Journal of Computational Social Science (2018) 1:241–260
https://doi.org/10.1007/s42001-018-0026-9

RESEARCH ARTICLE

On cyberbullying incidents and underlying online social relationships

Qianjia Huang[1] · Vivek K. Singh[2] · Pradeep K. Atrey[3]

Received: 21 March 2018 / Accepted: 6 August 2018 / Published online: 14 August 2018

Abstract

Cyberbullying is an important social challenge that takes place over a technical substrate. Thus, it has attracted research interest across both computational and social science research communities. While the social science studies conducted via careful participant selection have shown the effect of personality, social relationships, and psychological factors on cyberbullying, they are often limited in scale due to manual survey or ethnographic study components. Computational approaches on the other hand have defined multiple automated approaches for detecting cyberbullying at scale, and have largely focused only on the textual content of the messages exchanged. There are no existing efforts aimed at testing, validating, and potentially refining the findings from traditional bullying literature as obtained via surveys and ethnographic studies at scale over online environments. By analyzing the social relationship graph between users in an online social network and deriving features such as out-degree centrality and the number of common friends, we find that multiple social characteristics are statistically different between the cyberbullying and non-bullying groups, thus supporting many, but not all, of the results found in previous survey-based bullying studies. The results pave way for better understanding of the cyberbullying phenomena at scale.

Keywords Cyberbullying · Social networks · Bullying

✉ Vivek K. Singh
vivek.k.singh@rutgers.edu

Qianjia Huang
qhuan035@uottawa.ca

Pradeep K. Atrey
patrey@albany.edu

1 School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada

2 School of Communication and Information, Rutgers, The State University of New Jersey and Media Lab, Massachusetts Institute of Technology, Cambridge, USA

3 Department of Computer Science, University at Albany, State University of New York, Albany, USA

Introduction

Cyberbullying, which may be defined as “an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself” [1] has emerged as a serious social problem in the past few years. With the widespread use of the online social networks, cyberbullying has become increasingly more common and more harmful for both the victims and the bullies [2]. A recent research report by the Cyberbullying Research Center found that about a third of teenagers have experienced cyberbullying at some point of time and about 20% were bullied within the last 30 days [3]. This is especially worrying as multiple studies have reported that the victims of cyberbullying often deal with psychosomatic disorders [4], and a study in Britain has reported that at least half of suicides among teenagers are related to bullying [5].

Unlike traditional bullying, cyberbullying can happen anywhere and any time online, which makes cyberbullying often more harmful than traditional bullying. According to some studies [6], people who bully others on the Internet are less likely to feel regret, shame, guilt, or even fear of retaliation because they do not see the emotional reactions of their “targets”. At the same time, cyber victims may not recognize their experiences as bullying and they may not report it or seek help for the emotional difficulties. The reports such as [7] and [8] show that around 90% of young cyber victims did not tell others about their experiences. However, psychology scholars have reported that cyberbullying has a number of serious consequences for the victims, including lower self-esteem, increased suicidal thoughts, and a variety of emotional responses: retaliating, or being scared, frustrated, angry, and depressed (e.g., [8, 9]).

Based on Danah Boyd’s view, cyberbullying mirrors traditional bullying. For most online users, the internet mirrors the dynamics that take place offline. Troubled kids offline are often troubled kids online. According to related research [10], most victims and bullies online are victims and bullies in the real world. Further, most common perpetrators of bullying are “known” individuals in both online and offline worlds [11]. This implies that understanding traditional bullying could be useful when analyzing cyberbullying.

In the past, multiple research efforts have looked at cyberbullying from varying perspectives. While experts in social science have studied the phenomenon by analyzing the personality, social relationships, and demographic factors involving the bully and the victim, computer science researchers have developed automated methods to identify cyberbullying messages. However, the accuracy of these (typically textual feature-based) methods remains limited. We posit that, while cyberbullying happens over a technical substrate, at its core it is a social problem. Hence, understanding the social context surrounding cyberbullying messages is as important as (if not more important) analyzing the content of those messages.

Some recent studies (including our earlier work) have highlighted the value of using social features to improve cyberbullying detection performance [12, 13]. However, there are as yet, very few attempts to dig deeper and understand the phenomena of cyberbullying using computational methods. Such an understanding-based (rather than solely prediction-based) approach has been identified as one of the cornerstones of computational social science research [14]. Hence, this research investigates whether the findings in bullying-related social media research hold true when tested at scale using online social media content. This work makes a methodological and empirical contribution at the intersection of computational and social literature to advance the understanding of cyberbullying.

In this paper, we survey the relevant social science literature on bullying and define a number of hypotheses connecting social features (e.g., number of common friends, number of exchanged messages) and cyberbullying. We test these hypotheses by building relationship graphs of users sampled from the Twitter CAW 2.0 corpus (Content Analysis on the Web 2.0[1]). With human labeling, the considered dataset includes 2150 receiver–sender pairs with 4865 messages exchanged between them and the social properties computed based on localized social relationship graphs involving 176,869 users and a total of 247,215 messages. Our analysis found most, but not all, of the identified hypotheses to hold true over the very different user settings.

The organization of the rest of this paper is as follows. The next section surveys the related work, followed by which a set of hypotheses based on prior bullying literature that needs to be tested in online social settings is motivated and defined. The subsequent section describes the data set and the methods used to characterize the social relationships, paving way for hypothesis testing described later. The results of hypothesis testing lead to the discussion. The paper ends with the conclusions and suggestions for future work.

Related work

Academic work to tackle cyberbullying can be viewed from two perspectives: social science studies from survey-based analysis of demographic and social factors associated with cyberbullying, and computer science (mostly text mining) efforts at automatic detection of cyberbullying.

Multiple social science studies have analyzed the phenomenon of cyberbullying and identified factors associated with it. Cyberbullying comes in multiple variants: Gossip, Exclusion, Impersonation, Harassment, Cyberstalking, Flaming, Outing, and Trickery [15] and [1]. Different studies have reported different associations with age and gender. According to [16], "Although cyberbullying arises among all age groups in varying degrees, a large majority of the research is targeted at children and teens". Tokunaga points out that in his meta-analysis of over 60 studies, all but one article [15] exclusively investigated cyberbullying victimization among minors under the age of 18. In the same study, Tokunaga states that there is no conclusive evidence on the role of gender in cyberbullying.

[1] http://caw2.barcelonamedia.org/.

Multiple studies have also found social factors to be associated with cyberbullying. For example, the victims tend to be less popular than others and have fewer friends [17], are perceived as different from their peers, such as being overweight or underweight [18], wearing glasses or different clothing, being new to a school, or being unable to afford what kids consider 'cool', are perceived as weak or unable to defend themselves [19, 20], anxious [21, 22], have (or develop) lower self-esteem [9, 17], or antagonize others for attention. Those most likely to bully, on the other hand, tend to be well-connected to their peers, have social power, are overly concerned about their popularity, and like to dominate or be in charge of others. Other bullies are more isolated from their peers and may be depressed or anxious, have low self-esteem [9], be less involved in school [23], be easily pressured by peers, may not identify with the emotions or feelings of others, are aggressive or easily frustrated [24].

It is also common for many of those involved in bullying to also be victims who are trying to 'fight back' in the real world as well as on the Internet [15, 25]. As it is easier for people to get the 'online power' from the internet, many people are both cyber bullies and cyber victims [4].

Research efforts on automatic cyberbullying detection have mostly used text-based methods and employed user profile and sentiment features to improve the text mining system. For instance, [26] used the number, density and value of foul words as features to determine the cyberbullying messages. Similarly, Dinakar et al. found that building individual topic-sensitive classifiers and common sense reasoning help to improve the detection of cyberbullying messages [27, 28]. Recently, [29] also presented an improved model using the user-based features, i.e., history of the user's activities and demographic features. All these works are based on text mining. However, the accuracy of text-based cyberbullying detection methods is typically limited. For preprocessing the text content, other researchers have used semantic methods to get around variations in spelling and use of emoticons [30]. For instance, [31] developed a Semantic-Enhanced Marginalized Denoising Auto-Encoder (seSDA) which is used for denoising the autoencoder.

There have also been some recent efforts aiming at looking beyond text features for cyberbullying detection. Nahar et al. built a cyberbullying network graph with the users who had been previously labeled as cyber bullies and victims, then used a ranking method to identify the most active cyber bullies and victims [32, 33]). Hosseinmardi et al. [34, 35] collected data from Instagram with snowball sampling method and complemented textual information (LIWC2015) with human-labeled image content tags for an enhanced cyberbullying detection model. Hosseinmardi et al. [34] also showed that cyberbullying detection cannot only rely on the usage of profanity because more than 60% of the media session with profanity word are not identified as cyberbullying/cyberaggression.

Similarly, Huang et al., Squicciarini et al., and Hosseinmardi et al. have suggested using a combination of content and network features for better cyberbullying detection [12, 13, 36]. Huang et al. have found the combination of textual and social network features to be better at cyberbullying detection than a purely textual approach [12]. Squicciarini et al. utilize both text and social network features to undertake two tasks—detecting cyberbullies in online social networks and also identifying the pairwise interactions between users through which the influence of bullies spreads [13]. Hosseinmardi et al. use a combination of text, image, and network graph features to build "early predictors" for cyberbullying, i.e., identify image-based Instagram posts which are likely to become victims of cyberbullying in future [36].

Social media platforms often have huge amount of posts which require significant manual labeling effort. Even using online platforms with multiple people (i.e., Amazon Mechanical Turk), getting human annotations for a large corpus is often prohibitively expensive and time consuming. Hence, scholars have considered the use of semi-supervised deep learning methods. For instance, Wulczyn, Thain, and Dixon [37] in a work on online attack from Wikipedia blocked list, first labeled each comment for personal attack for a small fraction of the corpus, and then trained the labeled comments with a machine learning classifier. Chu, Jue, and Wang [38] tested three deep learning models: a recurrent neural network (RNN) with a long–short-term memory cell (LSTM) and word embeddings, a convolutional neural network (CNN) with word embeddings and character embeddings, and after testing with the Wikipedia corpus, they found that CNN with character-level embeddings has the best performance.

Other than detection, many of the recent efforts have also focused on *analysis* of cyberbullying content. This implies that they were not trying to automatically detect cyberbullying, but rather were interested in analyzing and reporting the trends associated with cyberbullying. For example, [39] compared the prevalence of cyberbullying in Twitter and in Weibo. They reported that in general, Weibo posts contained less bullying than Twitter posts. They also reported that the posts from Weibo contain more mentions of family than those from Twitter. The authors posit that this may be due to the greater emphasis on family in Asian cultures. In another effort, Cortis et al. [40] analyzed twitter data to identify the most popular hashtags and named entities used within cyberbullying tweets. Specifically, they focused on two important real-world events (Ebola virus outbreak in Africa and the shooting of Michael Brown in Ferguson, Missouri) and analyzed the associated discussion to identify how often cyberbullying took place in the associated discussion and what were the main keywords used.

In this research, we focus on understanding cyberbullying phenomena by testing hypotheses that have been identified based on social science literature but tested at scale using computational methods. A brief comparison of the proposed approach with existing works is presented in Table 1. Note that this research does not focus on demographic features. This work is based on a conscious decision to work only with anonymized data sets and mitigating some of the ethical dilemmas associated with connecting immutable demographic characteristics with cyberbullying.

This paper builds upon and significantly expands a previous publication [12] which was the first paper to propose the use of social features for cyberbullying detection. Since then, two other efforts have also used social network features for cyberbullying detection [13, 36]. This further validates this line of work. The current manuscript uses the previous publication [12] as the starting point but has a very different focus. While the previous paper focused on cyberbullying detection, this manuscript provides an in-depth analysis of the social science literature and

Table 1 A comparison of the proposed approach with the existing works on cyberbullying detection

The work	Hypothesis testing	Method	User demography	Social network features
Smith et al. [1]	Yes	Surveys	No	No
Kowalski and Limber [7]	Yes	Surveys	No	No
Patchin and Hinduja [9]	Yes	Surveys	No	No
Mishna et al. [25]	Yes	Surveys	No	No
Campfield [41]	Yes	Surveys	No	No
Navarro et al. [22]	Yes	Surveys	No	No
Festl and Quandt [17]	Yes	Surveys	No	No
Reynolds et al. [26]	No	Text mining	No	No
Dinakar et al. [27]	No	Text mining	No	No
Dadvar et al. [29]	No	Text mining	Yes	No
Nahar et al. [33]	No	Text mining	No	No
Hosseinmardi et al. [34, 35]	No	Text + image analysis	No	No
Squicciarini et al. [13]	No	Image + network analysis	No	Yes
Huang et al. [12]	No	Text + network analysis	No	Yes
Proposed approach	Yes	Social network analysis	No	Yes

tests multiple hypotheses. The aspect of hypothesis testing which is central to this work was not present in the previous paper.

Cyberbullying and users' online relationships

There exists significant social science research (e.g., [17]) that suggests that structural features of the social network surrounding user messages may provide vital clues to cyberbullying. As we mentioned before, a limitation of exclusively focusing on the textual features for studying cyberbullying is that one would miss the cyberbullying messages which have no direct textual evidence to suggest cyberbullying. In Fig. 1, the message 'Stop farting on people!' from sender(A) to sender(B) is an example which could not be detected as a bullying message using the textual feature model [12]. In this work, we also look at the characteristics of online relationship graphs. By comparing the structures of relationship graphs of communication that have been observed and labeled as cyberbullying and non-bullying, we analyze the connection between cyberbullying and online relationship graphs. For example, the number of nodes in this graph was 43, which is much closer to the average for the cyberbullying group (43.35) than the non-bullying group (59.68) [12].

Building on the previous work of social science, we have defined multiple hypotheses about the associations between cyberbullying and the social context surrounding it. Past studies have found that cyberbullying is more likely to occur in specific peer configurations [17, 42] and reported positive associations between network centrality and forms of social aggression [43].

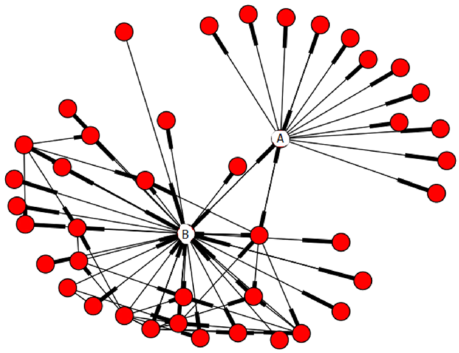

Fig. 1 Example of a social relationship graph between two users. The message "stop farting on people" exchanged between these users was missed by automated text-based bullying detectors, but the surrounding social structure (e.g., low number of nodes in the supporting network) provided vital clues for correct detection. Here, node A is the bully and node B is the victim. All nodes are anonymized

However, the social network created in all these studies was based on questionnaires (e.g., 'Who are your friends in school?'). Multiple studies have shown that self-reported surveys, while useful, are costly, retrospective, piecemeal, and may suffer from cognitive and observational biases [44–46]. While "forgetting" remains the most common explanation for the discrepancy between self-reported and logged data, it is by no means the only explanation. Multiple studies have reported different biases including attractiveness bias, sociability bias, and expansiveness bias as possible explanations and hence also considered different factors such as network size, salience, behavioral specificity and tie strength to interpret the differences found [47]. Using objective logged data such as the Twitter data considered here could be useful to counter for many of such biases. This logged dataset does not have to deal with human forgetfulness, nor many of the abovementioned cognitive biases [48]. Hence, the social structure derived from such data is likely more suitable to study phenomena like cyberbullying. Lastly, while the manual survey approach could typically be applied to only tens or hundreds of participants at a time, an online network approach can scale gracefully to millions of users if not more.

Hence, we propose to analyze multiple hypotheses using passive (objective) longitudinal data coming from a large number of users. For this, we conduct a comprehensive study of the literature and develop multiple relevant hypotheses.

In this paper, we divide the collected data and considered relationships into two groups: cyberbullying group and non-bullying group. For cyberbullying group, we refer to the message sender and receiver as 'cyber bully' and 'cyber victim'. For non-bullying group, we call the users from conversation as 'non-bullying sender' and 'non-bullying receiver'.

The first social network feature that we consider is the size of the relationship graph. As Fig. 1 shows, we define the relationship graph of two users by combining the users' ego network graphs.

Prior research has pointed out associations between the social status of the actors (both bullies and victims) and their propensity to be involved in cyberbullying. While Brighi et al. (2012) and Festl and Quandt (2013) have pointed out that the victims tend to have a smaller number of friends, there are mixed reports on the number of friends for the bullies [17, 49]. Looking at it as a *relationship* rather than the two individuals, prior research on romantic relationships and domestic violence (e.g., [50, 51]) suggests that the surrounding social support structure might play an important role in the success of a relationship. Considering these, we hypothesize that relationships with less social support may be more prone to cyberbullying.

H.1 Compared to the non-bullying group, the relationship graphs of the cyberbullying group will have lesser nodes (total).

Tong [52] has showed that in a proper interval number, more friends (opportunities for social interactions) lead to more sociometric popularity in the online social world, and it is also clear that social interactions on networks affect user's activity [53]. From a psychological perspective, Campfield [41] found that people involved in cyberbullying (cyber bullies, victims, and bully/victims) have lower positive social interactions. In social media, social interactions could be represented by the number of online connections since social interactions are effected by the number of peers (opportunities for social interaction). Hence, we hypothesize that:

H.2 Compared to the non-bullying group, the individuals in the bullying group (cyber bully and cyber victim) will have lesser connections (individual).

Further, H.2 can be operationalized into two specific parts: *H.2(a) The senders in the bullying group (i.e., bullies) will have lesser social connections than senders in the non-bullying group.* and *H.2(b) The receivers in the bullying group (i.e., victims) will have lesser social connections than receivers in the non-bullying group.*

Next network feature that we use to compare the cyberbullying and the non-bullying group is the relationship strength. Prior research efforts [11, 54–56] have reported that users are often cyberbullied by someone whom they know in real life but do not consider as close friends. While it has been shown that users are more likely to know each other in real life if they have higher tie strength in online social networks [57] it has also been found that 'close friends' exchange lesser number of messages via online networks [58]. Hence, we expect bullying to occur when individuals know each other in the real world but are not close friends, yielding the following hypotheses:

H.3 In the relationship graphs, the cyberbullying group will have a higher online relationship tie score than the non-bullying group (know each other in real world).

H.4 In the relationship graphs, the cyberbullying group will have more exchanged messages (not close friend) than the non-bullying group.

Note that the relationship tie score in H.3 refers to the ratio of common friends between the two users to the total number of friends that either of them has. (More formally defined in Sect. 4.2.)

Another social network feature that we take into account is the online social activity of users. In the early 1980s, Olweus suggested that externalized behaviors such as argumentativeness and disruptiveness as well as internalized symptoms (e.g., loneliness, low self-esteem, and emotional problems) may invite or reinforce peer victimization [59]. Campfield [41] found that cyber victims have emotional and behavioral difficulties as well as loneliness, which may be similar to the vulnerability factors that make youth easy targets for face to face settings. Furthermore, Morahan-Martin and Schumacher [60] found that loneliness has been associated with increased online activity. It is also known that a user with a higher out-degree centrality posts to more users (higher user-initiated social activity) than those who have lower out-degree centrality [61]. Therefore, we hypothesize that:

H.5 The social activity (out-degree centrality) of cyber victims is higher than that of those in other roles (cyber bullies, non-bullying senders, and non-bullying receivers).

Data set and methods

The description of the dataset of this research and how we use it to build the users' relationship graphs are as follows.

Data set

An important pre-requisite for understanding cyberbullying is the creation of a large, reliable data set. Creating such a data set in itself is a challenge because of the subjectivity involved in creating such labels, which are compounded by the use of 'internet slang', limited character lengths, and the size imbalance between positive (cyberbullying) and negative (non-bullying) samples. Existing public data sets do not include social features (e.g., number of common friends, edge centrality) surrounding the message exchange and thus are not appropriate for understanding the social context and interpersonal relationships associated with cyberbullying.

Twitter, one of the most-visited internet websites, has over 500 million registered users. Users can post with @ to mentioning or replying to other users. In this work, we considered the @ sign as the direct connections which can be used to create a relationship graph.

To study the issue of cyberbullying, we focus on a randomized sample of Twitter post from the CAW 2.0 dataset. We chose the CAW 2.0 because it has been widely used in previous literature for similar tasks [26] and it provides information on both the textual content and the social network. Other data sets in CAW 2.0, including those containing data from forums such as Slashdot or Youtube were rejected because there is no easy way to identify social network structure information in them. For the current analysis, we randomly selected 800 files and kept the comments posted with the @ symbol, which represents the direct paths between two users. This resulted in a data set of approximately 13,000 messages. In this paper, we have two steps for processing these messages so that they can be utilized for building the online relationship graph. On the one hand, we asked three students to label each message two times on whether the message could be thought of as a cyberbullying incident (or not). Each post was labeled as 'yes', 'no', and 'not sure' regarding the incidence of cyberbullying. The labelers disagreed or were undecided for 36 messages and these messages were removed from the dataset. This process led to 257 out of 13,000 messages being marked as bullying messages (around 2 percent of the total). Furthermore, since the social structure information could be erroneously effected by the incomplete graph if we could not get the interaction record from the receiver, we decided to keep only those messages for which both the sender and the receiver interaction histories exist in the CAW 2.0 database. This finally yielded a data set with 2148 pairs of users (relationship graphs similar to

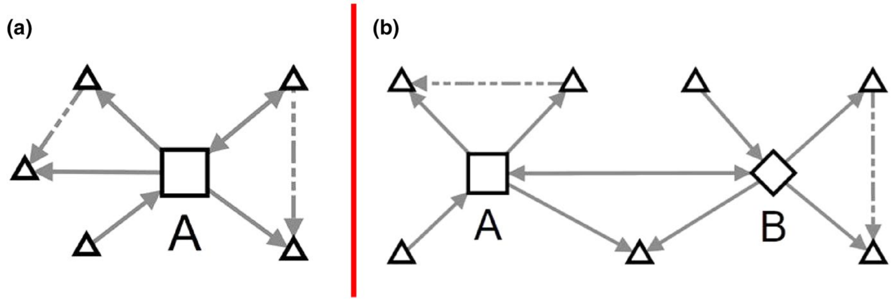

Fig. 2 **a** A 1.5 ego network, **b** a relationship graph defined by combining the 1.5 ego network graphs of the sender and the receiver.eps

Fig. 1) and 4865 messages between them (91 of them were labeled as bullying messages). Our presented results focus on this set of 4865 messages.

Characterizing social structure surrounding bullying messages

To understand the impact of social structure on the incidence of cyberbullying, we constructed the social network graph(s) and derived a set of features. The social features were derived from using the 1.5 ego networks [62]. Please note that the 'ego' refers to an individual focal node in the social network. A network could have as many nodes as it has egos.

In this paper, we denote the 'global social network' as a graph $G =< V, E >$ where V is the set of all nodes and E is the set of directed edges over those nodes. The 1.0 ego network of a node v as the graph $G_1(V_1, E_1)$ such that V_1 contains all of the nodes u, and there exists an edge (v, u) in E, and that E_1 contains all the edges from v to the nodes of V_1. Relatively, we denote by the 1.5 ego network the graph $G_{1.5}(V_{1.5}, E_{1.5})$ such that $V_{1.5}$ has the same nodes with V_1, but $E_{1.5}$ consists of E_1 plus all direct links between the nearest neighbors of v.

In Fig. 2(a), the ego node A is marked as a square, while the neighbors are marked as triangles. The edges of the 1.0-ego network, i.e., E_1 are shown using solid lines while the edges in $\Delta_{1.5}$ are shown via dashed lines. In this work, we focus on building 1.5 ego networks as they capture a reasonable level of social context (me, my friends, and the relationships between them) while still keeping the data requirements and computational complexity low. This is in congruence with the human processing limits and the social brain hypothesis proposed by [63] and also the recent results that have shown the value of 1.5 ego networks in identifying network phenomena [62].

As shown in Fig. 2(b), we combine the 1.5 ego networks of the two users, the sender(A) and the receiver(B) to determine their online relationship graph. This relationship graph between sender(A) and receiver(B) contains all one-path neighbors of either the sender(A) or the receiver(B), and all the edges between these nodes. We also use directed, weighted edges to represent all communication information for relationship graphs. By analyzing these graphs, we could characterize the position

Table 2 Results of H.1

	Number of nodes in relationship graph	
	Cyber bully	Non-bully
Median	37	44
Mean	43.345	59.709
p value	0.015 (not significant at the corrected 0.0083 level)	
Hypothesis	Reject	

of both the sender(A) and the receiver(B) in their respective 1.5 ego networks. To take an instance, it allows us to identify which users are more 'popular' or 'central' to their social network(s) and which users experience little to no interaction with their peers.

Hypothesis testing

We analyzed 4865 messages and 2148 pairs in conversations. To ignore the duplicates of social network graph information (e.g., if sender(A) sent four messages to receiver(B), the graph information about the four messages would be absolutely the same, which would disturb the final data), we kept only one case of graph information for each connection in the data set. Because we do not know whether the data follow a normal distribution, for H.1 through H.4, we used Mann–Whitney U test to test whether there are differences between cyberbullying relationship groups and normal relationship groups. For H.5, we performed Chi-square test to compare the activity from senders and receivers of each message. Since we used the same data set for testing the six hypotheses (including H.2(a) and H.2(b)), we used the Bonferroni correction which asks for the significance level of $\alpha / n = 0.05 \ / \ 6 = 0.0083$.

For clear and convenient reading, we explain the symbols which we use to describe the relationship graph beforehand; in the relationship graph G, for the sender(A) and receiver(B), all the one-path neighbors that they have are represented by Γ_A and Γ_B, respectively. The common neighbors for both sender(A) and sender(B) can be expressed as

$$CN_{A,B} = \Gamma_A \cap \Gamma_B. \tag{1}$$

We identify the strength of relationship tie between sender(A) and receiver(B) by the Jaccard Index [64]:

$$\mathrm{RT}_{A,B} = \frac{\Gamma_A \cap \Gamma_B}{\Gamma_A \cup \Gamma_B}. \tag{2}$$

H.1: Size of relationship graph and online social interaction

Upon analyzing the data, bullying relationship cases were **not** found to have statistically significantly lesser nodes in the relationship graph. In our data set, we had

Table 3 Results of H.2

	Individual neighborhood (nodes)	
	Cyber bully	Sender (normal group)
Median	12	17
Mean	18.51	24.10
p value	0.005	
Hypothesis	Accept	
	Cyber victim	Receiver (normal group)
Median	16	23
Mean	28.65	38.88
p value	0.045 (not significant at the corrected 0.0083 level)	
Hypothesis	Reject	

55 bullying relationship cases and 2093 normal relationship cases. The detailed numerical results are shown in Table 2. We found that there is a 0.05 level difference (0.015) between the cyberbullying group and the non-bullying group in the number of nodes in the relationship graph, but it is not significant at the corrected 0.0083 level.

H.2: Number of connections for individuals

For H.2(a), we found that there is a significant statistical difference between the cyber bullies from cyberbullying group and the senders from non-bullying group with regard to the number of neighboring connections (nodes). At the same time, we found that there was no significant difference between cyber victims and the receivers in the non-bullying group. From Table 3, the mean number of nodes for cyber bullies was 18.51, and the senders in non-bullying group had an average of 24.10 nodes in their relationship graphs. The *p* value of Mann–Whitney U test is 0.005, which suggests that the cyber bullies in cyberbullying group have significantly less nodes (connections) than the senders from non-bullying group. Similar to H.1, the difference between cyber victims in cyberbullying group and receivers in non-bullying group was not significant after Bonferroni Correction.

We can draw the conclusion that the cyber bullies in cyberbullying group have a smaller set of connections than the senders in non-bullying group (H.2(a)), but there is no significant evidence for the difference in either the overall size of the relationship graph (H.1) or the number of receiver’s connections (H.2(b)) between the cyberbullying group and the non-bullying group.

Table 4 Results of H.3

	Relationship tie	
	Cyber bully	Non-bully
Median	0.024	0
Mean	0.073	0.036
p value	0.007	
Hypothesis	Accept	

Table 5 Results of H.4

	Exchanged posts	
	Cyber bully	Non-bully
Median	5	2
Mean	8.255	2.985
p value	< 0.001	
Hypothesis	Accept	

H.3: Online relationship tie and offline relationship

Looking at the online relationship tie score, the cyberbullying group has a higher score than the non-bullying group in the relationship graph. For the average data in Table 4, the cyberbullying conversation (0.073) has a higher average score in the relationship tie than the non-bullying conversation (0.036). As Equation (2) which calculates the online relationship tie shows, it is clear that the higher score of the online relationship tie represents the higher frequency of common neighbors between the users in their social network graph. Upon considering the *Mann–Whitney U test* result ($p = 0.007$), it was found that the difference between the cyberbullying group and the non-bullying group is indeed significant.

H.4: Exchanged posts and not close friends offline

The Mann–Whitney *U* test result ($p < 0.001$) showed that the number of exchanged posts from conversation in the cyberbullying group was significantly higher than conversations in the non-bullying group. The detailed numerical results are shown in Table 5. The average number of exchanged posts between user in the cyber bully group was 8.255 and the non-bullying group was 2.985.

H.5: Online activity and loneliness

Comparing the social activity level of users from each message in our data set, the cyberbullying group was found to be different from the non-bullying group. To find the difference between the levels of out-degree centrality of the cyberbullying messages and non-bullying messages, we compared the out-degree centrality between the sender (cyber bully) and the receiver (cyber victim) for each message. We found

Table 6 Chi-square test for H.5

	Cyber bully	Non-bully	Total
$A > B$	29 (31.9%)	2471 (51.8%)	2500 (51.4%)
$A < B$	62 (68.1%)	2303 (48.2%)	2365 (48.6%)
Total	91	4774	4865
Hypothesis	Accept		

Chi-square = 14.1438, $p = 0.000169$

that the cyber victims have the highest out-degree centrality value (0.426) on average compared with the cyber bully (0.376), the non-bullying sender (0.408), and the non-bullying receiver (0.373).

Next, the Chi-square analysis (refer Table 6) shows that 29 (31.9%) cyberbullying messages belong to the "$A > B$", i.e., sender's out-degree centrality ≥ receiver's out-degree centrality group. By comparison, 2471 (51.8%) messages belonged to the "$A > B$" group for non-bullying messages; 62 (68.1%) cyberbullying messages belonged to the "$A < B$", i.e., sender's out-degree centrality < receiver's out-degree centrality group; and 2303 (48.2%) messages belonged to "$A < B$" for non-bully messages. In terms of the proportion of four groups, the Chi-square score of H.3 is 14.1438 and the p value is 0.000169. This indicates that cyberbullying messages are often received by the cyber victims who have higher out-degree centrality than the senders.

Discussion

On the whole, the results of hypothesis testing (a summary of results is shown in Table 7) open a door to understanding cyberbullying with automatically derived social network features which can help understand and interpret the related social science literature.

The results of H.1 were unexpected. While it was hypothesized that the relationships which involve cyberbullying will have lesser number of nodes, it was not found to hold based on the data considered. This is in contrast to the existing literature in social sciences which suggests that lower social support results in more conflicts in different social settings [50, 51]. This could be attributed to multiple factors. First, we note that much of existing social literature on bullying employs self-reported surveys to create social network graphs and thus typically involves much smaller N (of the order of 10s not 1000s). Further, the sample population studied is quite different in this work compared to most existing social science studies on cyberbullying (students below 18 years versus anonymous online population). Lastly, there is a clear difference between communication patterns (and hence number of nodes in social graph) between online and physical networks. While a higher tie strength in online networks indicates a higher likelihood of knowing each other in real life [57] it has also been found that 'close friends' exchange lesser number of messages via online networks [58]. The results hence motivate more work at scale on understand the

Table 7 Overall results for hypothesis testing

	Hypothesis	Method	Results for testing
H.1	Relationship graphs for the cyberbullying group have less nodes than the non-bullying group (total)	Mann–Whitney U test	Reject
H.2	Cyber bully/victim of cyberbullying group have less neighbors than sender/receiver of non-bullying group (individual)	Mann–Whitney U test	(a) Accept; (b) reject
H.3	The cyberbullying group has higher relationship tie score than the non-bullying group	Mann–Whitney U test	Accept
H.4	The cyberbullying group has more exchanged messages than the non-bullying group	Mann–Whitney U test	Accept
H.5	The social activity (out-degree centrality) from cyber victims is higher than other roles (cyber bullies, non-bullying senders, and non-bullying receivers)	Chi-square	Accept

cyberbullying phenomena across different population samples, network positions, tie strengths, and platforms.

The results from H.2(a) show that cyber bullies in the cyberbullying group have lower social interaction than senders in the non-bullying group. This is consistent with existing literature (e.g., [41]), which suggests that people who are involved in cyberbullying (cyber bully, victim, and bully/victim) have lower global self-esteem and overall sense of self-worth than the non-bullying group, and their positive social interactions (online and offline) are limited. The result for H.2(b) were in the unexpected direction but consistent with the findings for H.1 above. While the victims were found to have lower number of connections, the difference was not significant on its own. Hence, instead of using this feature as a predictor for cyberbullying in its own right, perhaps it makes sense to interpret and utilize it in combination with the other nuanced features of the social network. Like H.1 above, this result also motivates more work at scale on understanding the cyberbullying phenomena across different population samples, network positions, tie strengths, and platforms. Conceptually, these aspects may be connected with self-esteem, social support, and social capital [65–67].

In testing H.3, we have found that users in cyberbullying group have higher proportion of common online friends (as Eq. (2) represents) than non-bullying group, which could be seen as these users have higher possibility to know each other offline than non-bullying group. This result aligns with the findings of previous works [54, 55, 55, 55, 56] which state that people involved in cyberbullying typically know each other in real life.

A significant difference in the number of exchanged posts was also found for H.4. The cyberbullying group contains more exchanged posts than non-bullying group, this result indicates that users in cyberbullying group have less closer relationship in real world. The Chi-square test result for H.5 suggests that in conversations involving cyberbullying, victims often are more active than bullies. The reasons for this result could be found in the psychological literature. According to [68], the lonely individuals are likely to have more online activity for emotional support and heightened satisfaction. In effect, a user who has a higher out-degree centrality in the relationship graph of conversations is more likely to be bullied.

This research advances the state of the art in understanding cyberbullying beyond textual analysis to consider the social relationships in which these bullying messages are exchanged. We also found a large (but not perfect) alignment between previous survey-based cyberbullying studies and the current online study. Interestingly, this was based on a random sample of users found in an online corpus and hence did not necessarily pertain to teenagers and student, which have been the focus of most social science research on the topic. In fact, Tokunaga [16] identifies such a focus on a specific demographic segment as a peculiarity of existing cyberbullying research and hence moving beyond such a demographic focus might be worthwhile. Note that current social science literature either exists in non-bullying contexts or employs self-reported surveys to create social network graphs and thus typically involves much smaller N (of the order of 10s not 1000s). This motivates further similar studies that use large-scale online data to study cyberbullying.

While the current study scales up the analysis to order of N = 1000 s, it still does not scale to millions because of the requirement of human labeling of messages as cyberbullying. This remains a pertinent problem with all related cyberbullying work but might be addressed in future through accurate cyberbullying detection [12, 13, 36] or through the use of semi-supervised methods, which are able to synergize labeled and unlabeled data for creating bigger corpora [69]. The current analysis would also have a few limitations. For instance, as previously mentioned we do not analyze the demographic properties of the users. However, accessing and utilizing such information need to be traded-off with the compromise on anonymity, biases/errors in assigning demographic labels based on partial information, and the ethics of connecting cyberbullying with immutable demographic descriptors. Furthermore, the additional characteristics (e.g., time of post, geographic locations and the underlying context) could help in better understanding of such data in future.

The results also have implications in multiple ways: (1) they broaden the understanding of cyberbullying as a phenomena, especially on the social aspects associated with online bullying; (2) motivate the use of social features for better detection [12, 13] of cyberbullying; (3) lastly, based on the finding that cyberbullying is indeed affected by social relationships, we think that the same relationships can be used to intervene in cyber settings. In future, we plan to explore the social peer pressure, reflexive interface strategies, and delayed action methods [28], to make users aware of the social implications of their message and to investigate the effect of the such interventions on their decisions.

Conclusions

Cyberbullying is an important social challenge that takes place over a computational infrastructure. This work surveys social science-based literature on bullying to identify multiple hypotheses and tests them in a scalable manner using a computational approach. With human labeling, the considered dataset included 2150 receiver–sender pairs with 4865 messages exchanged between them and the social properties computed based on localized social relationship graphs involving 176,869 users and a total of 247,215 messages. The results point to a significant but not perfect alignment between expected hypotheses and the empirical evidence. The results expand the understanding of cyberbullying as a phenomena and motivate future work that uses social features for better detection as well as prevention of cyberbullying.

References

1. 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*(4), 376–385.

2. Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior*, *29*(2), 129–156.
3. Cyberbullying Research Center. (2015). Cyberbullying Data. https://cyberbullying.org/2015-data. Accessed 21 July 2018.
4. Sourander, A., Klomek, A. B., Ikonen, M., Lindroos, J., Luntamo, T., Koskelainen, M., et al. (2010). Psychosocial risk factors associated with cyberbullying among adolescents: A population-based study. *Archives of General Psychiatry*, *67*(7), 720–728.
5. BBC News. (2010). 'Bullying' link to child suicide rate, charity suggests. https://www.bbc.co.uk/news/10302550. Accessed 21 July 2018.
6. Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research*, *14*(3), 206–221.
7. Kowalski, R. M., & Limber, S. P. (2007). Electronic bullying among middle school students. *Journal of Adolescent Health*, *41*(6), S22–S30.
8. Mason, K. L. (2008). Cyberbullying: A preliminary assessment for school personnel. *Psychology in the Schools*, *45*(4), 323–348.
9. Patchin, J. W., & Hinduja, S. (2010). Cyberbullying and self-esteem. *Journal of School Health*, *80*(12), 614–621.
10. Boyd, D., Marwick, A., Aftab, P., & Koeltl, M. The conundrum of visibility: Youth safety and the internet.
11. Crick, N. R., & Nelson, D. A. (2002). Relational and physical victimization within friendships: Nobody told me there'd be friends like these. *Journal of Abnormal Child Psychology*, *30*(6), 599–607.
12. Huang, Q., Singh, V. K., & Atrey, P. K. (2014). Cyber bullying detection using social and textual analysis, in: Proceedings of the 3rd International Workshop on Socially-Aware Multimedia, SAM '14, ACM, New York, NY, USA, 2014, pp. 3–6. 10.1145/2661126.2661133. http://doi.acm.org/10.1145/2661126.2661133
13. Squicciarini, A., Rajtmajer, S., Liu, Y., & Griffin, C. (2015). Identification and characterization of cyberbullying dynamics in an online social network, in: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ACM, 2015, pp. 280–285.
14. Qiu, L., Chan, S. H. M., & Chan, D. (2018). Big data in social and psychological science: theoretical and methodological issues. *Journal of Computational Social Science*, *1*(1), 59–66.
15. Slonje, R., & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, *49*(2), 147–154.
16. Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, *26*(3), 277–287.
17. Festl, R., & Quandt, T. (2013). Social relations and cyberbullying: The influence of individual and structural attributes on victimization and perpetration via the internet. *Human Communication Research*, *39*(1), 101–126.
18. Wang, J., Iannotti, R. J., & Luk, J. W. (2010). Bullying victimization among underweight and overweight us youth: Differential associations for boys and girls. *Journal of Adolescent Health*, *47*(1), 99–101.
19. Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, *12*(4), 495–510.
20. Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review*, *31*(12), 1222–1228.
21. Campbell, M., Spears, B., Slee, P., Butler, D., & Kift, S. (2012). Victims perceptions of traditional and cyberbullying, and the psychosocial correlates of their victimisation. *Emotional and Behavioural Difficulties*, *17*(3–4), 389–401.
22. Navarro, R., Yubero, S., Larrañaga, E., & Martínez, V. (2012). Childrens cyberbullying victimization: associations with social anxiety and social competence in a spanish sample. *Child Indicators Research*, *5*(2), 281–295.
23. Minkkinen, J. (2013). Associations between school-related factors and depressive symptoms among children: a comparative study, Finland and Norway, School Psychology International. 0143034313511008.
24. Beran, T., & Li, Q. (2008). The relationship between bullying and school bullying. *The Journal of Student Wellbeing*, *1*(2), 16–33.

25. Mishna, F., Khoury-Kassabri, M., Gadalla, T., & Daciuk, J. (2012). Risk factors for involvement in cyber bullying: Victims, bullies and bully-victims. *Children and Youth Services Review*, *34*(1), 63–70.
26. Reynolds, K., Kontostathis, A., & Edwards, L. (2011). Using machine learning to detect cyberbullying, in: Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on, Vol. 2, IEEE, 2011, pp. 241–244.
27. Dinakar, K., Reichart, R., & Lieberman, H. (2011). Modeling the detection of textual cyberbullying., in: The Social Mobile Web.
28. Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiS)*, *2*(3), 18.
29. Dadvar, M. de Jong, E., Ordelman, R., & Trieschnigg, R. Improved cyberbullying detection using gender information.
30. Bigelow, J. L., Edwards, L., et al. (2016). Detecting cyberbullying using latent semantic indexing, in: Proceedings of the First International Workshop on Computational Methods for CyberSafety, ACM, pp. 11–14.
31. Zhao, R., & Mao, K. Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. IEEE Transactions on Affective Computing.
32. Nahar, V., Li, X., & Pang, C. An effective approach for cyberbullying detection, Communications in Information Science and Management Engineering.
33. Nahar, V., Unankard, S., Li, X., & Pang, C. (2012). Sentiment analysis for effective detection of cyber bullying. Web Technologies and Applications. Springer, pp. 767–774.
34. Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lv, Q., & Mishra, S. Detection of cyberbullying incidents on the instagram social network. arXiv preprint arXiv:1503.03909.
35. Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lv, Q., & Mishra, S. (2015). Analyzing labeled cyberbullying incidents on the instagram social network, in: International Conference on Social Informatics, Springer, pp. 49–66.
36. Hosseinmardi, H., Rafiq, R. I.., Han, R, Lv, Q., & Mishra, S. (2016). Prediction of cyberbullying incidents in a media-based social network, in: Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM International Conference on, IEEE, 2016, pp. 186–192.
37. Wulczyn, E., Thain, N., & Dixon, L. (2017). Ex machina: Personal attacks seen at scale, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, pp. 1391–1399.
38. Chu, T., Jue, K., & Wang, M. Comment abuse classification with deep learning.
39. Xu, J.-M., Huang, H.-C., Bellmore, A., & Zhu, X. (2014). School bullying in twitter and weibo: a comparative study. *Reporter*, *7*(16), 10–14.
40. Cortis, K., & Handschuh, S. (2015) Analysis of cyberbullying tweets in trending world events, in: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business, i-KNOW ’15, ACM, New York, NY, USA, 2015, pp. 7:1–7:8. 10.1145/2809563.2809605. http://doi.acm.org/10.1145/2809563.2809605.
41. Campfield, D. C. (2008). Bullying and victimization: Psychosocial characteristics of bullies, victims, and bully/victims, Graduate Student Theses, Dissertations, and Professional Papers (University of Montana). ProQuest, 2008.
42. Salmivalli, C., Huttunen, A., & Lagerspetz, K. M. (1997). Peer networks and bullying in schools. *Scandinavian Journal of Psychology*, *38*(4), 305–312.
43. Xie, H., Swift, D. J., Cairns, B. D., & Cairns, R. B. (2002). Aggressive behaviors in social interaction and developmental adaptation: A narrative analysis of interpersonal conflicts during early adolescence. *Social Development*, *11*(2), 205–224.
44. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879.
45. Bless, H., Fiedler, K., & Strack, F. (2004) Social cognition: How individuals construct social reality. Psychology Press.
46. Eagle, N., & Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268.
47. Singh, V. K., & Jain, A. (2017) Toward harmonizing self-reported and logged social data for understanding human behavior, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, pp. 2233–2238.

48. Giles, J. (2012). Making the links. *Nature*, *488*(7412), 448–450.
49. Brighi, A., Melotti, G., Guarini, A., Genta, M. L., Ortega, R., & Mora-Merchán, J., et al. (2012). Self-esteem and Loneliness in Relation to Cyberbullying in Three European Countries. Cyberbullying in the Global Playground: Research from International Perspectives (pp. 32–56).
50. Reis, H. T., & Sprecher, S. (2009). *Encyclopedia of Human Relationships* (Vol. 1). Thousand Oaks: Sage.
51. Cazenave, N. A., & Straus, M. A. (1979). Race, class, network embeddedness and family violence: A search for potent support systems. *Journal of Comparative Family Studies*, *10*, 281–300.
52. Tong, S. T., Van Der Heide, B., Langwell, L., & Walther, J. B. (2008). Too much of a good thing? the relationship between number of friends and interpersonal impressions on facebook. *Journal of Computer-Mediated Communication*, *13*(3), 531–549.
53. Raghavan, V., Ver Steeg, G., Galstyan, A., & Tartakovsky, A. G. (2013) Modeling temporal activity patterns in dynamic social networks, Computational Social Systems, IEEE Transactions on 1.
54. Casey-Cannon, S., Hayward, C., & Gowen, K. (2001) Middle-school girls' reports of peer victimization: Concerns, consequences, and implications., Professional School Counseling.
55. Mishna, F., Wiener, J., & Pepler, D. (2008). Some of my best friends experiences of bullying within friendships. *School Psychology International*, *29*(5), 549–573.
56. Mishna, F., Cook, C., Gadalla, T., Daciuk, J., & Solomon, S. (2010). Cyber bullying behaviors among middle and high school students. *American Journal of Orthopsychiatry*, *80*(3), 362–374.
57. Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social Networks*, *25*(3), 211–230.
58. Gilbert, G., & Karahalios, K. (2009) Predicting tie strength with social media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 211–220.
59. Olweus, D. (1980). Familial and temperamental determinants of aggressive behavior in adolescent boys: A causal analysis. *Developmental Psychology*, *16*(6), 644.
60. Morahan-Martin, J., & Schumacher, P. (2003). Loneliness and social uses of the internet. *Computers in Human Behavior*, *19*(6), 659–671.
61. Newman, M. (2010). *Networks: An introduction*. Oxford: Oxford University Press.
62. Altshuler, Y., Fire, M., Shmueli, E., Elovici, Y., Bruckstein, A., Pentland, A. S., et al. (2013). The social amplifier reaction of human communities to emergencies. *Journal of Statistical Physics*, *152*(3), 399–418.
63. Dávid-Barrett, T., & Dunbar, R. (2013). Processing power limits social group size: Computational evidence for the cognitive costs of sociality. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1765), 20131151.
64. Gupte, M., & Eliassi-Rad, T. (2012) Measuring tie strength in implicit social networks, in: Proceedings of the 3rd Annual ACM Web Science Conference, ACM, pp. 109–118.
65. Steinfield, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, *29*(6), 434–445.
66. Walker, K. N., MacBride, A., & Vachon, M. L. (1977). Social support networks and the crisis of bereavement. *Social Science & Medicine*, *11*(1), 35–41.
67. Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*(1), 28–51.
68. Lopez, C., & DuBois, D. L. (2005). Peer victimization and rejection: Investigation of an integrative model of effects on emotional, behavioral, and academic adjustment in early adolescence. *Journal of Clinical Child and Adolescent Psychology*, *34*(1), 25–36.
69. Nahar, V., Al-Maskari, S., Li, X., & Pang, C. (2014) Semi-supervised learning for cyberbullying detection in social networks, in: Australasian Database Conference, Springer, pp. 160–171.