantic interests as noted by the presence of words such as *crush* and *love*. The third sub-category captures confessions that convey feelings of malice or hatred towards the recipient, indicated by words such as *punch*, *vile*, *disgusting*, and *waste*. We identify that the *negative* and the *romantic* messages can potentially cause discomfort to the recipient.

Questions. In this conversation category, we find questions directed to the recipient. This category contains both general (i.e., non-intrusive) questions as well as intrusive and offensive personal questions. Note that in the examples under personal questions in Table 2, we find both intrusive personal questions such as *What is your phone number?* to offensive ones such as *Do you like an*l s*x?*.

Opinions on world issues. There are also opinions on world issues, especially sensitive ones, which users may not be comfortable sharing in a non-anonymous setting. Some examples of world issues we see in our data are related to: i) President Trump, ii) India-Pakistan partition, and iii) North Korean politics.

Inspirational. We also find some inspirational messages which may/may not be targeted towards the recipient.

We identify that bullying messages primarily occur in the *confession* and *questions* messages.

4.3 Bullying Categories

Using our LDA topics, we identify the different prominent bullying categories present in this data. The sensitive nature of this data restricts creating and sharing labeled data, making unsupervised methods lucrative for this problem. In order to overcome the need for message-level labels, we label the LDA topics based on the presence of bullying words in them. We find words related to the following bullying categories in the LDA topics.

Sexual. Messages in this category contain explicit sexually offensive words that are intended to harass, intimidate, or make the recipient uncomfortable. There are both confessions, which convey the sender's feeling towards the recipient and personal questions that are intended to make the recipient uncomfortable.

Hate. Messages in this category are intended to convey hatred and emotionally unsettle the recipient. Messages that convey hatred, death threats, and emotional/physical abuse belong in this bullying category.

Inappropriate flirting. The messages in this category are intended to convey a romantic interest toward the recipient. While this may not be considered bullying under normal circumstances, since the

Conversation Types	Sub-categories	Example Messages
Confession	Positive	I always had a crush on you, the way you smile the way how your beautiful eyes are. You're actually so pretty, people will always hate no matter what only listen to the opinions from people who matter most in your life.
	Romantic	Hi crush. I love you. And I love u even more when you tweet my messages. Love you more than anything
	Negative	I want to punch you. In your face. You are such a waste of oxygen. I think you're vile, disgusting, and deserve nothing but grief in your life. Pathetic waste.
Questions	General	What inspired you to photography? Food or volleyball?
	Personal	Do you have a girlfriend? Can I have Your Number? How many times you literally got *ss f*cked?
Opinions on World Issues		North Koreas trade China so Trump can threaten China Pakistani think Partition bad idea Jinnah Nehru didn't lose anything and in fact became legends in respective countries. We are the victims. The decision was made without our consent. Pakistan has started to fence its Afghan border and there is a deep profound message in it for Trump
Inspirational		Life is one big road with lots of signs. So when you ricling through the ruts, don't complicate your mind. Flee from hate, mischief and jealousy. Don't bury your thoughts, put your vision to reality. Wake Up and Live!

Table 2: Different conversation types found in Sarahah messages and example messages in each conversation type.

Coarse-grained Bullying Categories	Words
Sexual	s*xy, *ss, b*tch, gay, hot, f*ck, b**bs, d*ck, s*ck, seductive, b**ty, virgin, lesbian, bl*wj*b, straight, homosexual, b*tt, h*rny, h*es, trans, lick, bite, bed, naked, wh*e
Hate	punch, shoot, kick, fat, bullsh*t, beast, threat, fight, death, rude, ruin, sh*t, slap, ugly, abuse, betray, harm, size, ego, loathe, sad, cheat, trash, pain, tear, cry, emotion, breakup, trap, annoy, heartless, loser
Inappropriate Flirting	crush, dreams, appeal, stalk, babe, crave, love, proposal, hit, cheek, sweetie, baby, candy, babe, look, pie, cutie, hug, chick, romance, desires, pleasure, bomb
Admiration	beautiful, amazing, smile, awesome, kind, pretty, heart, great, gorgeous, nice, handsome, sweet, funny, hilarious, smart, strong, laugh, adorable, appreciate, proud, laugh, good, like, decent, positive, inspiration, perfect, blessings, genuine, courageous, brighten, honest, respect

Table 3: Coarse-grained topic categories in Sarahah and representative words in each category

sender is anonymous, this can potentially cause significant distress to the recipient. We see messages that imply stalking, divulging/asking for personal secrets, and usage of words implying romantic interests, which especially under circumstances when the sender is anonymous can unsettle the recipient. This makes it an important bullying category to study.

We present the different bullying categories and top words in each category in Table 3. If we filter the messages using the top words in the bullying categories 1 – 3 in Table 3, we find that around 20% of the messages contain one or more of these words, which is considerably high given that the messages are exchanged between users who are friends on the social network. In addition to the three bullying categories, we also find messages that convey admiration (category 4 in Table 3). While the messages in this category are

mostly positive ones, some messages do have a touch of flirting in them. The subtle bullying words that could possibly be present in messages in this category make it an important category to study.

Table 4 gives some example messages in each bullying category. Bullying words are highlighted in italics. Notice that the first two categories have profane/offensive/hurtful words making it easier to automatically flag them. However, messages in the flirting category have words such as *kiss*, *hug*, which in an anonymous setting could potentially cause discomfort to the recipient but is harder to detect automatically using vocabulary-based approaches. Also, these messages combine positive words such as *love*, with possibly concerning words such as *kiss*, *hug*, and *shape*, making it necessary to

understand the semantics of these messages and their corresponding user responses to accurately determine whether they could potentially unsettle the recipient.

Bullying Category	Example post
Sexual	Do you wanna s*x with me?
	Is small d*ck a turn off?
	Leak your d*ck pics?
Hate	I hate you please leave the twitter.
	Go cut yourself lol. Why you an ugly dumb sl*t?
	Why are you so pretty b*tch? It just makes me hate my genes.
Flirting	I'd love to steal a kiss from you one day.
	I wish I could hug and kiss you all day long.
	In love with the shape of you.

Table 4: Bullying topic categories and some example messages in each category.

5 DISCUSSION AND FUTURE WORK

In this work, we introduced data collected from a recently released anonymous web/mobile messenger application, *Sarahah*. We performed a preliminary linguistic analysis of messages and found that cyberbullying can occur in different forms, with or without the presence of profane words, calling for a fine-grained analysis. Our analysis is helpful in identifying the different ways in which anonymous applications are used and understanding the different types of bullying present in anonymous exchanges between people who already know each other. There are several exciting directions to go from here. The unique aspect of the user responses when recipients share these messages opens up possibilities for studying the varying levels of discomfort caused by bullying messages. Our data and subsequent analysis could potentially help in answering many important questions related to cyberbullying. The first and foremost question is what kinds of messages cause the most distress to recipients. Understanding the effect that bullying messages has on the recipient can be useful in identifying the severity of bullying. We observe that not all recipients react the same way to messages in the same bullying category. So, a related question is, can we take the recipients' characteristics and background to predict what kind of messages could possibly be hurtful to the recipient? Understanding some important characteristics of people who get bullied the most can be helpful in taking preventative action towards bullying. Lastly, can we design models that possess a superior semantic understanding of cyberbullying and can identify different types of bullying that can occur with/without the presence of profane/offensive words. Developing models that can detect the different subtle signals present in bullying messages and responses (i.e., sarcasm, emotional, and psycholinguistic signals) will help in accurately identifying different forms of bullying and foster a deeper understanding of their impact on recipients.

REFERENCES

- [1] [n. d.]. pypi.python.org Twitter API python library. ([n. d.]). <https://pypi.python.org/pypi/twitter>
- [2] [n. d.]. StopBullying.gov Webiste, Federal Govt. <https://www.stopbullying.gov/>. ([n. d.]). Accessed: 2017-09-12.
- [3] [n. d.]. StopBullying.gov Webiste, Federal Govt. <https://cyberbullying.org/sarahah>. ([n. d.]). Accessed: 2017-09-12.
- [4] [n. d.]. trends.google.com Sarahah trends by Google Trends. <https://trends.google.com/trends/explore?q=sarahah>. ([n. d.]). Accessed: 2017-10-30.
- [5] Elias Aboujaoude, Matthew W Savage, Vladan Starcevic, and Wael O Salame. 2015. Cyberbullying: Review of an old problem gone viral. *Journal of Adolescent Health* 57, 1 (2015), 10–18.
- [6] Amy Bellmore, Angela J Calvin, Jun-Ming Xu, and Xiaojin Zhu. 2015. The five ways of bullying on twitter: who, what, why, where, and when. *Computers in Human Behavior* 44 (2015), 305–314.
- [7] Albert Bifet and Eibe Frank. 2010. Sentiment knowledge discovery in twitter streaming data. In *Proceedings of the International conference on Discovery Science*.
- [8] Jacob L. Bigelow, April Edwards (Kontostathis), and Lynne Edwards. 2016. Detecting Cyberbullying Using Latent Semantic Indexing. In *Proceedings of the Workshop on Computational Methods for CyberSafety*.
- [9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (March 2003), 993–1022.
- [10] Lucie Corcoran, Conor Mc Guckin, and Garry Prentice. 2015. Cyberbullying or cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression. *Societies* 5, 2 (2015), 245–255.
- [11] Karthik Dinakar, Roi Reichart, and Henry Lieberman. 2011. Modeling the detection of Textual Cyberbullying. *The Social Mobile Web* 11, 02 (2011).
- [12] Homa Hosseinmardi, Amir Ghasemianlangroodi, Richard Han, Qin Lv, and Shivakant Mishra. 2014. Towards understanding cyberbullying behavior in a semi-anonymous social network. In *Proceedings of the Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.
- [13] H. Hosseinmardi, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra. 2016. Prediction of cyberbullying incidents in a media-based social network. In *Proceedings of the Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.
- [14] Qianjia Huang, Vivek Kumar Singh, and Pradeep Kumar Atrey. 2014. Cyber bullying detection using social and textual analysis. In *Proceedings of the Workshop on Socially-Aware Multimedia*.
- [15] Homa Hosseinmardi Shaosong Li, Zhili Yang, Qin Lv, Rahat Ibn Rafiq Richard Han, and Shivakant Mishra. 2014. A comparison of common users across Instagram and ask.fm to better understand cyberbullying. In *Proceedings of the Conference on Big Data and Cloud Computing (BdCloud)*.
- [16] Aravindan Mahendiran, Wei Wang, Jaime Arredondo Sanchez Lira, Bert Huang, Lise Getoor, David Mares, and Naren Ramakrishnan. 2014. Discovering evolving political vocabulary in social media. In *Proceedings of the Conference on Behavior, Economic and Social Computing (BESC)*.
- [17] Hendro Margono, Xun Yi, and Gitesh K Raikundalia. 2014. Mining Indonesian cyber bullying patterns in social networks. In *Proceedings of the Australasian Computer Science Conference*.
- [18] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M Carley. 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose.. In *Proceedings of the Conference on Social and Information Networks*.
- [19] Vinita Nahar, Xue Li, and Chaoyi Pang. 2013. An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering* 3, 5 (2013), 238.
- [20] Dung T Nguyen, Nam P Nguyen, and My T Thai. 2012. Sources of misinformation in online social networks: Who to suspect?. In *Proceedings of the Conference on Military Communications Conference (MILCOM)*.
- [21] Thomas Piehn and Allison Piehn. 2001. Voice enabled digital camera and language translator. (Feb. 20 2001). US Patent App. 09/789,220.
- [22] Charis Psaltis and Gerard Duveen. 2007. Conservation and conversation types: Forms of recognition and cognitive development. *British Journal of Developmental Psychology* 25, 1 (2007), 79–102.
- [23] Elaheh Raisi and Bert Huang. 2017. Cyberbullying Detection with Weakly Supervised Machine Learning. In *Proceedings of the Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.
- [24] Elaheh Raisi and Bert Huang. 2017. Cyberbullying Detection with Weakly Supervised Machine Learning. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.
- [25] Shari Kessel Schneider, Lydia O'donnell, Ann Stueve, and Robert WS Coulter. 2012. Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health* 102, 1 (2012), 171–177.
- [26] Ray Smith. 2007. An overview of the Tesseract OCR engine. In *Proceedings of the Conference on Document Analysis and Recognition (ICDAR)*.
- [27] Robert S Tokunaga. 2010. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior* 26, 3 (2010), 277–287.
- [28] William Warner and Julia Hirschberg. 2012. Detecting hate speech on the world wide web. In *Proceedings of the Workshop on Language in Social Media*.
- [29] Elizabeth Whittaker and Robin M Kowalski. 2015. Cyberbullying via social media. *Journal of School Violence* 14, 1 (2015), 11–29.