Abstract

This study explored social media representations of celebrities related to bullying by combining social science and computer science methods to analyze 1,280,151 bullying keyword tweets. Results show that Twitter users defined celebrities as bullies and victims, but also, and most frequently, as potential advocates against bullying. Some celebrities of all types fell into the advocate/confidant space and the sentiments within the tweets that defined celebrities as advocates were more likely to be positive versus negative. In sum, the results suggest that social media is a space that holds potential for promoting well-being around the topic of bullying.

Keywords: Social Media, Bullying, Twitter, Celebrity, Machine Learning, Sentiment Analysis

Celebrities Emerge as Advocates in Tweets About Bullying

When social media is considered in relation to bullying, it is often presented as a modern risk factor (David-Ferdon & Hertz, 2007; Hinduja & Patchin, 2008; Kowalski et al., 2014). Yet, the broader research conversation about social media in relation to psychosocial adjustment is more nuanced, with findings reflecting both benefits (Valkenburg, Peter, & Schoutem, 2006; Selfhout et al., 2009; Antheunis, Schouten, & Krahmer, 2014) and costs (O'Keeffe et al., 2011; Kross et al., 2013; Becker, Alzahabi, & Hopwood, 2013). The study described in this paper sought to describe the range of ways that Twitter users may reach out to celebrities about the topic of bullying with the expectation that users may do so in both positive and negative ways.

Social media is defined as websites and apps where users can contribute, retrieve, and/or explore content primarily generated or shared by themselves or fellow users (McGowan et al., 2012). Twitter is one such social media platform that is currently very popular with over 320 million monthly active users worldwide and 33% of American teens reporting using it (Twitter.com; Lenhart et al., 2015). Twitter is "a hybrid between a communication media and an online social network and hosts real time discussion of current topics of popular interest" in which users are limited to 140 characters (Lehmann, et al, 2012).

Bullying is commonly understood as an aggressive behavior that is directed from one person to another, usually within the context of an imbalanced power relationship and repeated over time (Olweus, 1993). Bullying extends beyond bullies and victims because it reflects a group process that incorporates all actors within a context. Other role players are defenders, who stand up for victims, and outsiders who are not directly involved in the bullying (Salmivalli, 1999). Just as bullying takes place in physical and social media environments, so too might all bullying roles exist within both environments (Bellmore, Calvin, Xu, & Zhu, 2015).

Social media is unique because it creates a dynamic, time-sensitive record of bullying interactions as well as providing a context for discussions of bullying events that may have taken place earlier online or offline. The ability to discuss bullying events on social media can enhance social support and help individuals cope (Gross, Juvonen, & Gable, 2002; Hampton, Goulet, Rainie, & Purcell, 2011; Hampton, Rainie, Lu, Shin, & Purcell, 2015). For example, when a fan messaged Taylor Swift about being bullied in school Taylor Swift responded with a genuine message of support. Following the response thousands of caring others reached out to encourage the bullied teen (Quinland, 2014). While social media may be seen my many primarily as an entertainment space, uses and gratification theory would suggest that users could also use these platforms for social utility and coping if that is their dominating need (Dominick, 1996; Christenson & Roberts, 1998; Neubaum, Rosner, Rosenthal-von der Putten, & Kramer, 2014).

Twitter is an example of a social media platform where the need for social utility and coping can be seen. Salmivalli (1999) identified four roles of individual's involved in bullying: bully, victim, reporter and defender. A recent analysis by Bellmore et al. (2015) found that the most common role represented on Twitter was the reporter, an individual who shares information about an episode. Results from the same study revealed that the number of social media posts made about a particular bullying event was associated with whether or not a celebrity was involved in the bullying event or mentioned the bullying event in a post. For example, following the September 2011 suicide of bulling victim Jamey Rodemeyer, Lady Gaga posted several tweets about the event. These tweets lead to giant spikes in tweets mentioning bullying (Xu, et al., 2012) and increased interest from Lady Gaga fans through social media to end bullying (Click, Lee, & Holladay, 2013). While Lady Gaga can be clearly defined as a celebrity independently of her Twitter following based on her six GRAMMY Awards (The GRAMMYs, 2016). However, there can also be users on Twitter who generate a large following, but may not be considered a celebrity. For example, Pattie Mallette, the mom of Justin Bieber has over three million followers but is not considered a celebrity. This study focuses on individuals who would be considered celebrities using Gabler's (2001) definition which states that celebrity is a function of being well known, capturing our interest and the interest of the media with their narrative, and depends on the idea of tangibility (i.e. actors, musicians, etc.) such as Lady Gaga, Oprah Winfrey, and Barack Obama who also have a presence on Twitter.

Based on the previously mentioned findings by Xu et al. (2012) and since people make meaning from celebrity culture as part of their daily lives, it appears that celebrities play a role in how bullying is reported and interpreted within our society (Marwick & boyd, 2011). There are several features of social media websites, such as Twitter, that afford the influence of celebrities. One such affordance within Twitter is that it is designed to be a public venue with most users indicating they have a public account and follow both users they know and do not know, which may facilitate a broader reach (Lady Gaga has 56.8 Million followers in March 2016) (Lenhart et al., 2015). Another affordance is that people can reach celebrities directly by mentioning them. Users are able to both view and respond to celebrities' posts (Marwick & boyd, 2011). Importantly, within the Twitter context, these interactions and posts are visible to other users, which allows any user to publically define a bullying role for a celebrity by reacting to what they see on Twitter.

Knowing how celebrities are defined by the public in relation to bullying is relevant at a practical level because many celebrities have taken the initiative to fight against bullying. Lady Gaga's Born This Way foundation works to empower youth through kindness and bravery. The Boston Vs. Bullies program is an anti-bullying initiative that connects professional athletes from

Boston with schools and youth groups with hopes that interacting with these athletes will resonate with youth about bullying (Dupont, 2012). These are only a few examples among many that draw on the connection that individuals feel to celebrities to mitigate the negative effects of bullying.

***The Present Study***

This study combines social science and machine learning methods to address three goals related to how individuals engage with celebrities on the topic of bullying. The analyses are exploratory in nature so no specific predictions were made. A large dataset (N = 25,370,824) of publically posted bullying tweets was collected between January 1, 2012 and December 31, 2012. This study focuses on a subset of those tweets that mention one of the 302 celebrity usernames mentioned the most in this dataset (N= 1,280,151). These 1,280,151 celebrity tweets are a small percentage of the total 25,370,824 bullying trace tweets collected in 2012. Yet, this subset of tweets is likely to be impactful and relevant to the discourse on bullying when considering the large number of followers each celebrity has and thus the large number of individuals who could potentially see their tweets. The 302 celebrities included in this study had between 1,767 to 84 million followers ($M$ = 5,790,198 followers, median = 1,090,000). By combining methods from two disciplines we were able to capitalize on real-world, real-time data about bullying, a topic of great social importance (APA, 2004). The ability to use "big data" provides promise for new insights into these important questions of our time (Shaw, 2014).

This study had three overall goals. Goal one was to identify the bullying roles that celebrities were assigned to by any users who tweeted using a bullying keyword within any single tweet in the year 2012. Both human annotators and machine learning algorithms were used to assign the bullying roles. It was expected that some similarities would be found between the bullying roles that were identified previously, Bully, Defender, Victim and Bystander (Xu, et al., 2012), but because the focus of this study was on the roles of celebrities who were mentioned within the tweet whereas the prior research focused on authors of tweets, it was also expected that some new, previously unidentified roles may emerge.

Goal two was to evaluate whether there was an association between bullying roles and "type" of celebrity and age of celebrity. Type of celebrity refers to the professional category they are known for (Turner, 2013), and includes actor, athlete, journalist, and religious figure among others. It was expected that some celebrity types would be more closely associated with their bullying roles on Twitter. For example, because they tend to promote kindness, we expected religious figures to be defined in supportive bullying roles like defenders. Whether age of celebrity was related to the bullying roles was investigated because bullying peaks in early adolescence (Nansel et al., 2001), social media use is ubiquitous among teens (Lenhart, 2015), and many bullying programs (including those mentioned earlier) target children, adolescents, and young adults. Thus, it was expected that many of the celebrities mentioned in the posts would also be young and that there may be differences in the roles assigned to young versus old celebrities.

Goal three was to measure the sentiment of the tweets. Fear, sadness, anger, and relief were found to be the most common emotions present within bullying posts in a prior study (Xu, Zhu, & Bellmore, 2012). It was expected that tweets that defined a celebrity as a bully or victim would reflect more negative sentiments such as anger and sadness whereas tweets that defined a celebrity as a defender would be more positive and reflect gratitude or admiration.

## Method

### *Data*

The data were derived from the public Twitter streaming Application Programing Interface (API) between the period of January 1, 2012 and December 31, 2012. The public streams available through the API are used by industrial developers and academic researchers (see https://dev.twitter.com/docs/streaming-apis). To address our goals, various aspects of the data present within the 140 characters in individual Twitter posts were coded and analyzed.

***Procedure***

Three steps were followed to obtain the data used in the analyses.

**Step 1: Identifying bullying tweets.** All posts that contained the keywords: “bully”, “bullied”, or “bullying” were collected from the public Twitter streaming API. Unless the maximum number of allowable tweets was hit in a given day (This only occurred on 12 days), all posts were received that satisfied this condition for 2012 were included as data. Then, re-tweets that did not contain any original content from the user who posted the tweet were removed. Finally, with all of the tweets that contained one of the bullying keywords a bullying traces classifier was used to narrow the dataset to just tweets that contained bullying traces. Based on previous research by Bellmore et al. (2015), bullying traces are defined as any mention of bullying within the context of a discrete episode. Recognizing that even though this method does capture the majority of tweets that include the word bullying, it does not capture all tweets related to bullying, which is why the tweets captured are referred to as “bullying traces”.

In total, we collected 25,370,824 tweets. From this collection, we extracted tweets that contained mentions (Twitter, 2016) by searching for @username mentions anywhere in the body of the tweet. For example, in the post, “@ladygaga You’re an inspiration about bullying!” @ladygaga is a user who was mentioned. Then all 5,049,121 unique usernames were sorted by the number of tweets in which they were mentioned to identify the top 1000 mentioned users in the collected bullying tweets.

**Step 2: Identifying celebrities from the top 1000 users.** The next step was to extract real celebrities from all users because this broader corpus of usernames includes individuals, groups or organizations, parody accounts, campaigns, or even spam bots. From the top 1000 users mentioned in the most number of tweets, human annotators identified whether the user was a celebrity or not based on Gabler's (2001) definition which states that celebrity is a function of being well known, capturing our interest and the interest of the media with their narrative, and depends on the idea of tangibility. All annotators were informed about the criteria for how to code the celebrities. The annotators began by viewing the public Twitter profile of the celebrity. If based on the profile, the account fit Gabler's criteria for a celebrity they were coded as a celebrity. If it was unclear whether the user was a celebrity based on the profile alone, then the user was searched for on famousbirthdays.com for clarification. We also specified that the celebrity account must be associated with a real individual, thus excluding "fake" profiles as well as those of groups or organizations. For example, the account for musician @LaurenJauregui, a member of the music group Fifth Harmony, would be classified as a celebrity but the account for @FifthHarmony would not. If a Twitter user simply included the name of a celebrity in their tweet without the @ symbol as part of the username or used the @ symbol with a misspelling of the celebrity, the mention would not be captured.

With these criteria, 302 among the top 1000 users within bullying posts were identified as celebrities. Two human annotators ($\kappa$ =.78 for all 302 users) identified the types of celebrities as Musicians (29%), Actors (22%), TV-personalities (20%), Authors (8%), Social-Media Celebrities (8%), Athletes (5%), Politicians (2%), and Other (6%), which was a group comprised of celebrity types that only appeared a few times including radio show hosts, entrepreneurs, fashion designers, and friends and family of celebrities. These categories were modified from Turner's (2013), previous research using celebrity twitter accounts (Marwick & boyd, 2011), and the professions that were specified on Twitter accounts. For celebrities who are famous in multiple categories, type was labeled based on the category the celebrity was most known for in 2012. Demi Lovato would be classified as a musician in 2012 because her music career was primary in that year even though she had previously acted (Grossberg, 2011). The age for each of the 302 most-mentioned celebrities on January 1, 2012 was obtained from famousbirthdays.com

**Step 3: Identifying bullying role of the 302 celebrities mentioned.** This step involved both human annotators and machine learning methods. Machine learning is a technique where computer systems learn from data (Bellmore, Calvin, Xu, & Zhu, 2015; Zhu & Goldberg, 2009). Machine learning has been used for practical applications such as detecting and tracking disease outbreaks by finding patterns in large datasets (Mitchell, 2006). Following the procedure of Xu (2015) to identify person-mention roles, first humans annotated individual bullying tweets that mentioned any of the 302 celebrity users mentioned to determine the bullying role of the celebrity within that tweet. Six-thousand posts were chosen randomly from the sample of 1,280,151 and annotated by coders so that a single bullying role was assigned to that celebrity user mentioned based on how they were mentioned in the tweet. In each tweet a celebrity was only mentioned once and the best fitting role was assigned. Based on prior experience annotating Twitter data that revealed that annotating several thousand tweets maximizes reliability of assignment to bullying roles when machine learning is used to classify tweets (Bellmore, Calvin, Xu, & Zhu, 2015). Drawn from prior research (Salmivalli, 1999), we expected four specific bullying roles to be assigned to the celebrities in the tweets. These roles were: Bully, Defender, Victim and Bystander (see Table 1 for sample tweets associated with these roles). A bully was defined as a celebrity being represented as, or called, a bully by the author of the tweet. A defender was a celebrity who was perceived as having defended a victim of bullying. A victim was defined as a celebrity being perceived as a target of a specific bullying incident or of bullying more generally in their past. A bystander was a celebrity who was called out for, or perceived as, not doing anything about an instance of bullying or bullying in general.

We also allowed for the identification of new roles that became evident through the annotating process (see Table 1 for sample tweets). A requested advocate was defined as a celebrity who was asked to do something (tweet or retweet about bullying, follow someone, read/watch a link, answer a question, etc.) regarding bullying as a societal issue in general or a specific bullying incidence. An advocate was defined as a celebrity portrayed as a general supporter of anti-bullying awareness. A confidant was defined as any celebrity user who was at the receiving end of personal information about bullying incidents by an author of a tweet. Tweets that were not entirely in English, those not connected to bullying even though a bullying keyword was included within the tweet, or those who did not include a role-player who fit a previously defined category were assigned to other. Most tweets mentioning celebrities fell clearly within the confines of one role. In the few instances where more than one role seemed plausible, the role that best described the celebrity's role in relation to the author within the context of the specific tweet was chosen. For example, in the tweet "*I've been bullied online everyday this year :'( my life has been miserable but seeing @ddlovato bringing attention to bullying makes me glad" @ddlovato* could be labeled as a confidant or an advocate but because the content of the tweet itself is confiding personal experiences *@ddlovato* would be best labeled as a confidant. With these parameters, two human coders (**κ** =.82 for 1000/6000 of the tweets) assigned a single bullying role to the user mentioned based on the text in each post.

Then a machine learning model was trained to automatically classify the celebrity role in the much larger collection of 1,280151 tweets. Specifically, the 6000 annotated tweets were used as training data to train a support vector machine. A support vector machine is a popular state-of-the-art classifier for text classification tasks (Zhu & Goldberg, 2009). Each tweet is represented by a vector of word counts known as a bag-of-words (BOW) representation with unigram (a single letter or word token) and bigram (two consecutive letter or word tokens) features (Zhu, Goldberg, Rabbat, & Nowak, 2008). The class label is the bullying role. The trained support vector machine is applied to the 1,280,151 tweets.

**Results**

***Goal 1: Identification of the bullying role players***

Using the human annotating method described in Step 3 of the methods section the top celebrity usernames mentioned in the six-thousand annotated tweets were assigned to one of the eight bulling roles (see table 3 for distribution). Next, the machine learning technique (also described in Step 3) was used to determine the distribution of the bullying roles across the subsample of tweets (N = 1,280,151; see Table 3 for final distribution). Then cross validation accuracy needed to be determined for using machine learning to assign bullying roles based on celebrity mentions. Before this was assessed several categories of roles were merged. The requested advocate role and the advocate role were merged in consideration of the similarities of their definitions and the limitations of machine learning. Also due to low frequency, the bystander role (N = 73) and the defender role (N = 12) were merged into the other role. After merging these categories, the machine learning method achieved 69.5% accuracy. A confusion matrix, a table used to illustrate the agreement between human annotating and the machine learning classifier for role assignment, can be found in Table 2. In the full dataset, using machine learning, the distribution of roles listed from most to least prevalent, was advocate (59.71%), bully (15.52%), confidant (9.27%), victim (8.60%), and other (6.90%).

***Goal 2: Identify whether celebrity type and celebrity age is associated with bullying role***

Principal Component Analysis (PCA) was conducted with the 302 most-mentioned celebrities within all 1,280,151 tweets to understand whether patterns of role involvement differed based on celebrity type and celebrity age. Principal Components Analysis is an orthogonal linear transformation that transforms the data to a new coordinate system such that the first principal component has the largest variance, and each following component has the highest variance possible given that it is orthogonal to the previous components. When using PCA to project high dimensional data into lower dimension space for visualization, it preserves the distances among points. If two points are relatively close in the original space, they will be close in the low dimension space. It is helpful to visualize and analyze points close to each other (Abdi & Williams, 2010; Pallant, 2013; Xu, 2015).

For this task each celebrity is represented by the union of text from tweets that mentioned them. The text is then converted into a standard bag-of-words feature vector. The pairwise distances are computed between all pairs of celebrity feature vectors and performed PCA to embed the celebrities in a two dimensional space. In Figure 1, each mark (both circles and squares) represents one of the 302 celebrities, and celebrities with similar feature vectors are closer to one another within the two dimensional space. Each panel shows one celebrity type, and all the panels have exactly the same ranges of values. The PCA did not use any category or age information.

The bullying roles embedded at different positions in the space were investigated, and the principal components analysis (PCA) revealed two components with eigenvalues over 1. The two components identified by PCA are related to the different roles. Principal Component 1 (x-axis) is related to more frequent assignment to advocate/confidant roles, and Principal Component 2 (y-axis) is related to more frequent assignment to the bully roles. The religious figure celebrity in the lower right corner was frequently mentioned as an advocate and confidant but never mentioned as a bully. At the other extreme, the political figure celebrity in the upper left corner was mentioned as a bully frequently and rarely mentioned as an advocate or confidant.

The PCA results also show that for some types of celebrity, the distribution of mentions across bullying roles in the same celebrity type tend to be similar. For example, many musicians were embedded in the lower right corner because they were confidants and advocates and not bullies. This was not true for all musicians as several also cluster in the upper left corner and many spots in between. Television celebrities and actor celebrities were also assigned to bullying roles across the spectrum. It is interesting to note that writers, journalists, and political figures fell into the confidant/advocate space as these are individuals with the influence and opportunity to shape the dialogue about bullying.

Examining celebrity age helped to further understand the bullying role patterns. Celebrities were grouped into one of two age categories—young and old. The mean age for the sample was 31.60 (SD = 13.81, median age = 29). Young celebrities were those who were 25 or younger and old celebrities were everyone else. The age 25 was chosen as it represents a period after both adolescence and emerging adulthood (Arnett, 2000). Within Figure 1, where young celebrities are represented with circles it is notable that in the musician and actor types where both confidant/advocates and bullies were represented, the young celebrities clustered in the confidant/advocate space and the older celebrities clustered in the bully space.

**Goal 3: Characteristics of Tweet Content Associated with Bullying Role**

Sentiment analysis was conducted to analyze the attitudes and valence of language within all 1,280,151 bullying tweets that contained a mentioned celebrity. Sentiment analysis has been widely studied in natural language processing and has been applied to many social science research domains (Calvin et al., 2015, Runge et al., 2013; Tausczik & Pennebaker, 2010). Sentiment algorithms learn a sentiment lexicon with weights, which are words and phrases commonly used to express positive or negative sentiments, from a coded list and/or existing corpus. The sentiment of a new document is then detected by aggregating the weights of the sentiment lexicon that appears in the new document with a set of rules designed from the earlier coded list. Recent algorithms also take informal spelling and emoticons into account to improve the performance on social media posts. In this study, SentiStrength software (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010) was used as it has been shown to have human-level accuracy for short social media posts in English (Thelwall, Buckley, & Paltoglou, 2011). This algorithm allowed us to produce the number of negative (e.g., "@CELEB_USER u r a MEAN SICK BULLY! :( ), neutral (e.g., "Bullying Needs To Stop @CELEB_USER"), and positive tweets (e.g., "@CELEB_USER You are such a role model for standing up against bullying :) thank you <3") associated with each bullying role by assigning a sentiment to each tweet associated with a celebrity username.

Once the number of tweets that was assigned to each sentiment was obtained, a chi-square analysis was conducted to determine the number of tweets for each sentiment in each role (see Table 4 for residuals). The overall chi-square test, $X^2(8)= 31247.33$, $p<.0001$, indicated that the distribution of tweets by sentiment differed significantly from the distribution that was expected. Based on the objectives of this question we decided to use a chi-squared test but we acknowledge it is sensitive to sample size. The standardized residuals within each cell show that the distribution of posts aligned with our expectations that tweets that included celebrities in a bullying or victim role were more likely to be negative and fewer were likely to be positive. The bully and victim groups differed for the neutral sentiments; tweets with bully roles were more likely to be neutral whereas tweets with victim roles were less likely to be neutral. Tweets with an advocate role were less likely to be negative and were more likely to be positive or neutral than expected. Tweets containing a confidant role were more likely to be negative and positive than expected. Tweets that named the other category were less likely than expected by chance to be negative or positive and more likely to be neutral.

## Discussion

This study explored social media representations of celebrities related to bullying by combining social science and computer science methods to analyze bullying keyword tweets. Results show that Twitter users defined celebrities as bullies and victims, but also, and most frequently, as potential advocates against bullying. Some celebrities of all types fell into the advocate/confidant space and the sentiments within the tweets that defined celebrities as advocates were more likely to be positive versus negative. In sum, the results suggest that social media is a space that holds potential for promoting well-being around the topic of bullying.

The bullying roles that were identified in this study overlap with four of the roles identified in previous research (Salmivalli, 1999). However, the distribution of roles was different from prior research that relied on Twitter posts about bullying (Bellmore et al., 2015). This study showed that bullies were more represented than were victims whereas that study showed the opposite. A reason for this difference may be the perspective from which the role-players were studied (as subjects versus authors of posts) as well as the fact that this study focused on celebrities. What these findings underscore is that it is impossible to understand prevalence rates without strong consideration of the sampling procedure used to derive the rates.

In addition to the four bullying roles that overlapped with previous research new bullying roles of advocates and confidants were identified in this study (Salmivalli, 1999). The confidant role was associated with more negative and positive sentiment than would be expected and was often associated with younger celebrities. This may be related to the main idea of uses and gratification theory that individuals will use social media in different ways depending on what need is dominating (Dominick, 1996; Christenson & Roberts, 1998; Neubaum, Rosner, Rosenthal-von der Putten, & Kramer, 2014). Thus, victims of bullying who are seeking a way to cope with the negative experience may label a celebrity as a confidant and therefore share their bullying story with them on Twitter as a way to cope.

Advocates were associated with positive sentiments by users. The advocate role is similar to a defender role but differs in that an advocate might only have been sought after and not yet acted as such. Advocates may be similar to allies in that they may publicize a general stance that is "anti-bullying" but not necessarily take measures to defend victims directly. Advocates may be unique to celebrities as they have the potential to influence the public through their visibility and status. In line with the uses and gratifications theory the new advocate role identified in this study may exists so that users can achieve their need for social utility and adding social capital to try and ameliorate bullying. Advocates may also be a better fit for social media because of its potential reach. Differentiating between defenders, advocates, and confidants, roles that may promote well-being, can help to direct individuals with influence (e.g., celebrities) into the role most effective for them as role models.

**Limitations and Directions for Future Work**

While thee methods allowed the exploration of novel questions about bullying within social media, there are limits to this approach.  These conclusions are tied to the sampling approach. This paper attempts to balanced theory and practicality yet, it is essential to underscore that the method relied on bully keywords and bullying traces to extract public tweets and then identified most-mentioned celebrities who had been named within the tweets. This limited the social media content to one specific context, bullying and narrowed the number of tweets analyzed. Additionally, while requiring @ username mentions limited the relevant posts that included celebrities, this requirement meant that each celebrity was notified of the tweets containing their names. This factor may be important in understanding how users and celebrities interact on social media especially when considering the number of followers each celebrity has and the potential to go beyond a parasocial interaction. It is also important to note that this study is descriptive and does not allow for understanding the processes through which users came to define the celebrities on the social media. It is not known whether celebrity behavior offline or online preceded or followed the posts that were studied. Relatedly, the calculations of the frequencies in which celebrities were assigned to roles, did not examine whether celebrities were mentioned few times by many social media users or many times by only a few. Each possibility is worthy of investigation.

**Implications for Translation**

The topic of bullying is one that lends itself naturally to translation because so many different types of stakeholders beyond scholars, including parents, schools and their staffs, government agencies, and professional organizations that focus on the psychosocial and physical welfare of individuals, are interested in eliminating bullying and its effects. Recognizing the group-process nature of bullying, many modern interventions have begun to emphasize the significance of social group norms and all role-players (Salmivalli, Kärnä, & Poskiparta, 2010). However, this work typically focuses on members of school and family communities and does not extend to media-based communities. As members of the broader social group in which bullying resides, the results of this study show that individuals do interact with celebrities through social media. Knowing that celebrities can act as advocates, defenders, and role-models in the fight against bullying, and knowing how large their reach can be through millions of social media followers indicates that celebrity influence should be considered in the design of novel interventions to combat bullying.

**References**

Abdi, A., & Williams, L.J. (2010). Principal component analysis. *WIREs Computational Statistics,* 2 .Retrieved from https://www.utdallas.edu/~herve/abdi-awPCA2010.pdf

American Psychological Association, Council of Representatives (2004). APA resolution on bullying among children and youth. Retrieved from <https://www.apa.org/about/policy/bullying.pdf>.

Antheunis, M.L., Schouten, A.P., & Krahmer, E. (2014). The role of social networking sites in early adolescents' social lives. *The Journal of Early Adolescence,* 1-24.

Arnett, J.J. (2000). Emerging Adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, 55(5), 469-480.

Becker, M.W., Alzahabi, R., & Hopwood, C.J. (2013). Media multitasking is associated with symptoms of depression and social anxiety. *Cyberpsychology, Behavior, and Social Networking, 16(2),* 132-135.

Bellmore, A., Calvin, A.J., Xu, J., Zhu, X. (2015). The five ws of bullying on twitter: Who, what, why, where, and when. *Computers in Human Behavior*, 44, 305-314.

Calvin, A., Bellmore, A., Xu, J., & Xhu, X. (2015). #bully: uses of hashtags in posts about bullying on twitter. *Journal of School Violence*, 14(1), 133-153.

Christenson, P.G., & Roberts, D.F. (1998). It's only rock & roll. Creskill, NJ: Hampton Press.

Click, M.A., Lee, H., & Holladay, H.W. (2013). Making monsters: Lady Gaga, fan identification, and social media. *Popular Music and Society*, 36(3), 360-379.

David-Ferdon, C., & Hertz, M.F. (2007). Electronic media, violence, and adolescents: An emerging public health problem. *Journal of Adolescent Health, 41(6),* S1-S5.

Dominick, J.R. (1996). *The dynamics of mass communication* (6th ed.). New York: McGraw-Hill

Companies.

Dupont, K.P. (2012). Sports Museum Launches ‘Boston vs. Bullies’ Initiative. The Boston Globe.

Gabler, N. (2001). *Toward a New Definition of Celebrity*. Los Angeles, CA: Normal Lear Center, University of Southern California Annenberg School for Communication.

Gross, E.F., Juvonen, J., & Gable, S.L. (2002). Internet use and well-being in adolescence. *Journal of Social Issues*, 58(1), 75-90.

Grossberg, J. (2011, April 11). Demi Lovato Quitting *Sonny With a Chance*. ENews.

Hampton, K., Goulet, L.S., Rainie, L., & Purcell, K. (2011). Social networking sites and our lives. Pew Research Center, Washington, DC.

Hampton, K., Rainie, L., Lu, W., Shin, I., & Purcell, K (2015). Social media and the cost of caring. Pew Research Center, Washington, DC.

Hinduja, S. & Patchin, J.W. (2008). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior*, 29(2), 129-156.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin, 140,* 1073–1137. http://dx.doi.org/10.1037/ a0035618.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D.S., Lin, N., et al. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS ONE, 8(8), e69841. doi:10.1371/journal.pone.0069841*

Lehman, J., Goncalves, B., Ramasco, J., & Cattuto, C. (2012). Dynamical classes of collective attention in twitter. In *Proceedings of the 21st international conference on World Wide Web (WWW),* 251–260. ACM.

Lenhart, A., et al. (2015). Teens, social media, and technology overview (Report 2015). Washington, DC: Retrieved from Pew Research Center website: http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf

Marwick, A., boyd, d. (2011). To see and be seen: Celebrity practice on twitter. *Convergence: The International Journal of Research into New Media Technologies, 17(2)*, 139-158.

McGowan, B.S., Wasko, M., Vartabedian, B.S., Miller, R.S., Freiherr, D.D., & Abdolrasulnia, M. (2012). Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. *Journal of Medical Internet Research*, 14(5).

Mitchell, T. M. (2006). The discipline of machine learning. Carnegie Mellon University, School of Computer Science, Machine Learning Department. Retrieved from <http://www-cgi.cs.cmu.edu/~tom/pubs/MachineLearningTR.pdf>.

Nansel, T. R., Overpeck, M., Pilla, R.S., Ruan, W.J., Simons- Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. The *Journal of the American Medical Association*, 285(16), 2094–2100.

Neubaum, G., Rosner, L., Rosenthal-von der Putten, A.M., & Kramer, N.C. (2014). Psychosocial functions of social media usage in a disaster situation: a multi-methodological approach. *Computers in Human Behavior*, 34, 28-38.

O'Keeffe, G.S., Clarke-Pearson, K., Council on Communications and Media. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), 800-804.

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Oxford, UK: Blackwell.

Pallant, J. (2013). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS.*

Berkshire, England: McGraw Hill.

Quinland, E. (2014, September 4). *Taylor Swift tells bullied teen fan to 'keep walking in the sunlight'*. Retrieved from http://www.today.com/popculture/taylor-swift-sends-touching-instagram-message-bullied-teen-fan-1D80126920

Runge, K. K., Yeo, S. K., Cacciatore, M. A., Scheufele, D. A., Brossard, D., Xenos, M. A., et al. (2013). Tweeting nano: How public discourses about nanotechnology develop in social media environments. *Journal of Nanoparticle Research*, *15*, 1381.

Salmivalli, C. (1999). Participant role approach to school bullying: Implications for interventions. Journal of Adolescence, 22, 453–459.

Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In S. R. Jimerson, S.M. Swearer, & D. L. Espelage (Eds.), Handbook of bullying in schools: An international perspective (pp. 441–454). New York, NY: Routledge.

Selfhout, M. H., Branje, S. J., Delsing, M., Ter Bogt, T. F., & Meeus, W. H. (2009). Different types of Internet use, depression, and social anxiety: The role of perceived friendship quality. *Journal of Adolescence, 32*, 819– 833.

Shaw, J. (2014, March). Why "Big Data" Is a Big Deal. Harvard Magazine, 30-75.

Tausczik, Y., & Pennebaker, J. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29,* 24-54.

The GRAMMYs (2016). *Lady Gaga*. Retrieved from http://www.grammy.com/artist/lady-gaga

Thelwall, M., Buckley, K., Paltoglou, G. (2011). Sentiment in twitter events. *Journal of the American Society for Information Science and Technology* 62(2):406–418.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*, 2544–2558.

Turner, G. (2013). *Understanding Celebrity*. SAGE Publications.

Twitter. (2016). *What are replies and mentions?* Retrieved from https://support.twitter.com/articles/14023

Valkenburg, P.M., Peter,J., & Schouten, A.P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology and Behavior*, 9, 584-590.

Xu, J.-M., Jun, K.-S., Zhu, X., & Bellmore, A. (2012). Learning from bullying traces in social media. In Proceedings of the conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 656– 666). Montreal, Quebec, Canada: Association for Computational Linguistics. Retrieved from <http://www.proceedings.com/15547.html>.

Xu, J.M., Zhu, X., & Bellmore, A. (2012). Fast learning for sentiment analysis on bullying. In *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining (WISDOM 12′;* pp. 1–6). New York, NY: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2346676

Xu, J. (2015). *Understanding and fighting bullying with machine learning* (unpublished doctoral dissertation). University of Wisconsin-Madison, Madison, WI.

Zhu, X., Goldberg, A.B., Rabbat, M., & Nowak, R. (2008). Learning bigrams from unigrams. In Proceedings of the conference of the North American chapter of the association for computational linguistics: human language technologies. Columbus, Quebec, Canada: Association for Computational Linguistics. Retrieved from <http://pages.cs.wisc.edu/~jerryzhu/pub/bigram.pdf>.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 3,* 1-130.

*Table 1.*
*Examples of Bullying Roles Within Tweets*

| Bullying Role | Example Tweets |
|---|---|
| Bully | This b*tch @ CELEB_USER is a f*ckin Bully!! B*tch grow up!<br>@ CELEB_USER, you are the biggest bully! Why are you tormenting us |
| Defender | @ CELEB_USER just hugged a bullied boy & told him there's absolutely nothing wrong with him<br>@ CELEB_USER I saw your video about what you did for @VICTIM. How you stood up to bullying and let me just say: youre much respected! #life |
| Victim | I feel sorry for whoever bullied my girl @ CELEB_USER<br>People have got to stop bullying @ CELEB_USER!! We get that you're jealous. Why bullying & making up rumors about her? Lifeless, much? |
| Bystander | @ CELEB_USER I am very disappointed in U as a woman allowing her to BULLY @Victim, all for TV ratings how sad and what a bad example<br>@ CELEB_USER why don't you stop your fans from bullying everyone online?!? |
| Other | @CELEB_USER are you ready for your fight with bully ray? #WWE<br>@ CELEB_USER - New slow jam! Bedroom Bully (OFFICIAL VIDEO)!! |
| Requested Advocate | @ CELEB_USER Will u help me and my friends with our anti-bullying campaign? Please get back :)<br>@ CELEB_USER its bullying awareness week! Please tweet a picture of "LOVE" on your wrist it would mean a lot. |
| Advocate | @ CELEB_USER says bullying is really not cool. It will never be cool to ruin someone's life.<br>I'm impressed that @ CELEB_USER watched the bully movie and is sharing how bad bullying is |
| Confidant | @ CELEB_USER I've been bullied through public school and have been depressed bc of it. #StayStrong<br>@ CELEB_USER i have been bullied by fake friends & your song, "Mean" really helped me I <3 u!!! |

*Table 2.*
*Confusion matrix of person-mention roles using the 6000 annotated tweets*

| Annotated as | Predicted as | | | | |
|---|---|---|---|---|---|
| | Advocate/Requested Advocate | Bully | Bystander, Defender, Other | Confidant | Victim |
| Advocate | 2116 | 101 | 73 | 158 | 47 |
| Bully | 209 | 521 | 68 | 32 | 44 |
| Other | 186 | 126 | 871 | 26 | 19 |
| Confidant | 271 | 48 | 20 | 499 | 26 |
| Victim | 151 | 131 | 32 | 60 | 165 |

*Note. The diagonal indicates agreement between assignment based on human coders and machine learning. The accuracy of assignment was 69.5%.*

*Table 3.*
*Distribution of bullying roles in bullying tweets that contained top celebrity usernames in 2012*

| | Initial Bullying Role Labels (6000 tweets; hand annotating) | | Final Bullying Role Labels (6000 tweets; hand annotating) | | Final Bullying Role Labels (all 2012 tweets; machine learning) | |
|---|---|---|---|---|---|---|
| | n | % | n | % | n | % |
| Bully | 874 | *14.57* | 874 | *14.57* | 192023 | *15.52* |
| Defender | 12 | *.02* | merged with Other | | merged with Other | |
| Victim | 539 | *8.98* | 539 | *8.98* | 115213 | *8.60* |
| Other | 1141 | *19.02* | 1226 | *20.43* | 89610 | *6.90* |
| Requested Advocate | 1256 | *20.93* | merged with Advocate | | merged with Advocate | |
| Advocate | 1241 | *20.68* | 2497 | *41.62* | 768091 | *59.71* |
| Confidant | 864 | *14.4* | 864 | *14.4* | 115214 | *9.27* |
| Bystander | 73 | *1.22* | merged with Other | | merged with Other | |

*Note. The requested advocate role and the advocate role were merged in consideration of the similarities of their definitions and the limitations of machine learning. Also due to low frequency, the bystander role and the defender role were merged into the other role.*

*Table 4.*
*Number of tweets per sentiment type for each celebrity bullying role*

| Role | Sentiment | | |
|---|---|---|---|
| | Number of Negative Tweets (%) | Number of Neutral Tweets (%) | Number of Positive Tweets (%) |
| Bully | 41944 (21.1%) | 145109 (73.0%) | 11626 (5.9%) |
| Z | 17.16* | 5.95* | -42.59* |
| Victim | 20293 (23.0%) | 62201 (70.4%) | 5825 (6.6%) |
| Z | 21.12* | -3.22* | -22.46* |
| Advocate | 143209 (18.7%) | 546817 (71.6%) | 74294 (9.7%) |
| Z | -21.15* | 2.06* | 25.81* |
| Confidant | 35300 (32.0%) | 59462 (54.0%) | 15369 (14.0%) |
| Z | 91.25* | -68.17* | 57.06* |
| Other | 12105 (10.2%) | 100867 (84.9%) | 5730 (4.9%) |
| Z | -74.43* | 55.6* | -46.52* |

$X^2(8)= 31247.33$, $p<.0001$*

*Note. Numbers in parentheses are percentages of tweet sentiment type computed separately for each bullying role. Cell Z scores that exceed or fall below the critical value of ±1.96 are significant at p < .05. Significance is indicated with a *.*

*Figure 1.*
*Principal Components Analysis of Celebrity Type, Bullying Role, & Age*

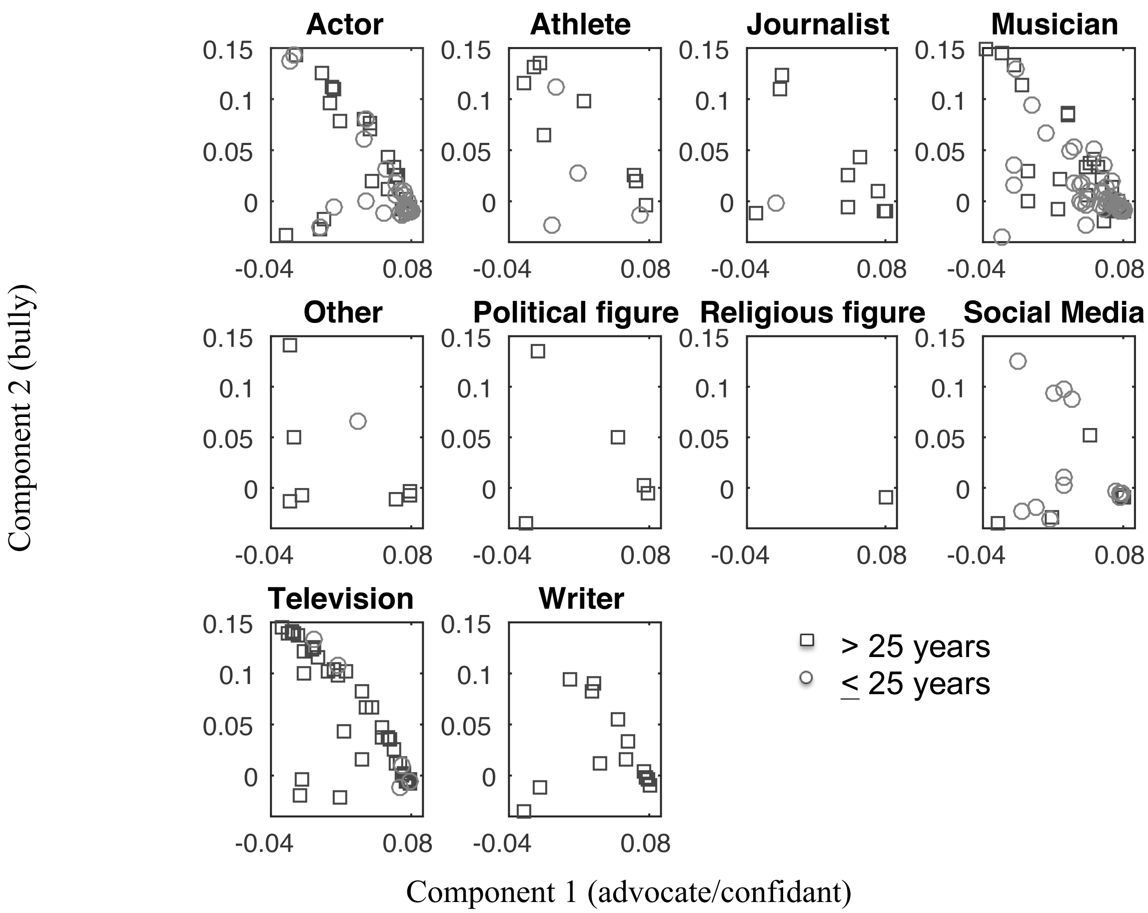


*Figure 1.* Principal Components Analysis (PCA) of bullying role. This is a representation of the 302 celebrities mentioned within each of the 10 types of celebrity categories along Principal Component 1 (x-axis), which is related to the advocate and confidant roles and Principal Component 2 (y-axis), which is related to the bully role. Age is represented by the shape of the point with squares representing celebrities older than 25 and circles representing celebrities age 25 or younger.