Insights into Adolescent Online Conflict through Qualitative Analysis of Online Messages

Megan L. Ranney, MD MPH
Rhode Island Hosp./Brown Univ.
megan_ranney@brown.edu

Kyler R. Lehrbach
Rhode Island Hosp./Brown Univ.
kyler_lehrbach@brown.edu

Nicholas A. Scott
Brown University
nicholas_scott@brown.edu

Nicole R. Nugent
Brown University
nicole_nugent@brown.edu

Alison Riese
Brown University
alison_riese@brown.edu

Jeff Huang
Brown University
jeff_huang@brown.edu

Grant Fong
Brown University
grant_fong@brown.edu

Rochelle K. Rosen
The Miriam Hosp./Brown Univ.
rrosen@lifespan.org

Abstract

Given adolescents' widespread use of online messaging and social media, as well as the prevalence of cyberbullying, analyzing adolescents' online message-based communication topics and patterns is relevant to public health. To better describe conflict in adolescent online communication, this paper analyzes patterns of conflict in a dataset of adolescent online messages. We describe a qualitative methodology for analyzing these complex data, to expand understanding of adolescents' online conversations, and to identify how best to categorize conflict within online media datasets. In this study, 14,239 messages from 20 adolescents in the Northeast United States (of which 1,911 were coded) were analyzed using thematic analysis. Several distinct kinds of conflict and responses were identified. Conflict was either direct or indirect, serious or non-serious; it most often was indirect and serious, referenced either insults or romantic contacts, and was frequently related to in-person fights. Coding relied on understanding both textual contexts and referents.

1. Introduction

Social media and online messaging program use is nearly ubiquitous among adolescents. Approximately 95% of teens own a smartphone, and 45% say that they are "always online" [1]. YouTube, Instagram, Snapchat, and Facebook are used by the majority of teens, and almost all teens report texting with friends [1].

Social media and online messaging can have either positive or negative effects on adolescent well-being. These tools can be a helpful way to connect with friends or family and may provide an opportunity to share positive messages [1]. However, they can also reduce in-person social interactions and provide a platform conducive to conflict [1, 2]. In a recent national survey, 59% of teens report experiencing at least one form of cyber-harassment or bullying, and 90% of youth say online bullying is a problem [3]. Cyberbullying has been recognized as an urgent public health issue [4]. Exposure to cyberbullying and cyber harassment has also been linked longitudinally to depressive symptoms and engagement in other risky online behaviors, such as sexting [5, 6].

Researchers and clinicians are inconsistent in their ways of both *describing* online conflict [7], and *measuring* the different levels of violence that can occur online [8]; common terms include, for instance, "cyber stalking," "cyber aggression," "online harassment," "internet bullying," "cybervictimization," and "electronic aggression" [4]. Cyberbullying is generally described as acts of aggression that occur through the medium of the internet or social media and involve an imbalance of power between the victim and the aggressor [2]. However, debate exists about the centrality of several elements of cyberbullying – including repetitiveness and intent to harm [9]. Reflecting this lack of agreement as to what does and does not constitute "cyberbullying," there are a wide range of instruments, mostly surveys and questionnaires, used to measure cyberbullying. There is no gold standard for self-report of incidence or severity of online bullying or harassment [4] nor is there clear understanding, among adolescents or researchers, about what constitutes problematic online discourse.

Direct evaluation of social media content, through either qualitative analysis or automated quantitative measurement, may provide alternatives to self-reported experiences of online conflict. For example, preliminary work has described hand-coded patterns of online conflict among gang-involved adults, suggesting that it is possible to identify high-level patterns of conflict using social media [10]. Others have measured the sentiment of communities, as expressed through social media, after exposure to violence [11, 12]. The LIWC (Linguistic Inquiry Word Count) method measures occurrences of words defined in the dictionary to comprise a finite set

HĬCSS

of categories, such as negative emotion, positive emotion, anxiety, anger, conflict, or sadness, and may be able to be applied to identify online conflict [13]. However, while automated analysis is more efficient, the loss of context may reduce accuracy. For example, past research found a positive but weak correlation between the two methods in analyzing the sentiments of tweets [13]. Moreover, machine learning is only as good as the initial classification scheme on which it is developed [14]; a necessary step in developing automated coding methods is to first develop accurate human-level coding. To our knowledge, qualitative coding has not been systematically applied to classification of adolescent online drama or conflict.

As a first step in developing more accurate coding schemes for online conflict, we describe the methods used to obtain and hand code online conflict among adolescents' online messages. In comparison to other studies that measure hypothetical situations [15], draw inferences from survey results [5, 6], or describe ecological level data [13, 16], this paper gives insight into how conflict is referenced – both directly and indirectly – and responded to, in adolescents' online communication. We also comprehensively discuss the process of obtaining and analyzing online message transcripts.

2. Methods

Data Download: As part of a larger study to develop a cyberbullying-prevention intervention, adolescents in a pediatric primary care clinic in an urban, Northeastern US teaching hospital, who had reported prior experience with online conflict or bullying, were asked to allow pseudonymized downloading of four weeks of online messages during an in-person follow-up [8]. At the time of written assent and consent for the larger study, teens were told that they would be asked for permission to download their online messages at a later time. At the time of the follow-up interview (eight weeks after study enrollment), teens who completed their interviews in person (N = 34) were asked if they would be interested in allowing researchers to download their data. The Sochiatrist software, developed by the author JH, was briefly explained to the participants, and their confidentiality was assured. Teens were given the option to only allow downloading of certain messages or message platforms. Of the thirty-four participants completing in person follow-ups, twenty (59%) consented to the data download. Participants' mean age was 14.7, 65% female, 85% non-white, and 50% Hispanic or Latino. There were no significant differences between those who did and did not permit data download in age, gender, race/ethnicity, or self-reported cyberbullying scores.

Sochiatrist software was used to extract messages from participants' cell phones [17]. The Sochiatrist system enables researchers to collect social media and messaging data from multiple web and mobile-based platforms through the use of an automated script. The data extractor can be used to collect direct messages from Facebook, Instagram, Twitter, Kik, WhatsApp and SMS (iOS's Messages) on both Android and iOS phones. These platforms are the most used messaging platforms in the US [1]. Participants must be physically present to input their login information for each messaging platform. Monitoring of participants outside of the extraction is not possible with this system. In order to extract messaging data from these platforms across the different devices participants may have, multiple extraction strategies must be used ranging from directly using an application programming interface (API) (Twitter), using a data export tool (Facebook), to reading data off of a phone backup (iOS).

After extraction, the messages are aggregated and compiled in a consistent format using a script, and the researcher is prompted to specify the time range of messages for their final Comma-Separated Value (CSV) file. In this study, data from two separate two-week time periods were extracted. Only data within the specified time range is saved. Identifying names in the data are pseudonymized as detailed below. In addition to being able to delete any interactions prior to the download, the participant could also request for certain conversations to be removed from the data after it was downloaded. At the end of the process, all backups, data downloads, and non-pseudonymized files are deleted from the lab computer.

Pseudonymization: After the extraction and compilation process, the message content, medium used, recipient of the message, and time/date were pseudonymized.

Multiple methods of pseudonymization were used to protect the identity of participants and their conversation partners, including the removal of names and potential identifiers. Names and numbers were replaced respectively with hashes or substitute symbols. In order to pseudonymize numbers, a simple regular expression search was run over the messages, replacing numbers with the # symbol. This preserves the context and the general form of the numbers, so one is able to still interpret the type of a specific number, such as a dollar amount or telephone number. Names are replaced by a pseudo-random string that is internally consistent, mapping the replacement name across platforms. The Sochiatrist pseudonymizing capabilities are imperfect, as messages that contained shortened names, obvious nicknames, or abbreviations of locations could be missed by the automated pseudonymization process. Therefore,

before any coding or interpretation of these messages began, Research Assistants (RAs) read through each message to confirm pseudonymization and further remove or pseudonymize any missed names, locations, or other identifying information by hand. Finally, the messages were grouped by conversation (e.g., between pseudonymized IDs) and ordered by time.

Hand Coding: Once the message data were cleaned and pseudonymized, two RAs (KRL & NAS), in collaboration with investigators RKR and MLR, iteratively developed a coding scheme specifically focused on topics relating to online conflict and drama in teens. Both research assistants, in their early twenties, were familiar with adolescent slang, including common euphemisms and texting abbreviations. The coding scheme was initially deductive, based on pre-identified themes of interest to the ongoing research and drawn from current literature. This initial set of codes was then revised, based on inductive or emergent topics, negative case analysis, and areas of variability or disagreement amongst the coders. Once the coding scheme was finalized, the two RAs coded all transcripts independently, then met to come to consensus on the coding. Remaining coding conflicts were resolved through discussion with RKR and MLR. Agreed-upon double coded data were entered into NVivo qualitative data analysis software [18]. A summary of the codes used and number of references per code is displayed in Table 1.

Because consent was provided by our participants to access their online conversations, only their own sent messages were coded. Once all general codes were entered into the software, they were grouped into "conversations" which were identified by both time and content. Overall, of the 14,239 sent message lines, 1,911 (13.4%) were identified as meeting criteria for coding.

Table 1. Partial coding structure:
Number, title and total references

Code	Title	Total References (N=2,521)
1. Online Conflict		
1.1	Direct Online Conflict	501
1.1.1	Direct Online Conflict of a Non-Serious or Sarcastic Nature	333
1.2	Indirect Online Conflict	1161
1.2.1	Insulting Someone not in the Conversation	107
2. Discussion of Deviant Behaviors		
2.1	Discussion of Violence	55
2.1.1	Inciting Offline Violence Online	9
2.2	Discussion of Drugs	87
2.3	Discussion of Other Deviance	27
3. Reactions to Conflict		
3.1	Exacerbation of Conflict	72
3.2	Attempt at Resolution of Conflict	83
3.3	Unclear Response to Conflict	86

Qualitative Coding Summaries: Once all coding was complete, summaries were written for each relevant code. Messages within a code were grouped into categories, and trends within codes were tracked. Double codes were also identified (that is, instances where a single message was independently coded into two or more separate codes), allowing analysts to draw conclusions about potential relationships between various codes. The conversation codes were used to identify whether codes occurred more frequently within the same conversation or if they occurred independently. Coding Summaries contained detailed explanations of observations and trends within each code, any applicable categorization of coded messages, statistics regarding categorization, double-coding and conversation coding, and examples of messages.

Analytical memos captured key patterns and themes within the data overall. Observed patterns were identified by paying attention to similarities and differences within the data including the theme, setting, and purpose of the messages and conversations.

In the description of results below, messages have been slightly changed to avoid unintentional violation of confidentiality and privacy, as per best practices for social media research [19]. To further guard against participant identification, we have not identified speakers, but examples were drawn from a variety of our participants.

3. Results

Our thematic content analysis of these adolescent online messages identified several distinct kinds of social messaging conflict, or response to conflict, each of which had specific subject and context considerations. In the

sections below, we address 1) the object of the conflict, specifically whether it directly or indirectly references the persons in the conversation and key topics that are relevant to the conflict, including insults and romantic attachments; 2) reactions and responses to conflict; and 3) references to what we have labeled "deviant behavior" such as substance use and truancy. Details of each of these themes are provided below.

3.1. Direct and Indirect Online Conflict

Online conflict identified in message transcripts was interpreted as either direct or indirect. Direct conflict occurred between the participant and the recipient. Indirect conflict referenced third party persons not in the online messages we analyzed. The decision was made to label conversations as "conflict" rather than "bullying" because of difficulty interpreting intent, power differentials, and other relevant variables.

3.1.1. Direct Online Conflict. Messages coded as "Direct Online Conflict" were further divided into either a) messages that were serious, important, and even potentially detrimental to the relationship between participant and the recipient or b) messages that were clearly non-serious or of a sarcastic nature. For example: *"Don't talk to me no more we ain't friends"* was coded as serious direct online conflict and *"If you leave you'll be back anyways so I guess I'll see you in three months"* was identified as sarcastic, non-serious direct conflict.

Both the length (defined by number of messages in the same conversation) and tone signaled the significance of the conflict and its serious or non-serious nature. Clearly written definitions and coding comparison helped coders use contextual inference, tone, and length of conflict to determine whether a direct online conflict message was "serious" or "non-serious". Brevity of the conflict and terms that signaled laughter or joking usually indicated a "non-serious" conflict.

While brevity was often key to understanding messages as non-serious, so were the habits of individual message writers (for example, understanding their patterns of interactions with other teens or in non-conflict situations). Not all joking comments indicated conflict, and accurate coding required attention to sometimes subtle differences. In many cases it was not the initial message itself but subsequent response(s) that made the serious or joking tone clear to the coders. 50% of the teens in the study provided data that contained direct online conflict codes. Specifically, 57% of the male participants and 46% of the female participants contained at least some messages coded as direct online conflict. 76% of the total direct online conflict codes were sent

by female participants. It is worth noting that females were disproportionately more likely to be involved in direct conflict (6.6% of all messages sent by females were coded as direct conflict, compared to 4.4% of all sent messages by males.

3.1.2. Indirect Online Conflict. Many online discussions referenced conflicts with someone who was not part of the conversation. Nearly all instances of indirect online conflict referenced specific people and situations (for example: *"Don't let her fool you…"*). A select few, however, made general statements about groups of people or about broad issues, for example: *"I'm so done with boys."* Instances of indirect conflict could be between the participant and a third party (96%), or could be between two third parties who were not involved in the conversation (representing only 4% of all indirect online conflict). The majority of these insulting messages occurred within short discussions.

Third party insults were generally brief, often out-of-context. By definition, they referred to a person not in the conversation. Third party insults commonly referenced physical appearance *"[name2] is a midget",* personality characteristics, *"shawty mad goofy she be acting its annoying"* or past behaviors:

"-Yes the whole party she was pissing me off
-Then she was screaming at the top of her lungs
-She always makes a scene
-Bc she's a dirty b*tch".

Some insults about third parties were more general, e.g.: *"LMAO I hate him."* Codes were split relatively evenly between these types of indirect conflict.

Forty-seven out of 107 (44%) third-party insults were interpreted as "deep rooted," or reflecting prior or longstanding conflict. Some of these messages referenced bothersome behavior *"corny ass like [name8] wasn't talking sh*t bout [name5] all the time"* or occurred in longer conversations, such as the example above. 40% of the teens in the study provided data that contained indirect online conflict codes. Specifically, 29% of the male participants and 46% of the female participants contained at least some messages coded as indirect online conflict. With this being said, nearly all (89%) of the indirect online conflict codes were sent by female participants. Females were disproportionately more likely to be involved in indirect conflict: 11.8% of all sent messages by females were coded as indirect conflict, compared to 3% of all sent messages by males.

3.1.3. Dating Relationships and Online Conflict. Dating or sexual relationships played an important role in both direct and indirect online conflict. More than half (52%) of the indirect online conflicts and two-thirds

(66%) of the serious direct online conflicts involved past, present, or future dating or sexual relationships or attachments (messages coded as non-serious or joking direct online conflict were not analyzed for romantic references). These messages either described actual conflict directly within the relationships ("*I don't know why I believed for a second that you actually gave a sh*t*"), or referenced indirect conflict regarding a real or theoretical sexual/dating relationship ("*Try and flirt with [name4] so I have a reason to leave him*"). Indirect conflict related to dating/sexual relationships could occur between a member of the conversation and their romantic partner, or could be a conflict related to a relationship. An example of the latter is "*well her friend didn't want me and her to be together*".

3.2. Reactions to Conflict

Adolescents' online reactions to conflict were separated into three categories: 1) exacerbation or encouragement of the existing conflict, 2) attempts to resolve the conflict, 3) unclear or ambiguous responses to conflict.

"Exacerbations of conflict" messages were identified as either substantive (27%) or non-substantive (73%). An example of a substantive message is "*want me to make a fake account and expose her?*"; this message clearly offers to exacerbate an existing situation. An example of a non-substantive message is "*Damn*" or "*Hahahahahahahahaha*." Non-substantive messages implicitly or explicitly encouraged and reinforced the conflict behavior of the recipient, but did not exacerbate it. Many non-substantive replies were responses to an indirect conflict or third-party insult which, by providing the recipient a laugh or validation, is implicitly rewarding and encouraging further insults.

Attempts at online resolutions of conflict employed one or more of the following strategies (see Table 2):

Table 2. Strategies for conflict resolution

Strategy	Example	% of messages
Apologizing	*"sorry babe, next time"*	42
Offering an explanation or excuse	*"Damn my fault for making u come here for no reason I was knock n these kids didn't even wake me up"*	22
Reassuring	*"I'll pay you back but okay good night"*	16
Defusing the Situation	*"Calm down"*	12
Avoiding/ changing the subject	*"Ima leave it alone" "I'm done talking to her"*	8

Many (36%) online reactions to conflict were coded as "Unclear"; these messages were short, one-word responses such as "*nevermind*," "*it's nothing*," "*okay*," "*whatever*," "*oh*," "*lol*," "*eh*," and "*wow*." Many of these "unclear responses" ended a conversation. Some appeared dismissive, suggesting that the participant did not have a fruitful or more substantive response. We saw some repetition of these phrases among several participants, suggesting that the phrases serve as common responses when teens respond to conflict.

When analyzing the relative occurrence of participant reactions, coders noted a strong pattern with respect to the indirect or direct nature of the conflict. Conflict exacerbation was usually (81%) expressed in response to indirect online conflict, particularly discussion of conflicts with a third party. Conflict resolution and unclear responses, on the other hand, often occurred in response to direct conflict: 87% of conflict resolution messages and 64% of unclear responses were in response to a message coded as direct conflict.

3.3. Conflict and Deviant Behavior

Codes were also developed for instances of "discussion of deviant behaviors", such as substance use or truancy, which are theoretically related to in-person and online conflict [20]. Deviant behaviors found in message transcripts were divided into three separate codes: "discussion of physical violence", "discussion of drugs", and "discussion of other deviance", such as stealing or skipping school, respectively.

The "discussion of violence" code referred to actual in-person fights. These messages sometimes discussed prior physical violence, or sometimes (explicitly or implicitly) encouraged future in-person violence. This physical violence sometimes involved both the participant and the other person in the conversation ("*I'll smack you*"), sometimes referenced physical violence between the participant and one or more persons not in the conversation ("*[name8]'s sister wants to fight me*"), and sometimes discussed remote physical violence (not involving either participant ("*lmao they wish to fight on Tuesday*"). Some conversations discussed violence that had already happened (12%), while others discussed plans for future fights (88%). It was not always clear whether online messages referenced literal or figurative violence. For example, "*I'm going to kill someone*"

could be a suggestion of physical violence, or it could just be an expression of frustration. The surrounding context of these messages was therefore important for coding accuracy.

All of the physical violence codes co-existed with one of the online conflict codes. The majority (69%) were double coded with "Indirect Online Conflict" and the remainder were double coded with "Direct Online Conflict" or "Direct Online Conflict of a Non-Serious or Sarcastic Nature".

Nearly all of the "discussion of alcohol and drug use" codes were about marijuana (82%), with a handful of codes referencing cocaine (9%) or alcohol (9%). Only 10% of these messages were double-coded with any type of online conflict, suggesting a minimal connection between online conflict and alcohol or drug use.

Most of the "other deviant behavior" messages involved a discussion of teens being somewhere they should not, often trying to sneak out of home to go see people or places, sneaking into the movies or a club, or skipping school. Similar to the alcohol and drug use codes, the minority of these messages (15%) co-existed with any type of online conflict, suggesting a minimal direct connection between these deviant behaviors and conflict.

3.4 Social media and online communication in Online Conflict

Because these data come from teens who previously self-reported having been involved in online drama/conflict, coders anticipated that social media itself would be an important, even frequent, topic referenced in both direct and indirect online conflict. Indeed, 24% of the total messages related to indirect conflict (1,159) referenced some aspect in which social media or online communication was a component of the conflict/drama. For example, conversations would discuss conflict associated with someone receiving a text message, posting a picture, or sharing someone else's social media post, such as: "*she's apparently talking to [name2] she posted it on her story*", or "*lmao look who texted me out of nowhere*". However, this was not the case with the direct online conflict codes, where only 5% of the total messages in which two participants expressed conflict with each other pertained to a conflict involving social media or online communication. The team found it difficult to determine whether these direct conflict codes would meet expert definitions for "cyberbullying."

4. Discussion

This qualitative study of online message interactions between adolescents provides novel insights into the real-time characteristics of conflict in adolescent social media use. It both outlines the patterns of conflict and provides insights into the ways in which adolescents' own messages could be used to better identify conflict and deliver preventive messages. A clear and correct interpretation of conflict is essential for understanding and addressing its role in cyberbullying, as well as for future efforts to develop automated identification of online conflict.

4.1. Expressions of Conflict in Adolescents' Online Messages

Despite the fact that all youth in this study reported involvement in cyberbullying, indirect online conflict (in which participants discussed or exacerbated a conflict with someone else; 46% of total codes) was more common than direct conflict (in which participants actively argued with or bullied each other; 20% of total codes). In our sample, youth were more likely to "gossip" or "vent" about their conflicts on social media, than they were use social media to initiate an argument with someone they were in direct contact with. This finding supports recent literature based on qualitative interviews with teens [2, 21], suggesting that the role of online media in adolescent conflict may be more nuanced than previously suspected.

Interestingly, adolescents' response to online conflict depended on circumstances. They often exacerbated or encouraged episodes of indirect online conflict; when involved in direct online conflicts, however, teens were more likely to de-escalate or resolve the situation. A next step for this research is to investigate whether, and how, the direct or indirect nature of online conflict is connected to the initiation, maintenance or response to cyberbullying. This information could directly inform future efforts to decrease both in-person and online conflict (and its consequences) among adolescents.

Discussions of dating and relationships appeared in a significant portion of instances of both direct and indirect online conflict. The majority (63%) of direct conflict appeared to occurr between teens who were currently or previously involved in a dating or sexual relationship, and a large portion (52%) of indirect conflict discussed conflict with sexual or dating partners. Here, too, further investigating how (or if) relationship issues transfer to cyberbullying will be a component of future analyses.

Interestingly, in this data set we did not find that direct or indirect conflict occurred often in messages about drugs, alcohol or truancy. This finding contradicts others' work [22] and requires further investigation.

The format of conversations was another surprisingly critical element of online conflict in this dataset. In this analysis, short indeterminate responses regularly served to end a fraught conversation. It is possible that such unclear responses reflected adolescents' not knowing how else to respond to conflict-ridden messages. Alternatively, indeterminate responses may be more useful as an online conversational style, in the midst of difficult communications, than either a non-response or a more direct response. Again, this hypothesis requires further testing and confirmation of conversational conventions.

4.2. The Role of Inference and Context in Coding

In addition to the revealing that the **direction**, **topics** and **form** of online messages are relevant, our research demonstrates how important the methodological process for qualitative coding is. Our coders' ability to identify tone and context of the messages was integral to their interpretation of the direct or indirect direction of the messages. Additionally, the qualitative coders attended to individual participants' writing characteristics, including their proclivities for sho rt or long messages.

For example, severity and intensity in a message was important for understanding both its meaning and the presence or direction of conflict within it. In order to avoid coding messages between friends where they merely called each other names in a meaningless, or even affectionate fashion, coders needed to understand the relationship context as well as the tone of words used in the messages. Inferences in tone were also important to drawing distinction between literal and hypothetical discussions of deviant behaviors, substantive and non-substantive third party insults, and substantive and non-substantive exacerbations of conflict. Finally, only messages that were deemed substantial were considered "Direct Online Conflict". Each of these distinctions indicate how important it is to have multiple coders, each of whom have a good grasp of adolescent language use, including referents and metaphors, and who work closely together to interpret the codes and to reconcile differential coding.

Another hurdle to analysis was an abundance of slang, euphemisms, and texting abbreviations in the lexicons of our participants. The research team's use of coders who were similar in age to the participants helped overcome some of these barriers, but even our young RAs occasionally encountered usages that they were unfamiliar with or unable to interpret. In some cases, repeated usage of unfamiliar words across multiple transcripts allowed RAs to infer their meaning in different contexts. For example, the word "souped" or "soupt" appeared in several transcripts, and our coders

were ultimately able to deduce that – in these transcripts– it referred to someone who is "pent up" and full of intense positive or negative emotion. For example: "*Lmao her sister got hyped on the phone and was like 'if anyone has anything to say to … you or [name7] I will be showing up at your school Tuesday and dragging you out of the classroom' and i was like 'yeah you're f*ing souped cuz that's not happening'*". Understanding the meaning of "souped" in these participants' messages became important to identifying instances of conflict. It would be difficult for a coder to identify the message "*now ya really gettin souped*" as an example of direct online conflict without knowing the meaning of the word "souped" in this context. Due to quick evolution and locally unique qualities of slang, existing slang dictionaries may not be able to identify the meaning of these new words [23]. Regarding "souped," for example, UrbanDictionary.com has multiple meanings ranging from "excited" to "angered" to "aroused" to getting beaten up [24]. Ultimately, automated sentiment analysis may be unable to make these deductions regarding novel usage, and thus miss, or misidentify, meanings that rely on an understanding of emergent, ephemeral teen vocabulary.

4.3. Implications for big data

Automated sentiment analysis has been advanced as a potential solution to the challenges of hand-coding of online messages. In theory, use of techniques such as VADER (Valence Aware Dictionary for sEntiment Reasoning) or more advanced natural language processing could allow analysis of much larger data sets, and could be used to inform delivery of just-in-time adaptive interventions [25-27]. However, the accuracy of such codes will only be determined through comparison with human inference. Whether automated coding can accurately engage with the subtle interpretations of content and tone that were key to coding these data is yet to be determined. For instance, techniques such as LIWC, which count occurrences of words, may miss the importance of figurative inferences or joking tone, and may miss the slang and its misspellings; as a result, it may be unable to make the types of distinctions that we believe are important to a thorough understanding of the data [13].

The importance of this question is epitomized by recent popular debates about online messaging services' "policing" of bullying and hateful rhetoric; if even large technology companies cannot regularly identify hateful language, it will likely be challenging to accurately identify adolescent conflict online [28, 29]. Other alternative techniques, such as letting adolescents "tag" incidents of online conflict, also have ethical challenges [30]. Ultimately, while automated sentiment analysis

has the advantage of minimizing time and human effort, and holds potential for real-time monitoring and intervention delivery, a mixed-method analysis in which hand review and qualitative coding of a substantive and representative sample that is subsequently analyzed with machine learning automated analysis may be the best approach to very large data sets.

4.4. Limitations

A number of limitations shaped the data and our analysis of them. Our extractor software was unable to extract Snapchat messages, which nearly all of our teens reported using (90%). Some teens opted to only submit a portion of their data, as they were not required to download messages from all of their potential outlets. Some teens may have, for example, only downloaded Facebook messages but still engaged in conversation in their text messages. The pseudonymization of links and removal of photographs and GIFs prevented coders from analyzing the full picture of the conversation. Additionally, of the thirty-four teens eligible to download their data, only twenty agreed (59%). The twenty teens in this sample were primarily minority youth, and living in a single state. The same methodology applied to this sample should be applied to a larger, more representative sample of the adolescent population for generalizability.

Consistent with our commitment to protect participants, we only coded data from persons from whom appropriate informed consent/assent was collected. This necessarily limits our potential understanding of these conversations. Our intention is to use these methods, in combination with LIWC and VADER machine learning analysis on larger data sets which either include consent from all parties or which are publicly available.

5. Conclusions and next steps

High quality hand-coding of conflict provides insight into the role of direct and indirect conflict and identified key topics in conflict as well as a nuanced understanding of relevant teen language use and linguistic forms. We believe this approach will also be essential to better inform automated analysis of prevalence and severity of cyberbullying and online harassment, as well as to just-in-time adaptive interventions, by providing real examples of direct and indirect online conflict between teens. Intervention techniques may be better informed by using real examples of cyberbullying in conjunction with reported statistics and teen preferences on solutions. Future research will entail developing more accurate automated measures of online conflict and cyber-bullying, as a necessary next step in the process of intervention development.

6. References

[1] Pew Research Center (2018). "Social Media Fact Sheet." from https://www.pewinternet.org/fact-sheet/social-media/.

[2] Englander, E. (2019). "Back to the Drawing Board With Cyberbullying." JAMA Pediatrics **173**(6): 513-514.

[3] Pew Research Center, September 2018, "A Majority of Teens Have Experienced Some Form of Cyberbullying"

[4] Aboujaoude, E., et al. (2015). "Cyberbullying: Review of an Old Problem Gone Viral." Journal of Adolescent Health **57**(1): 10-18.

[5] Brailovskaia, J., et al. (2018). "Cyberbullying, Positive Mental Health, and Suicide Ideation /Behavior." Psychiatry Research **267**: 240-242.

[6] Fahy, A.E., et al. (2016). "Longitudinal Associations Between Cyberbullying Involvement and Adolescent Mental Health." Journal of Adolescent Health **59**(5): 502-509.

[7] Moreno, M.A., et al. (2018). "Stakeholder Perceptions of Cyberbullying Cases: Application of the Uniform Definition of Bullying." Journal of Adolescent Health **62**(4): 444 – 449.

[8] Berne, S., et al. (2013). "Cyberbullying Assessment Instruments: A Systematic Review." Aggression and Violent Behavior **18**(2): 320-334.

[9] Pew Research Center (2018). "Teens, Social Media, and Technology, 2018."

[10] Van Ouytsel, J., et al. (2019). "Longitudinal Associations Between Sexting, Cyberbullying, and Bullying Among Adolescents: Cross-lagged Panel Analysis." Journal of Adolescence **73**: 36-41.

[11] Jones, N.M., et al. (2016). "Tweeting Negative Emotion: An Investigation of Twitter Data in the Aftermath of Violence on College Campuses." Psychological Methods **21**(4): 526-541.

[12] Doré B., et al. (2015) "Sadness Shifts to Anxiety Over Time and Distance from the National Tragedy in Newtown, Connecticut." Psychological Science **26**(4): 363-73.

[13] Choi, D., et al. (2014). "Tracing Trending Topics by Analyzing the Sentiment Status of Tweets." Computer Science and Information Systems **11**(1): 157-169.

[14] Rajkomar A., et al. (2019) "Machine Learning in Medicine." New England Journal of Medicine **380**(14):1347–58.

[15] Leung, A.N., et al. (2018). "You Are What You Read: The Belief Systems of Cyber-bystanders on Social Networking Sites." Frontiers in Psychology 9: 365.

[16] McHugh, M.C., et al. (2019). "OMG U #Cyberbully! An Exploration of Public Discourse About Cyberbullying on Twitter." Health Education & Behavior 46(1): 97-105.

[17] https://sochiatrist.cs.brown.edu/ [Accessed June 15th, 2019]

[18] QSR International Pty Ltd. NVivo qualitative data analysis Software. Version 10, 2012

[19] Pagoto, S. and Nebeker, C. (2019). "How Scientists Can Take the Lead in Establishing Ethical Practices for Social Media Research." Journal of the American Medical Informatics Association 26(4): 311-313.

[20] Ranney, M.L., et al. (2018). "Problem Behaviors and Psychological Distress Among Teens Seen in a National Sample of Emergency Departments." American Pediatrics 18(6): 650-654.

[21] Ranney, M.L., et al. "What Counts?: A Qualitative Study of Adolescents' Lived Experience with Online Victimization and Cyberbullying." Academic Pediatrics: under review

[22] Cunningham, R.M. et al. (2015). "Violent Reinjury and Mortality Among Youth Seeking Emergency Department Care for Assault-related Injury: A 2-year Prospective Cohort Study." JAMA Pediatrics 169(1): 63-70.

[23]Kaufman, L. (2013) "For the Word on the Streets, Courts Call Up an Online Witness" New York Times [Accessed June 15th, 2019] https://www.nytimes.com/2013/05/21/business/media/urban-dictionary-finds-a-place-in-the-courtroom.html?pagewanted=all&_r=0&gwh=11CB0CAF3EC6279269CC184D90BC4B79

[24] https://www.urbandictionary.com/ [Accessed June 15th, 2019]

[25] Bae, S., et al. (2017). "Mobile Phone Sensors and Supervised Machine Learning to Identify Alcohol Use Events in Young Adults: Implications for Just-in-time Adaptive Interventions." Addictive Behaviors 83: 42-47.

[26] Wahle, F., et al. (2016). "Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild." JMIR Mhealth Uhealth 4(3): e111.

[27] Hutto, C.J. & Gilbert, E. (2015). "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.

[28] Sharif, D. (2019). "We Got 'Zucked': Facebook Censors Speech That Calls Out Racism, Black Activists Charge." The Root [Accessed June 15th, 2019] https://www.theroot.com/we-got-zucked-facebook-censors-speech-that-calls-out-1834286507

[29] Lopez Tena, A. (2017). "Twitter Has Gone From Bastion of Free Speech to Global Censor." Business Insider [Accessed June 15th, 2019] https://www.businessinsider.com/twitter-has-gone-from-bastion-of-free-speech-to-global-censor-2017-6

[30] Lee, D. (2014). "Samaritans Pulls 'Suicide Watch' Radar App." BBC News [Accessed June 15th, 2019] https://www.bbc.com/news/technology-29962199