An education-based approach to aid in the prevention of cyberbullying


In this paper, we outline a multi-faceted technological approach for addressing the problem of cyberbullying, focussing on actions which can be taken within educational institutions, for students in elementary and secondary schools. Our proposals include a specific program to introduce in schools, one that will engage students to discourage cyberbullying. This is accompanied by a few distinct approaches for analyzing data within online social networks, which may help to identify potential victims and bullies, resulting in additional insights for students. We discuss as well other parties who can partner with educational institutions in curtailling cyberbullying, notably ones who can assist in providing safer social networking environments. We conclude with a reflection on the responsibility of all individuals, learning from student years through to adulthood, inspired by lessons revealed from our reflections on this important topic.

Keywords: cyberbullying, social media, education
Corresponding Author: R. Cohen
Email: rcohen@uwaterloo.ca

Introduction

To all those who have witnessed cyberbullying online and who wonder how we can change the mindset of our youth today, this paper provides some initial reflection, from a group of authors who are primarily young graduate students. In this paper, we examine the problem of cyberbullying and provide some new directions to reduce the prevalence of this practice. Our primary focus is to reach out to elementary and secondary school students, with an approach that is especially meaningful and engaging for this generation. We additionally explore the specific context of online social media, outlining some strategies that can be introduced in order to mitigate cyberbullying, leading to added insights for those delivering education on this topic to students.

While traditional bullying has been defined as "an aggressive, intentional act or behaviour that is carried out by a group or an individual repeatedly and over time against
a victim who cannot easily defend him or herself," cyberbullying refers to "use of electronic forms of contact for the purposes of realizing bullying acts". Bullying is "unwanted, aggressive behavior that involves a real or perceived power imbalance". The impact of this behaviour is felt considerably within today's social networks, due to the prevalence of an online presence from children and young adults, with viral spread of comments and an ability to sometimes hide behind alternate identities. In fact, some research suggests that cyberbullying is positively correlated with time spent online.

As there is evidence of very strong usage of social media by people between the ages of 18 and 29 currently in the United States, it becomes quite important to consider this environment as one about which to engage students in education about cyberbullying. Facebook in particular has been identified as a platform of interest and becomes the focus of our investigation of additional techniques and strategies to assist students. We continue to be concerned as well through reports that suggest that the severity of bullying online may be even more disconcerting than traditional bullying environments.

Indeed, students within our chosen demographic have been shown to have awareness already of cyberbullying incidents, making the need for initiatives to deliver education all the more important. For instance, according to a report from EU (European Union) Kids Online in 2010, twenty-nine percent of 11 to 16 year olds had seen one or more of the negative forms of user-generated content listed. Perhaps even more disconcerting is the growing number of students reporting being victims of cyberbullying, with evidence in the United States of the problem rising considerably as children enter middle school.

Such statistics suggest the merit of reaching out to students at a relatively early age, as we propose.

Specific cyberbullying cases have attracted media attention in recent years, in Canada. One such case is that of 15 year old Amanda Todd from British Columbia, who committed suicide after being repeatedly harassed online. Weeks prior to her death, she

---

posted a video telling her story and what led to her depression. The Rehtaeh Parsons case in New Brunswick resulted in efforts by the provincial government to enact stricter cyberbullying laws. In Ontario, Jamie Hubley’s loss of life due to cyberbullying caused his parents to lead an impassioned plea for the federal government to examine these cases more significantly. American cases that have made headlines include those of Tyler Clementi (an undergraduate student in New Jersey) and Jessica Logan (a secondary school student in Ohio), whose tragic deaths led to bills in their names being passed in their states, to attempt to address cyberbullying. The fact that specific cases continue to draw the public's attention motivates our suggestion of introducing one of these cases, that of Amanda Todd, into our effort to educate Canadian students about cyberbullying. As we discuss towards the end of this paper, proper curation of content when cases like these are introduced into the curriculum needs to be done by educators, in order to gauge the proper maturity level and tolerance of the students.

An educational program to reduce cyberbullying

At the heart of our proposal is a strategy to enable elementary and secondary school students to understand both the technology and the social implications of the cyberbullying problem. Once educated, potential observers and victims would be able to identify a bullying problem in order to react to it promptly. Our teaching plan also leverages computer technology to educate students. The focus is on children aged 8 to 16 (grades 4 to 12): the aim is to engage students when they may first be exploring the Internet. Basic computer literacy skills (mouse and keyboard, browsing) would be assumed. We would also expect the school educators to acquire some background knowledge of cyberbullying. The importance of education is emphasized well by a recent OECD (Organisation for Economic Co-operation and Development) report on bullying. The claim is that educated individuals are more likely to be actively engaged in society and through development of cognitive, social and emotional skills, may be more resilient to attempts of victimization.

Program Outline

The proposed program has been designed to fit into one session of a Computer Science class, of 1 hour 20 minutes duration, aiming for a class size of at most 25. The content of the program is structured such that the students are given an introduction to the issue of

12 "Gay Ottawa teen who killed himself was bullied." CBCnews. CBC/Radio Canada, 19 Oct. 2011.
15 Skills for social progress: the power of social and emotional skills. K. Miyamoto et al. OECD Publishing. 2015.

ACM Computers & Society | Volume 47 | Issue 4 19
cyberbullying and its various manifestations, and given the opportunity to discuss an example shown to them. In addition to this, there is also an interactive module where students can participate in a simulated conversation with a chatbot to get a clear idea of what is exactly involved in cyberbullying.

i. Presentation on Cyberbullying

The session will begin with a presentation describing cyberbullying and the way it can occur on different social media platforms like Facebook, Twitter, Instagram, YouTube, Tumblr etc, or in texting-based applications like Snapchat, Messenger, Whatsapp, Skype or SMS. It will also contain information about the resources available to children who might be facing cyberbullying, for instance, KidsHelpPhone, a hotline for children to reach out for help, which also has a website (www.kidshelpphone.ca) with games relating to cyberbullying incidents, which might be useful to spread awareness; GetCyberSafe (www.getcybersafe.gc.ca), a website with many useful articles about different cybersecurity risks, and different resources for both parents and children to help combat cyberbullying.

ii. Class Discussion

After the presentation, we want to ground discussion in a real world case. For example, after viewing the video of Amanda Todd, students could be asked: a) Why do you think people treated Amanda as they did? b) What could have been done to help her? c) Do you think she deserved what happened to her? d) What would you do if you found yourself in that situation? e) Some people feel that children should be monitored while on the Internet, so as to prevent cyberbullying. What do you think about this?

This video has garnered widespread attention because of Amanda’s suicide, and has also been the subject of several debates among young YouTube commenters, so it is expected that it might provoke controversial opinions among the students in this session, as well. One possible strategy is to treat the discussion as a supervised sharing of opposing viewpoints, where students acquire skill in expressing their opinions and see the value of this kind of engagement. If successful, we may even observe students starting to post online in a manner that is more well planned and considerate.

iii. A Chatbot for Simulated Conversation

Sometimes, children might not report instances of cyberbullying to their parents or an adult, despite knowing that they ought to, if they do not realize that what they are experiencing is, in fact, cyberbullying. It is challenging to present examples of cyberbullying to students, as this might in turn provoke behaviour that is undesirable. On the other hand, it may be more problematic to simply leave it to students to experience these same exchanges online without the benefit of accompanying class discussion to provide advice and insights. Below we sketch some ideas for a chatbot, but urge important deliberation on choice of content, before the bot is introduced. The chatbot could additionally be used to suggest how students can cope with cyberbullying, what they can do to avoid being bullied online, and to spread awareness about not bullying
others. In fact, one interesting idea would be to construct an affirmation bot, with specific exchanges that promote self-healing exercises.

There are several online platforms where chatbots can be created and customized easily. We use Rebot.me to create a customized chatbot called CyBully, which is programmed to respond to most user inputs with insulting remarks, as matching the profile of real-life cyberbullies who make comments on the appearance or intellect of the victim. Alternately, CyBully also tries to get the user to send photos or to divulge information about where they live, mimicking the behaviour pattern of cyberstalkers. When presenting negative comments, the content of CyBully’s responses can be censored to suit the age of the users. We believe that having children interact with a simulated cyberbully would help students to identify what sort of speech characterizes a bully, in part to help them avoid using such word themselves. Since a teacher or adult would oversee the session, we trust that their guidance would leave students with a positive lesson.

One sample conversation with CyBully might be, for instance:

CyBully: Oh, I live in Canada too. Shall we meet tomorrow?
User: No
CyBully: Anyone who told you to be yourself couldn’t have given worse advice
User: What? That’s mean
CyBully: No, I’m being honest

Discussion

We believe that using a chatbot for simulated conversation could prove to be very effective in the long run. Sometimes students are hesitant to discuss what they have experienced online with their parents and advisors, either out of embarrassment or uncertainty about whether an incident could be called cyberbullying or not. Even if students might not feel free to express their true opinions in a supervised environment like a classroom it is hoped that such hesitation would not be universal among students, and that listening to other opinions about the subject could have positive effects on the viewpoints of students who prefer to be uninvolved in the debate.

We note as well that including a required module on cyberbullying into the school curriculum makes this education transparent to parents, who have an important role to play in this discussion. According to an article published in Australia, cyberbullying is at the top of parents’ minds when teens go back to school. If parents are confident that educators are aware of cyberbullying and taking active measures to prevent it, they can be more confident about the safety of their children in schools.

Our approach for delivering education about cyberbullying can be contrasted with other efforts in our country of Canada. One government initiative comes from the Royal Canadian Mounted Police (www.rcmp-grc.ca) where a brief 3-minute video outlines 16 "Cyberbullying top of parents' minds as teens go back to school." J. Herald and J. Wong. The Age; January 30, 2017.
different forms and dangers of cyberbullying, including prevention advice. This material
could certainly complement the module we envisage offering within schools; it also
reinforces the belief that students will find this medium an appealing method for
delivering education. The government of Canada has also established the National Crime
Prevention Strategy (www.publicsafety.gc.ca) which includes examination of bullying
among children, proposing evidence-based measures to be introduced within individual
schools. This also provides support for our approach of exposing students to real
elements of cyberbullying, and to offering a targeted curriculum within each school.

Education from insights about social media

Since social media environments are ones in which cyberbullying messages may prevail,
it is of value to examine how to detect instances of bullying in these contexts. In this
section, we outline some strategies which may assist in revealing cyberbullying,
accompanied by recommendations for how to improve social media environments in
order to increase awareness of cyberbullying and to provide users with more control. We
begin with some discussion of the use of sentiment analysis to examine more closely the
content of social media posts. From here, we propose some additional measures which
serve to encourage users to reflect on their comments when they can be seen as
potentially harmful towards an individual. We also outline extensions which would
enable victims to have greater control over their content, thus reducing their vulnerability.

A role for sentiment analysis with added data

Although the area of sentiment analysis for bullying is still emerging, experts in the field
of cyberbullying and its prevention have positive views on the desirability of automatic
monitoring.\textsuperscript{17} Many of these solutions are built on the assumption that there is no
available labeled training data, and instead rely on hand-picked seed words, hand-picked
terms, and explicit mentions of bullying in online text.\textsuperscript{18} We note, however, that
additional data for a sentiment analysis phase could potentially be obtained if we were to
introduce extensions to current social media environments which enable reporting of
activity.

We illustrate what we have in mind for the context of Facebook. Currently, Facebook has
systems for reporting posts, photos, and links. For sentiment analysis, we focus only on
the reporting system for posts. In the process of reporting a post, Facebook asks the user
for further information regarding the reasons for reporting. The user is guided through
several categorization options: (1) It’s rude, vulgar, or uses bad language (2) It’s sexually
explicit (3) It’s harassment or hate speech (4) It’s threatening, violent, or suicidal. For
cyberbullying detection, we are most interested in categories 3 and 4. Of particular
interest are posts that are classified as hateful towards an individual, or as a credible

\textsuperscript{17} Automatic monitoring of cyberbullying on social networking sites: From technological feasibility
\textsuperscript{18} Fast learning for sentiment analysis on bullying. J. Xu et al. In Proceedings of the First
threat of violence. This user-labeled data could provide a classifier with an ever-growing training data set that incorporates the many topics found on Facebook.

Proposed Reporting Process

To deal with offensive comments on their posts, users currently have an option of removing the comment. While a report link is also supported, this process can be improved. For example, even the simple removal of a comment could prompt the user for more information. And since comments tend to be primarily directed at the original poster (or perhaps other users who comment as well), we can imagine modifying the reporting process. Before submitting a report on a comment, the user could be asked a final question: “Is this content bullying”? The prompt might ask the user for a binary response or allow the user to move a slider to indicate to what extreme the post exhibits bullying characteristics. These steps would not only support the anti-bullying effort but might also help to increase the size of the data set for sentiment analysis.

Additional options towards greater awareness and filtering

In an effort to further identify possibly damaging sentiment expressed in social media, we have advocated greater reporting. We now suggest additional options to control the distribution of content.

Requiring confirmation before posting

One aid to filtering content and thus possibly lessening the impact of cyberbullying would be to allow posters to think before finally distributing their messages. To explain, the problem of cyberbullying is exacerbated by feeling a lack of responsibility for what is posted.19 Anything can be said with ease, and the poster is far removed from the recipient’s reaction. To mitigate this, we recommend the automatic analyzing of text sentiment before it is posted. If flagged as bullying, the poster will encounter a pop-up window that emphasizes the gravity of cyberbullying, and encourages editing of the post. By confirming, the user is forced to actively accept responsibility for the flagged post. This strategy of encouraging potential bullies to reconsider their flagged messages has been previously explored by Google Science Fair finalist Trisha Prabhu, who claims that adolescents reconsidered their messages 93% of the time in her ReThink project.20

Two designs are proposed to present this notification. The first is to indicate “This content may violate the following Facebook terms & conditions” and would ask the user whether they would like to proceed or not. This flag would appear if the sentiment analysis deemed a message as a potential violation of Facebook’s Terms of Service (http://www.facebook.com/terms). The second part of the notification will consist of the term or condition that best matches the user’s behaviour, to encourage users to engage in behaviour that aligns with the Statement of Rights and Responsibilities set by Facebook.

This could include any of the following under the Safety section of Facebook’s Terms of Service: a) You will not bully, intimidate or harass any user b) You will not post content that: is hate speech, threatening or pornographic; incites violence; or contains nudity or graphic or gratuitous violence c) You will not use Facebook to do anything unlawful, misleading, malicious or discriminatory.

The second proposed notification is to require the user who is commenting to think about how this might be affecting the initial poster. The exchange would have the header “This comment may contain sensitive material” and the user would be asked “How would you feel if this comment were made on your post?”, then asked “Would you like to proceed?”. This challenges the user to accept responsibility for whatever emotion will be invoked by their posting.

Allowing recipients to review and deny distribution

With processes in hand which aid in the flagging of messages as possible instances of cyberbullying, we suggest as well that the social media environment be configured to allow a view of disconcerting comments, before these comments are posted for the world to see. Even if a potential bully confirms their intention to post a flagged comment, the recipient has the chance to review it and accept or reject it, before it actually becomes a comment on said post.

The original poster will receive a notification. The poster is then presented with three options for addressing the comment: a) Review: the user can review the comment which includes viewing the attempted posted comment b) Delete: the user can immediately delete the comment, so that they do not have to even view c) Review Later: the user can ignore and review at a later time. The rationale for these options was to ensure that the user has complete control. The delete option was included so that the user does not have to view a comment that could be triggering to their mental health or that they simply do not want to view.

A final suggested extension to the social media environment is to enable restrictions on distribution, in an effort to stem the tide of widespread sharing which may exacerbate the effects of the cyberbullying, thus possibly reducing the anxiety of the potential victim. The posting mechanism would integrate yet another ‘Feedback’ option, which will open up a sub-component, similar to the ‘Friends’ option, specifying who can view a post. In a ‘Feedback Type’ sub-component, users can specify what kind of feedback (e.g likes, comments) they would like to receive on a particular post, which are indicated by check boxes.

A sub-component to “Feedback Type” is proposed to give users even more control. This includes options, which each open to take user input. These are: a) Not permitted to like: will open up a list specifying which users will not be permitted to like the specific post b) Not permitted to comment: will open to a list specifying which users will not be permitted to comment on the specific post c) Disallowed words: will open to a list specifying words the user wishes to ban from occurring in comments on the specific post. These designs were essentially chosen to give users or in this case, potential bullying victims, ‘ultimate control’ to reduce the bully’s power.
Next steps

While Facebook currently hosts a valuable hub as a resource for its users (http://www.facebook.com/safety/bullying), there are additional responsibilities that all social media companies could be encouraged to assume, as part of the effort to reduce the effects of cyberbullying. It has been suggested that removing an offender’s access to Internet and technology is viewed as an effective punishment, both by cyberbullies and their non-bully peers.\(^{21}\) Even if a victim is reluctant to directly report a bully, a bully may be automatically reported if their flagged posts have been rejected by a reviewer a pre-specified number of times. Facebook itself already declares "If you violate the letter or spirit of this Statement, or otherwise create risk or possible legal exposure for us, we can stop providing all or part of Facebook to you. We will notify you by email or at the next time you attempt to access your account.". It should thus be possible for Facebook to review the posts which have been labelled as offensive, to then consider disabling or terminating the bully's account. There are challenges in enacting this procedure, however, as discussed in greater detail below.

Social media organizations like Facebook could also adopt the practice of partnering with schools to put their concerns for the welfare of students in these environments quite front and centre. If some of the added features for mitigating cyberbullying that we advocate here were to be introduced, representatives of social media could appear at schools to make everyone aware of the changes and to champion these new directions.

Governments and researchers could also have increased awareness of Facebook's efforts to address cyberbullying, if the environment were used to support additional discussion on this topic, for example within a bullying prevention hub. Feedback from participants about the effectiveness of the current design could also be helpful to suggest revisions and extensions, for the future.

We also envisage opportunities for social media participants themselves to spread the word about the new features that we are proposing. Previous study outcomes indicate that people are easily attracted to the hot topics, following the trends and then participating. Since the trending section is always prominent on the website, it can have some potential for the prevention and awareness of cyberbullying.

Other directions for the future would be to expand the algorithms for detection of cyberbullying within social media environments. Certainly, offensive behaviour may be observed not only in direct postings but also in comments made about profiles or pictures online. It may also be quite valuable to examine the larger-scale behaviour of certain social media participants. For example, cyberbullies are often involved in high-frequency communications with their targets, creating a constant concern from the victims that attacks are imminent.\(^{22}\) Evidence of mass distribution of suspect messages may be another warning flag that the communication merits greater scrutiny. In addition, patterns

---


of deactivating accounts and blocking of friends may serve to signal some distress on the part of a social media participant, suggesting that this may be a potential victim. Perhaps looking more into the future, an interesting direction would be to analyze not just texts but also videos in order to discern negative emotions, inspired by existing research on image analysis.23

If technology can assist in spotting bullies and victims, it may be possible to create some kind of manual, with example cases, which could serve as an added element to the educational module delivered to students in classrooms. The aim would be to encourage students to recognize when their friends are involved in cyberbullying, and thus increase awareness. We recognize the need to carefully curate this content, as it is less likely that students will simply opt in to having their postings viewed by parents or educators.

**Challenges to consider**

We note that if student awareness of cyberbullying is increased, additional support for certain individuals may also be needed, within the educational institution, such as assistance from school psychologists to assist those who are most fragile. Those who self-identify as victims can be encouraged to visit a consultant in order to benefit from individual attention. Those who self-identify as bullies would also benefit enormously from individual consultation; greater insights into what causes this behaviour may be obtained and passed along to the educators (of course, in a totally anonymous and aggregated fashion). Students in distress who do not volunteer for assistance (or make their parents aware of a need for greater support) may potentially be identified by educators through other external behaviour during the in-class sessions. While at-risk students may potentially be identified through an analysis of their social media postings, actually sharing those postings with educators or parents is challenging, due to privacy concerns.

**Challenges in enhancing social media**

The specific extensions that we have proposed for social media environments may face certain roadblocks as well. It may be possible for users to abuse the report function to flag inappropriate content. Addressing these cases of misinformation could then also become an additional burden to Facebook. Apart from unfair reporting, defining the inappropriate content is another area of difficulty, as there may be valid differences of opinion between different users as to what constitutes inappropriate communication.24 Additional challenges may arise in conducting sentiment analysis when the language used is not English, as cultural differences exist about offensive behaviour. And our suggestion of introducing disallowed words may even generate additional harm, as it makes users more aware of these options. Finally, we also acknowledge that simply alerting a user to potentially sensitive comments and asking whether they want to

reconsider whether this is offensive may not result in any changes. Users may click through the warning without any notice of what has happened, much as they tend to merely flip through the terms of agreement for using the social network in the first place.

**Challenges in delivering the best educational content**

We also notice some unique challenges in ensuring that the educational content assists students in properly reflecting on cyberbullying. Dramatic content such as that of the Amanda Todd case\(^\text{10}\) may be determined to be too strong to place in front of children at a younger age, resulting in difficulties for the school boards that are offering the cyberbullying program. One may argue that this content was already available online for the students to discover and that our education would truly help to provide a better context of understanding, but some dialogue with parents, school board administrators and perhaps government may be needed.

We also acknowledge that due to the anonymity of many online environments, there may be students who have completed our proposed course and who still express inappropriate behaviours online. This remains a separate concern for educators and governments to explore. Another interesting challenge is trying to provide education for students in developing countries, where Internet access may be sporadic. We certainly imagine that some adaptation to local culture may be required with the development of the curriculum.

Finally, the chatbot we have introduced to keep students engaged about the topic has only been outlined at a fairly high level, and our current implementation for instance was only trained on a very small set of sentences. A greater exploration of how to operate this chatbot, with far more extensive training, is another challenge for the future.

**Conclusion**

In this paper, we have brought attention to a need to provide education about cyberbullying to youth in elementary and secondary schools. We have drawn on literature and cases in Canada, to imagine a possible program for our country, but feel that many of the strategies that we advocate can be effective for other nations, as well. We have also emphasized the role that social networks play in exacerbating the problem, suggesting methods for mitigating effects in these contexts and avenues for obtaining additional educational material.

While it is important to be reaching students at an early age, we are well aware of the challenges in properly introducing sensitive material. Certainly, educators need to be part of the team to properly design the instruction that we envisage, partnered with computer scientists who can continue to conceive of the proper technology to employ. We have suggested for instance the use of a chatbot as a method for effectively engaging the students and innovative methods such as these are intended to increase the educational value.

We have discussed briefly additional roles that can be played by those who are managing popular online social networks like Facebook, and tactics that can be adopted by
governments, in order for those in this sector to be effective partners in the effort to address cyberbullying. We have pointed out as well the prevalence of social media of bullying incidents in these environments, suggesting that education and awareness should not simply be initiated in earlier grades and then become out of focus. Indeed, as students progress in age, the hope is that they will draw from the lessons learned and continue to act responsibly, into adulthood. In fact, if one day these students become parents, they could be truly valuable contributors in the joint effort to assist youth towards a more productive online existence.

While challenges still remain in order to move forward with the proposals outlined in this paper, we feel that as long as there is sufficient engagement from all parties in the future, then we can make the important strides forward which are so vital to addressing the social concern of cyberbullying. And as we have tried to argue, computer scientists have a very significant role to play in this effort.