Computers in Human Behavior 28 (2012) 861–867

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Anonymity and roles associated with aggressive posts in an online forum

Michael J. Moore [a,*], Tadashi Nakano [b], Akihiro Enomoto [c], Tatsuya Suda [d]

[a] Frontier Research Center, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Frontier Research Base for Global Young Researchers, Frontier Research Center, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
[c] University of California – Irvine, Department of Computer Science, Irvine, CA, USA
[d] University Netgroup Inc., P.O. Box 1288, Fallbrook, CA 92088, USA

ARTICLE INFO

Article history:
Available online 9 January 2012

Keywords:
Online aggression
Cyberbully
Defender
Anonymity

ABSTRACT

Cyberbullying is a growing concern in online communications. Cyberbullying has negative impacts such as distress or suicide of a victim. One common type of cyberbullying attack utilizes aggressive forum posts to insult or threaten a victim. Automated tools to classify cyberbullying may aid in avoiding or reducing the negative impacts of cyberbullying. One approach to produce an automated tool is to identify features of forum posts which may be indicators of cyberbullying. One feature of a forum post is the role of the author of the forum post, such as a bully, victim, or defender. Another feature is whether the forum post insults or threatens an individual (e.g., contains insults directed at a victim). Attackers may use aggressive forum posts to attack someone and defenders may use aggressive forum posts to retaliate against attackers. Another feature is whether the communication is anonymous (e.g., sending forum posts with no identifier) since cyberbullies utilize anonymity to reduce the ability of the victim to defend themselves and to shield the cyberbully from social consequences. In this paper, forum posts were labeled in an online forum for these features. Text matching techniques had some success in identifying aggressiveness forum posts including both attacks and defends. Anonymity of forum posts (i.e., forum posts with no identifier) was identified as a criterion to distinguish attackers (more anonymous relative to non-aggressive communications) from defenders (less anonymous relative to non-aggressive communications).

1. Introduction

Recent research has identified the internet and wireless communications as a new environment for bullying behavior (Kowalski & Limber, 2007; Li, 2007; Ybarra & Mitchell, 2004). Traditional bullying (i.e. face-to-face bullying) occurs when a person is “exposed, repeatedly and over time, to negative actions on the part of one or more other persons” (Olweus, 1990). Face-to-face bullying has been found to occur in a variety of forms in many countries (Smith, Cowie, Olafsson, & Liefooghe, 2002; Taki et al., 2008). Cyberbullying is a new form of bullying which can occur through email messages, forum posting, web site publishing, cell phone text messages, chat rooms, computer hacking, and any other means of electronic communication (Kowalski & Limber, 2007; Li, 2007; Ybarra & Mitchell, 2004). For example, cyberbullies use email, cell phone text messages, or forum posts to send insults or death threats to victims. Another example is a cyberbully creating web pages to insult a victim or perform polling to vote negatively about a victim (Li, 2007). Yet another example is a cbyerbully using cell phones and video recording to attack through “happy slapping”, which is filming face-to-face bullying attacks, such as slapping, and circulating the videos through cell phones (Smith et al., 2008). Yet another example is socially attacking a victim by recruiting others to send anonymous e-mail or cell phone text messages (Dooley, Pyzalski, & Cross, 2009), by posting contact information of a victim, or by spreading rumors about a victim Li, 2007.

Recent research shows that cyberbullying has several differences from face-to-face bullying. In cyberbullying, communications can be anonymous (e.g., communication is associated with an unknown identifier) and attacks can come from a broader audience relative to face-to-face bullying (e.g., public forums may invite individuals from other communities to join in attacks) (David-Ferdon & Hertz, 2007; Li, 2007). Cyberspace lacks supervision, attacks can be technical in nature (e.g., attackers can lock or take control of devices of a victim), and cyberbullies can use technology to coordinate attacks (Patchin & Hinduja, 2006). Cyberbullying attacks can happen anywhere and anytime (e.g., an individual can receive messages at home at night) and attacks are impersonal compared to face-to-face bullying (Kowalski & Limber, 2007). Attacks using forums or websites have some degree of permanency and thus it is possible for a single cyberbully incident to achieve an effect similar to repetitive attacks (Dooley et al., 2009). Visual

[*] Corresponding author. Tel.: +81 06 6879 7199.
E-mail addresses: mikemo@wakate.frc.eng.osaka-u.ac.jp (M.J. Moore), tnakano@wakate.frc.eng.osaka-u.ac.jp (T. Nakano).

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.12.005

and audio media (e.g., calls through mobile phone and video postings) can also have a larger impact on victims than email or cell phone text messages since cell phone attacks and video postings were more known among peers (Smith et al., 2008).

Cyberbullying has resulted in significant psychological and social impact for both victims and bullies. There are several publicized suicides which involved bullies attacking through cyberbullying (Hinduja & Patchin, 2010). Cyberbullying has been identified as a contributing factor for depression and suicide of victims (Hinduja & Patchin, 2010) as well as feelings of anger by victims (Li, 2007). Victims of cyberbullying may feel distress as the result of online interactions and cell phone text messages which threaten or embarrass the victim (Ybarra, Mitchell, Wolak, & Finkelhor, 2006). Victims can become very distressed in certain situations such as youth being cyberbullied, adults cyberbullying children, requests for pictures, or bullies attacking online coupled with offline contact (e.g., bullies attacking online coupled with bullies calling the phone or visiting the home of a victim) (Ybarra et al., 2006). In addition to victims, cyberbullies are also at risk for psychosocial problems such as poor parent/child relationships, substance use, delinquency and depression (Ybarra & Mitchell, 2004).

1.1. Research framework

The objective in this paper is to progress towards developing tools to automatically detect potential cyberbullying. An automated tool to detect cyberbullying may help reduce the occurrence of cyberbullying or negative impacts associated with cyberbullying. The automated tool should analyze, for example, a set of forum posts and produce whether cyberbullying occurs in the forum posts. In order to facilitate the detection of cyberbullying, the automated tool should provide annotated forum posts which most strongly indicated the potential for cyberbullying. One approach to identify cyberbullying is to build a classifier to identify patterns in the set of forum posts which match patterns from examples of cyberbullying. A first step in building a classifier is to extract features from forum posts which may be relevant to cyberbullying, such as the aggressiveness or anonymity of forum posts. Then, the classifier is trained with sets of features and whether the forum posts have an incident of cyberbullying. Finally, the classifier takes an unclassified set of forum posts and determines whether the forum posts are similar to forum posts with incidents of cyberbullying.

Cyberbullies use a very broad and ever-moving technology. It is important to begin understanding and characterizing the various social interactions, such as forum posts, on the internet which are harmful. In this paper, the authors evaluate forum posts from the online forum Formspring.me. Formspring.me was chosen since it has been associated with an incident of bullying resulting in suicide (Lewin, 2010). Understanding cyberbullying in forum posts may help build an understanding of cyberbully attacks and provide a step towards modeling cyberbully attacks. Also, forum posts are important since they appear in a variety of attacks and have features which may increase the damage of an attack such as anonymity, permanency (i.e., one post can be viewed many times), and public visibility (e.g., public viewing allows other individuals to join an attack). Analysis of forum posts is also a promising direction since there are a variety of automated techniques for identifying features of text contained in forum posts. Furthermore, forum posts can appear in response to other media (e.g. photos or video), thus identifying cyberbullying in forum posts may lead to identifying other damaging media.

This paper focuses on the first step of extracting several features which are expected to be relevant to identifying cyberbullying. Cyberbullying has been associated with name-calling or verbal attacks in textual messages (Li, 2007), referred to as "aggressive forum posts" in this paper. One possibly relevant feature is the use of profanity in a forum post which may indicate that the forum post is aggressive. Another feature is the target of an aggressive forum post. The forum post may be an aggressive attack against a victim or an aggressive defense of the victim. Cyberbullying is also described as often being anonymous. Thus, another relevant feature is whether an attacker uses an identifier or has some degree of anonymity.

Research issues in this paper include measuring (1) how automatically labeled aggressiveness correlates with attacks and defends, (2) how automatically labeled anonymity correlates with aggressiveness and roles, and (3) how roles correlate with each other. These research issues help build an understanding of the features related to cyberbullying and provide hints towards criteria which may be useful to automatically identify cyberbullying.

2. Online forum analyzed

In this paper, the authors evaluate forum posts from the online forum Formspring.me. Formspring.me was chosen since it has been associated with an incident of bullying resulting in suicide (Lewin, 2010). In the website each user manages their own forum page. Each user has an identifier in the online forum and the identifer is not required to match their real life identity. The intended use of the forum is for interviewing the owner of the forum page, thus the topic of the forum page is typically about the owner of the forum page. The owner of a forum page can restrict access to their forum page by a white list or black list.

Each forum page has a list of forum posts and a list of connections (i.e. other users to whom the owner of the page frequently posts). A forum post is a textual message placed on a forum page. A user can send a forum post to the owner of the forum page. An individual can send the forum post without their identifier (i.e., the forum post appears anonymous). The owner of the forum page can respond with a textual message in which case the forum post and response appears on the forum page of the owner. If the owner of the forum page does not respond, the forum post does not appear on the forum page. The owner of the forum page is responsible for moderating their forum page and for choosing what appears on their forum page.

This paper focuses on cyberbullying, however the techniques described in this paper may be related to other potentially problematic uses. Examples of other problematic uses include harassment (aggression targeting known individuals), trolling (untargeted inflammatory forum posts), impersonation of individuals, unauthorized gathering of private information, adult predators sexually targeting youth, and spamming. For example, "trolling" involves posting inflammatory forum posts to illicit a response from anyone (Baker, 2001). Cyberbullying is different from trolling since cyberbullying targets a specific individual (i.e., the owner of the forum page).

2.1. Method used to gather forum posts

Public forums pages were downloaded from Formspring.me to perform analysis. A number of random forum pages were chosen from search results for forum pages with common English names. Forum pages were required to have at least 4 forum posts, forum posts on the forum page generally in English, and to be open to the public. A total of 26 forum pages were downloaded. A total of 5230 forum posts were downloaded. The number of forum pages is not expected to affect labeling of individual forum posts on the forum pages. Only the first 1000 forum posts of any given forum page were considered for labeling. Pages closed to the public (not considered in this paper) may be more strongly moderated by their owners and have less anonymity; therefore, the analysis of anonymity discussed in this paper may not apply to forum pages closed to public. The acquired forum pages are a snap shot. During the time necessary to acquire forum pages, forum post posts and users accounts may be added or deleted from the system. It is relevant to perform analysis even when data is not complete, since authorities analyzing a forum may need to make conclusions about cyberbullying with incomplete data.

3. Aggressive messages

Cyberbullying has been associated with name-calling or verbal attacks in textual messages (Li, 2007), referred to as “aggressive forum posts” in this paper. The aggressiveness of a forum post is expected to be a useful feature for identifying cyberbullying.

In this paper, the authors apply manual and automated techniques to identify aggressive forum posts.

Manual labeling applies human processing to identify aggressive forum posts and is a time-consuming approach to label communications. For this paper, the authors defined aggressive forum posts as verbal insults or attacks directed at an individual or someone associated with the individual (e.g. a friend or relative of the individual). The authors used this definition to subjectively label each forum post as aggressive or not aggressive. Aggressive forum posts identified included text describing threats of violence or stalking, sexual acts, and insults about physical or social characteristics of individuals.

Automated labeling applies text analysis techniques to identify aggressive forum posts and is an alternative approach to manual labeling. Yin et al. (2009) identify relatively aggressive posts within a thread (a “thread” is a set of grouped textual messages which is similar to the concept of “forum page” in this paper) as posts containing specific sentence structure along with profanity words, second person pronouns (e.g., you, your, yourself, etc.), third person pronouns, and relatively more profanity than other posts in the same thread. The automated labeling technique may be inaccurate when an aggressive forum post does not contain profanity, when profanity or pronouns are misspelled, or when the forum posts are not in the language matching the profanity words. Attackers may specifically misspell words in aggressive posts to prevent text filters from automatically identifying and removing the aggressive words. The automated technique does not distinguish between aggressive attacks against a victim or aggressive defends of a victim.

For this paper, the authors applied a simplified version to automatically label aggressive forum posts as follows. Since forum posts are generally directed towards the owner of a forum page in the online forum considered, a forum post with second person pronouns and profanity was considered a sufficient indicator for an aggressive forum post directed at the forum page owner. The second person pronoun typically implies the forum page owner, and profanity often implies negative sentiment towards the forum page owner. A list of profanity was acquired from an open source profanity list for a web browser profanity filter. Sentence structure is ignored since many forum posts lack grammar and punctuation which causes identifying sentence structure to be very complex. Profanity on a forum page was not normalized relative to the other forum posts in the forum page. The simplifications in this paper are not as strict as the technique described in Yin et al. (2009). As a result, the automated technique applied in this paper is expected to label relatively more forum posts as aggressive. This is likely to produce more posts correctly labeled and falsely labeled as aggressive. Automated text processing is still expected to produce useful results for a classifier which handles noisy features.

4. Roles in cyberbullying

In a face-to-face bullying situation, each individual in the situation takes on a role (Sutton & Smith, 1999). For example, one classification of roles is bully, victim, bully assistant, bully reinforcer, victim defenders, or outsider. A bully physically or mentally harms a victim through, for example verbal or physical attacks. A bully assistant actively participates in bullying (e.g. assists in holding down a victim). A bully reinforcer supports the bullying but does not directly attack a victim (e.g. encourages the bully to attack or laugh at a victim). An outsider does not participate in the bullying.

Similarly in cyberbullying, some individuals help the cyberbully to attack (e.g. follow instructions to send forum posts to attack a victim), some may encourage the cyberbully (e.g. encourage the bully or laugh at the victim through forum posts), some may help the victim to defend themselves (e.g. attack a cyberbully or encourage a victim to ignore cyberbullying through forum posts), and some individuals may observe and take no action (e.g. a neutral bystander who reads a forum). Roles in cyberbullying are not necessarily fixed and may change from situation to situation. For example, an individual may consistently be a bully (i.e. a stable bully), may sometimes be a victim or bully, or may cyberbully to retaliate in response to face-to-face bullying (Camodeca, Goossens, Schuenge, & Terwogt, 2003). Understanding the roles in cyberbullying may be beneficial as the specific roles and interactions between roles may provide hints for features of forum posts which can identify the existence of cyberbullying.

4.1. Method used to identify roles

In this paper, the authors manually analyze forum posts to identify roles. The owner of the forum page is considered a victim if there exists forum posts on their forum page which attack the owner of the forum page. A cyberbully and cyberbully assistant may send forum posts to target a victim. If a forum post is aggressive and targets a victim, the identifier associated to the forum post is labeled as an “attacker” which includes both cyberbully or cyberbully assistants. The authors did not find forum posts corresponding to individuals acting in the role of bully reinforcer. A bully reinforcer in this particular online forum may be rare as a result of the following. Firstly, posts are expected to be questions directed at the forum page owner and thus statements directed at a cyberbully may be out of place. Secondly, cyberbullies can be anonymous and it is more difficult to encourage someone who has no identifier. Thirdly, it is also difficult to textually “laugh” without explaining the context of the laugh. If a bully reinforcer explains the reason for a laugh, then they would likely be labeled as a bully assistant. A victim defender may send aggressive forum posts to target a cyberbully. Thus, if a forum post is aggressive and targets a cyberbully, the identifier associated to the forum post is labeled as a “defender”. An outsider includes individuals who view the cyberbullying but do not do send forum posts related to the attack or defend. Forum posts by outsiders were labeled as “neutral”. This does not include outsiders who passively view a forum without posting. In the case that a forum page does not include cyberbullying, all the posts are labeled as “neutral”. Based on these criteria labels for roles included in this data analysis are “vicitms”, “attackers”, “defenders”, and “neutral”. Other more detailed classification of roles may be possible; however, the authors focus on a simple classification criteria and leave definition and classification of other types of roles in cyberbullying as future work.

5. Anonymity of communications

Forum posts can be labeled with a feature of anonymity to describe whether a victim is unable to identify the individual posting a forum post.

A cyberbully may have psychosocial motivation for being anonymous. Anonymity may free the cyberbully "from normative and social constraints on their behavior" (Patchin & Hinduja, 2006) or allow the cyberbully to avoid social punishment through school or criminal law (Li, 2007). Thus, anonymity may result in a cyberbully being more "scathing, hurtful" (Li, 2007) or in "heightened aggression and inappropriate behavior" (Ybarra & Mitchell, 2004). Anonymity also provides the cyberbully with an imbalance of power which limits the ability of the victim to apply ordinary techniques to prevent aggressive behavior (David-Ferdon & Hertz, 2007).

In a face-to-face bullying situation, a defender may have psychosocial motivation for being identified as a defender. The defender may be satisfying socialized norms such as "social responsibility" (i.e. to help those in need) or "reciprocity" (i.e. to help those who help us) (Thornberg, 2007) or the defender may be "relating to others in a prosocial or cooperative manner" (Salmivalli, Lagerspetz, Bjorkqvist, Osterman, & Kaukiainen, 1996). In face-to-face bullying, defenders are often popular individuals (Salmivalli et al., 1996). If social norms related to face-to-face bullying also apply to cyberbullying, then defenders may be more likely to use identifiers. For example, defenders send forum posts with an identifier so that their identifier becomes associated with social responsibility, prosocial or cooperative manners, and behaviors of popular individuals.

5.1. Method used to identify anonymity

In this paper, the authors consider two methods to measure anonymity.

The first method to measure anonymity in this paper is to label a forum post as "anonymous" if the individual posting a forum post does not include their forum identifier with their forum post. The label of anonymity may provide some information about the psychosocial state of the author of a forum post. The individual posting the forum post makes a choice about whether or not to include an identifier, and as described in Section 5, the choice may be influenced by their psychosocial motivations. The label of anonymity may not match with true anonymity. A forum post with an identifier may still not be identifiable. If a victim does not know to whom the identifier is associated, then the individual posting the forum post is still anonymous even with an identifier. A forum post without an identifier may still be identifiable. The victim may be able to identify the individual posting the forum post based on specific content or writing style contained in the forum post. Although the label of anonymity does not necessarily match with true anonymity, the label may still be useful for a classifier which handles noisy data.

In the second method in this paper, the relative anonymity of an individual is measured as the number and strength of relationships the individual has with other individuals. The number of relationships an individual has is how many identifiable individuals post to the forum managed by the individual. The number of forum posts exchanged between a pair of individuals is a measure of the strength of the relationship between the pair of individuals. If an individual has a large number of strong relationships visible in the forum, then more information is likely to be known about the individual and the individual is relatively less anonymous. The ability of a victim to partially identify an attacker or characteristics of an attacker may be important. If the victim has information about the attacker, then the imbalance of power resulting from a cyberbully being anonymous may be reduced.

6. Numerical results

6.1. Automated labeling of aggression

The first research issue is to measure how automatically labeled aggressiveness correlates with attacks and defends.

Table 1 summarizes the ability of the automated technique to label attacks directed at a victim relative to non-attack forum posts. Forum posts on the forum pages were labeled for aggressiveness automatically and manually as described in Section 3. Out of a total of 5230 forum posts, the manual technique labeled 230 + 122 forum posts as attacks, 61 + 54 forum posts as defends, and 4552 + 208 forum posts as neutral. The automated technique labeled 230 + 4613 forum posts as neutral and 122 + 262 forum posts as aggressive. Table 1 also summarizes the ability of automated labeling of aggressive forum posts to identify defends of a victim relative to non-defense forum posts. Table 1 also summarizes the combined attacks and defends for automated and manual labeling of forum posts. Correlation coefficients are measured using Pearson's phi coefficient.

Automated labeling of forum posts as aggressive appears to have some use for identifying attacks and defends. Automated labeling identified a reasonable ratio of the attack forum posts (0.35 ratio) and defend forum posts (0.47). Automated labeling had a small level of correlation for identifying manually labeled attacks (0.28 correlation coefficient) and defends (0.23 correlation coefficient). The reasons that automated labeling may not have achieved a high correlation coefficient are as described in Section 3. Automated labeling does not appear to distinguish between attacks and defends, since both have a similar correlation with automated labeling.

Based on these observations, automated labeling appears to have some use as a tool for identifying forum posts which may be related to cyberbullying. Forum posts which may be related to cyberbullying are the combined set of attacks and defends. Automated labeling had a medium level of correlation (0.36 correlation coefficient) in identifying the combined set of forum posts versus neutral forum posts.

Automated labeling may also be useful for administration as a time-saving tool. A cyberbullying situation includes some number of forum posts. If the automated labeler can select at least one of the forum posts related to the situation, then the administration may be able to evaluate other related forum posts which are from the same cyberbullying situation. Of the 5230 total forum posts, automated labeling identified 4843 forum posts as neutral which corresponds to filtering a ratio of 0.94 of forum posts. At the same time, automated labeling kept a ratio of 0.37 of the 384 forum posts related to cyberbullying (i.e., combined set of attacks and defends). Although the ratio of forum posts incorrectly labeled can be high (ratio of 0.63), the number of neutral forum posts read by the administrator to identify potential cyberbullying can be much less.

Table 1
Correlation of manual labeling of attacks and defends versus automated labeling of aggression. Groupings for manual labeling include attacks against a victim versus all non-attack forum posts; defends of a victim versus all non-defend forum posts; and forum posts which may be related to cyberbullying (both attacks and defends) versus all neutral forum posts.

		Auto		Cor. coeff.
		Neutral	Aggress.	
Manual	Attack	230	122	0.28
	Non-attack	4613	262	
Manual	Defend	61	54	0.23
	Non-defend	4782	330	
Manual	Attack + defend	291	176	0.36
	Neutral	4552	208	

Table 2
Number of forum posts with identifiers and without identifiers (i.e., anonymous) for each type of automatically and manually labeled forum post.

Forum post type	Identifier	Anonymous	Chi-squared
Auto aggressive	72	312	7×10^{-17}
Manual attack	55	297	4×10^{-20}
Manual defend	62	53	0.0016
Manual neutral	1883	2877	–
All	2002	3228	.058

6.2. Correlation of anonymity with aggressiveness and role

The second research issue is to measure how automatically labeled anonymity correlates with aggressiveness and roles.

Table 2 summarizes the number of forum posts with identifiers and without identifiers (i.e., anonymous) for each type of manually and automatically labeled forum post. Significance of anonymity is relative to neutral forum posts. The manual technique distinguished attacks ("manual attack"), defends ("manual defend"), and neutral forum posts ("manual neutral"). The automated technique labels forum posts as aggressive ("auto aggress.") and does distinguish between attacks and defends. "All" is for all types of forum posts combined.

As seen in Table 2, both manually labeled attack posts and automatically labeled aggressive posts were significantly more anonymous relative to neutral forum posts. This may be caused by an attacker having psychosocial benefits for being anonymous as described in Section 5. Manually labeled defends were significantly less anonymous relative to neutral forum posts. This may be caused by the defender having psychosocial benefits for being identifiable as described in Section 5.

Fig. 1 illustrates the ratio of anonymous forum posts for manually and automatically labeled forum posts. The ratio of anonymous forum posts for each type of forum post is defined as the number of anonymous forum posts labeled to the type divided by the total forum posts labeled to the type.

Anonymity may be a useful feature for classifying an aggressive forum post as either an attack or a defend. Manually labeled attacks had a medium correlation with being anonymous (correlation coefficient 0.38) compared to manually labeled defends. Anonymity does not achieve high levels of correlation since around half of defends are still anonymous. If a forum post is labeled as aggressive and it is anonymous, then the forum post is more likely to be an attack. If a forum post is labeled as aggressive and it has an identifier, it is more likely to be a defend.

Another measure of the anonymity of an individual is the number and strength of relationships the individual has in the online forum. Note that this measure is per individual and not per forum post. The three types of individuals posting onto forum pages are attackers, defenders, and neutrals. The total numbers of attackers, defenders, and neutral individuals identified were 27, 30, and 868 respectively.

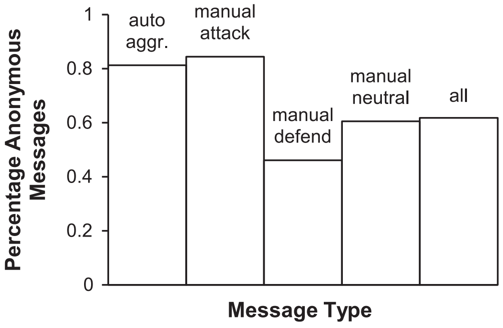

Fig. 1. Ratio of anonymous forum posts for various types of labeled forum posts.

Table 3
Number of relationships per forum page grouped by attacker, defender, or neutral individual.

Role	Relationships	St. dev.
Attacker	16	22
Defender	12	12
Neutral	20	36

Table 4
Number of forum posts per relationship grouped by forum page of an attacker, defender, or neutral individual.

Role	Posts per relationship	St. dev.
Attacker	5.7	11
Defender	8.3	13
Neutral	6.0	10

Table 3 summarizes how many relationships an attacker, defender, or neutral individual have. Table 4 summarizes the strength of relationships for each type of individual.

Defenders appear to have fewer, but stronger relationships relative to attackers and neutral individuals. However, the measure of relationships and forum posts per relationship did not significantly distinguish between attackers and defenders (chi-squared of 0.40 for relationships and chi-squared of 0.35 for forum posts per relationship). The low accuracy may result from the high variance in number of relationships and forum posts per relationship. The high variances may be typical variance in the number of relationships per user (i.e., some interact with few individuals and others interact with many individuals) as well as variance in the number of forum posts created by users (i.e., some use the forum rarely and other use the forum frequently).

6.3. Correlation between attacks and defends

The third research issue is to measure how roles correlate with each other.

A forum page contains some set of attacking forum posts and defending forum posts. Table 5 summarizes the number of manually labeled attacks and defends per forum page. For example, there were 21 forum pages with attacks, and there were a mean 17 (standard deviation of 19) attack forum posts on the each of the 21 forum pages.

The ratio of attack and defend forum posts per forum page varies widely. One reason for the variance may be the existence of cyberbullying on some forum pages but not on others. Another factor affecting the variance is the variance in the overall number of forum posts per forum page. As future work, it would be useful to understand what characteristics of a forum page or forum page owner may lead to more attacks on the forum page.

Fig. 2 illustrates the distribution of attacks and defends per forum page. Note that this measure is per forum page and not per forum post. In Fig. 2, each data point represents a forum page and there are a total of 26 forum pages. From Fig. 2, the number of

Table 5
Number of attacks and defends per forum page.

Post type	Pages with	Forum posts per page	
		Mean	St. dev.
Attack	21	17	19
Defend	13	9	10

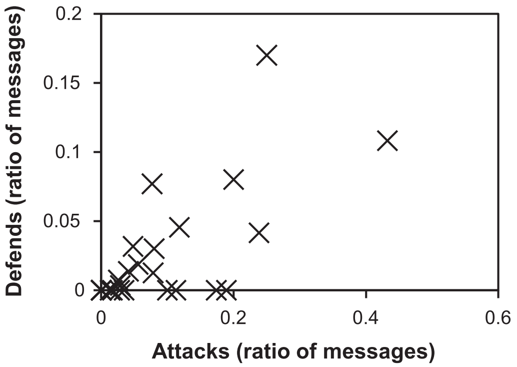

Fig. 2. Ratio of attack and defend forum posts for each labeled forum page.

forum posts by defenders on a forum page was generally less than the number of forum posts by attackers on the same forum page. This is expected since defender forum posts are in response to an attack forum post.

The ratio of attackers and defenders may be of interest. Defenders participating in a forum may reduce the imbalance of power of a cyberbully and may compensate for some of the negative impact of cyberbullying. Thus, forum pages with a low ratio of defends may be more at risk for negative impacts from cyberbullying.

7. Conclusions and future work

Cyberbullying is a growing concern as the usage of the internet is growing among youth. One step towards identifying cyberbullying in online forums is the identification of features in forum posts which may be related to cyberbullying.

In this paper, the authors applied an automated technique to label aggressiveness and anonymity forum posts. The authors also manually labeled the aggressiveness and role of the forum posts (i.e., attack, defend, neutral). The automated labels were correlated with the manual labels to determine the ability of the automated techniques to identify the aggressiveness and role of forum posts. Based on the online forum analyzed in this paper, several conclusions were made. Automated techniques to label aggressiveness had some success in identifying aggressive attacks. Automated techniques to label aggressiveness also had some success in identifying aggressive defends. Automated techniques to label anonymity had some success in distinguishing attack forum posts (relatively more anonymous) from defend forum posts (relatively less anonymous).

One future work is to identify the correlation of other features with cyberbullying. For example, the frequency of exchanging forum posts may correlate with being less susceptible to cyberbullying. This may be the case since individuals in face-to-face bullying with higher reciprocity (i.e., reflected as many forum posts between two individuals in the case of cyberbullying) are less likely to be victims (Mouttapa, Valente, Gallaher, Rohrbach, & Unger, 2004). Or for example, individuals who often receive attack forum posts but do not receive defend forum posts may correlate with negative consequences of cyberbullying. Another extension of the research is to consider how website designs, policies, and audiences impact the occurrence of cyberbullying.

Another future work is to quantify the level of anonymity of an individual and anonymity between a pair of individuals. One body of research has analyzed how anonymous shared social network data is (i.e., individuals and the network relationships among them). For example, in Sweeny (2002), *k*-anonymity is a measure of the degree of anonymity and is defined as whether an individual cannot be distinguished from at least *k* other individuals. If *k* is larger, then the shared data better prevents the mapping from shared data to real life individuals. Other examples include measuring anonymity based on network structure (Hay, Miklau, Jensen, Weis, & Srivastava, 2007; Singh & Zhan, 2007) or measuring anonymity based on social networks combined from multiple data sources (Narayanan & Shmatikov, 2009).

Another future work is to analyze social network features of online forums (e.g., measuring features such as centrality, clustering coefficients, and group cohesion) to improve labeling accuracy. Social network features may be preferred to text analysis since social network features can be measured without using private text contained within communications. It would be interesting to determine if other aspects of social interactions in face-to-face bullying appear in cyberbullying. For example, face-to-face bullying in a group increases the damage, perpetuates the abuse, and may satisfy a bully's desire for power and recognition (Dooley et al., 2009). Also for example, bullies have been correlated with social network features of "centrality", larger group size, and having aggressive friends (Mouttapa et al., 2004).

There exist a large number of techniques and directions which may apply towards automated tools to identify cyberbullying in online forums. It may be useful to develop many techniques since a different set of techniques may apply to each online forum being analyzed.

References

Baker, P. (2001). Moral panic and alternative identity construction in usenet. *Journal of Computer-Mediated Communication, 7*.

Camodeca, M., Goossens, F. A., Schuenge, C., & Terwogt, M. M. (2003). Links between social information processing in middle childhood and involvement in bullying. *Aggressive Behavior, 29*, 116–127.

David-Ferdon, C., & Hertz, M. F. (2007). Electronic media, violence, and adolescents: An emerging public health problem. *Journal of Adolescent Health, 41*, S1–S5.

Dooley, J. J., Pyzalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A theoretical and conceptual review. *Journal of Psychology, 217*, 182–188.

Hay, M., Miklau, G., Jensen, D., Weis, P., & Srivastava, S. (2007). Anonymizing Social Networks. Technical Report, No. 07-19, University of Massachusetts Amherst, Computer Science Department.

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research, 41*, 206–221.

Kowalski, R. M., & Limber, S. P. (2007). Electronic bullying among middle school students. *Journal of Adolescent Health, 41*, S22–S30.

Lewin, T. (2010). *Teenage insults, scrawled on web, not on walls*. New York Times. May 5.

Li, Q. (2007). New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior, 23*, 1777–1791.

Mouttapa, M., Valente, T., Gallaher, P., Rohrbach, L. A., & Unger, J. B. (2004). Social network predictors of bullying and victimization. *Adolescence, 39*, 315–335.

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. In *30th IEEE symposium on security and privacy* (pp. 173–187).

Olweus, D. (1990). *Health hazards in adolescence* (pp. 259–296). Walter De Gruyter Inc.

Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice, 4*, 148–169.

Salmivalli, C., Lagerspetz, K., Bjorkqvist, K., Osterman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior, 22*, 1–15.

Singh, L., & Zhan, J. (2007). Measuring topological anonymity in social networks. In *IEEE international conference on granular computing* (pp. 770–774).

Smith, P. K., Cowie, H., Olafsson, R. F., & Liefooghe, A. P. D. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen-country international comparison. *Child Development, 73*, 1119–1133.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, 49*, 376–385.

Sutton, J., & Smith, P. K. (1999). Bullying as a group process: An adaptation of the participant role approach. *Aggressive Behavior, 25*, 97–111.

Sweeny, L. (2002). *k*-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10*, 557–570.

Taki, M., Slee, P., Hymel, S., Pepler, D., Sim, H., & Swearer, S. (2008). A new definition and scales for indirect aggression in schools: Results from the longitudinal comparative survey among five countries. *International Journal of Violence and School*.

Thornberg, R. (2007). A classmate in distress: schoolchildren as bystanders and their reasons for how they act. *Social Psychology of Education, 10*, 5–28.

Ybarra, M. L., & Mitchell, K. J. (2004). Youth engaging in online harassment: associations with caregiverchild relationships, internet use, and personal characteristics. *Journal of Adolescence, 27*, 319–336.

Ybarra, M. L., Mitchell, K. J., Wolak, J., & Finkelhor, D. (2006). Examining characteristics and associated distress related to internet harassment: Findings from the second youth internet safety survey. *Pediatrics, 118*, e1169–e1177.

Yin, D., Davison, B. D., Xue, Z., Hong, L., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on web 2.0. In *Proceedings of the content analysis in the Web 2.0*.