Harnessing big data for social justice: An exploration of violence against women-related conversations on Twitter

Jia Xue1 | Kathy Macropol2 | Yanxia Jia2 | Tingshao Zhu3 | Richard J. Gelles4

1Factor-Inwentash Faculty of Social Work & Faculty of Information, University of Toronto, Toronto, Ontario, Canada
2Department of Computer Science & Mathematics, Arcadia University, Glenside, Pennsylvania
3Institute of Psychology, Chinese Academy of Sciences, Beijing, China
4School of Social Policy & Practice, University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

Social media, including Facebook, Twitter, Instagram, and Snapchat offer new means of communication, networking, and community building. Social media are mechanisms by which millions of people spread, share, and exchange information—ranging from sports and politics, to health and illness. Twitter users, in particular, also build communities on topics of interest. This paper examines Twitter content to examine the extent to which the topic of “violence against women” is posted and disseminated. We know very little about the intersection of social media and the social problem of “violence against women.” Is Twitter being used to advance advocacy efforts, seek information and assistance, and/or build communities among advocates and/or victims? First, we need to know whether and to what degree Twitter contains posts on the topic of violence against women (VAW). This paper offers the first exploration into Twitter postings related to the topic of VAW. We collected 2.5 million tweets posted from 2007 through 2015. We then classified postings (referred to as “Tweets”). We compared posting on the topic of VAW to posting related to nine topics: politics, entertainment, sports, women, relationships, fashion, kids, school, and food. We found a small but actively engaged community that Tweets about VAW. Twitter users who post on the topic of VAW reply to one another in each conversation thread, but they rarely disseminate conversations through Retweeting. Our exploratory findings suggest that more might be learned from future studies that investigate the use of social media on the topic of VAW.

KEYWORDS

conversation thread, Twitter, violence against women

1 | INTRODUCTION

New communication technologies increase the possibilities for how people send and receive information (Westerman, Spence, & Van Der Heide, 2014). Social media are technologies that facilitate information distribution as well as conversations. Social media refer broadly to the set of online media tools, such as social networking sites (e.g., Facebook) and microblogs (e.g., Twitter) that foster collaborative participation and engagement for end users (Westerman et al., 2014). Social media are mechanisms by which people seek and share information and also discuss personal concerns (Fox, 2011). In addition, social media are also tools that people use to create online communities.

Twitter is an “imagined community” (Gruzd, Wellman, & Takhteyev, 2011) for learning, meetings (Borau, Ullrich, Feng, & Shen, 2009), scientific conferences (Ebner & Reinhardt, 2009), forming opinions, answering questions (Dunlap & Lowenthal, 2009), and for public health information and data (Signorini, Segre, & Polgreen, 2011). Researchers find that individual users seek out health-related information on Twitter because users consider Twitter a rich environment for...
spreading health information, exchanging medical information, communicating health information, promoting positive behaviors, and seeking advice (Greene, Choudhry, Kilabuk, & Shrank, 2011; Paul & Dredze, 2011; Scanfeld, Scanfeld, & Larson, 2010). In the area of public health, Twitter users create communities in areas such as influenza, obesity, insomnia, antibiotics, depression, and cancer. In addition, researchers find value in examining the content of Twitter postings. When Twitter users Tweet about their personal health information, millions of such messages can reveal trends about certain health problems in a region or country (Lampos & Cristianini, 2010; Paul & Dredze, 2011). For instance, Tweets have been used to determine the extent of the H1N1 outbreak (Chew & Eysenbach, 2010). Culotta (2010) found that monitoring influenza-related Tweets provide cost-effective and quick health status surveillance.

Violence against women (VAW) is global public health problem. The World Health Organization (WHO) reports that 35% of women worldwide have experienced intimate partner violence or nonpartner sexual violence (WHO, 2015). For decades, VAW scholars have collected data about the nature of VAW from interviews with victims, surveys that employ in-person interviews or questionnaires, and by analyzing official data, such as crime statistics or medical records (Gelles, 2000).

Given the widespread use of social media, the question arises—are there important data to be obtained by analyzing social media postings, such as Tweets on Twitter? Do social media in general, and Twitter in particular, provide a new window into the nature of VAW? We know that 53% of 261 agencies serving abused and assaulted women have social media links on their websites, and 23% of the agencies use Twitter for advocacy (Sorenson, Shi, Zhang, & Xue, 2014). Do victims of partner violence post on Twitter, seek information, and/or attempt to build communities? In addition, given that there are Tweets on the topic of violence toward women, how do the number and distribution of Tweets on violence toward women compare to the number and dissemination of tweets on other topics?

Given the importance of the social problem of VAW and the growing and rather substantial use of Twitter, there is a reasonable argument for exploring the intersection of VAW and Twitter. To date, there is limited research that examines conversations posted on Twitter by those who are interested in the topic of VAW. Twitter could be a resource for victims and advocates (and perhaps offenders) of VAW, but it remains unknown and untracked how Twitter is actually employed as a resource.

The purpose of this study is to explore VAW conversations on Twitter in order to better understand Twitter’s possible contribution to knowledge building, advocacy, service provision, and public policy.

2 | LITERATURE REVIEW

2.1 | Twitter

As one of the most widely used social media platforms, Twitter serves as a public viewing platform for gathering information, disseminating messages, and generating large amounts of contents. Twitter has over 288 million monthly active users, with over 500 million Tweets sent each day (About Twitter, 2016). Twitter users can post 140-character messages under any topic, known as “Tweets.” While each Tweet message consists of only up to 140 characters, the aggregate of millions of Tweets provides unprecedented amounts of information (Padmanabhan et al., 2013). In addition to the microblogging function, Twitter users can reply or Retweet (repost) others’ Tweets. Open, public Tweets on Twitter can be read by anyone, as long as the individual organization’s account is set as “public” (Marwick, 2011). The publicly available Tweets are spread further through Twitter’s functions of Retweet, allowing for nearly an infinite number of users to read the original Tweets.

2.2 | Conversation threads

A group of Twitter messages composes a Twitter conversation, including the original Tweet and its succeeding replies and Retweets, arranged in a hierarchical structure, with the original Tweet being at the top. Research demonstrates that 12.5% messages on Twitter are parts of conversations (Mischaud, 2007).

A list of Tweets and the subsequent replies/Retweets compose what is referred to as a “conversation thread.” On Twitter, all conversation threads begin with a message or a post. “Reply” refers to the responses to a Tweet that appears in sequence below each Tweet. The function of “reply” indicates the degree and length of responses for Tweet conversations on Twitter. “Retweets” refer to reposts of others’ Tweets, which can be used as a measure to estimate the information diffusion in Twitter spaces. Thus, the number of Retweets, number of responses, and length of discussions are indicators of the size of conversations presentation on Twitter.

Within each conversation thread, Tweets are related to each other and related to the same topic compared to other Tweets outside the particular conversation thread (Cao et al., 2012). Understanding conversational threads is helpful to learn about specific topics on Twitter because conversation threads vary by topic. For instance, different conversation topics attract different numbers of replies and repliers, and thus the length of conversation threads is different across different topics. Twitter users can broadcast their opinions to a potential population of readers through conversation threads (Himelboim, Gleave, & Smith, 2009). Some Twitter users do not create posts to start their own conversation threads because of their concerns for privacy of social media or the sensitivity of the topics (Meeder, Tam, Kelley, & Cranor, 2010). However, users who wish to maintain privacy do contribute many messages, but to threads of others’ conversations. Therefore, conversation patterns in social spaces like Twitter are specialized according to various topics.

3 | AIM OF THE STUDY

Social media are changing the way in which people communicate about health topics, seek help, engage in collaborations, and built
Our study aims to explore whether and how VAW-related conversations are presented on Twitter. In order to achieve this goal, we collected and analyzed the conversation structure from a variety of different topics (see methodology section for a discussion of the selection process for the other topics). Our study is the first one to explore social media conversations on the topic of VAW on Twitter. By examining and analyzing Tweets, Retweets, and Twitter conversations, we may learn whether there are any conversations about VAW on Twitter, and the breadth and depth of those conversations compared to other topics. This initial exploration does not examine the content of conversations, as our first goal was to determine whether the topic of VAW actually appears on Twitter.

3.1 | Research questions

Overall, our main question is whether those interested or impacted by VAW are posting on Twitter and whether such posts create a breadth and depth of community through conversation threads. We examined volume and the lengths of Twitter conversations as a first pass at understanding how the topic of VAW is presented in the 140 characters world of Twitter. We will compare the volume and length of Twitter conversations to other "trending" topics on Twitter. Our specific questions are:

RQ1:

1. How many Tweets per conversation thread are on the topic of VAW?
2. Are there any differences regarding numbers of Tweets across different topics and subject matter?

RQ2:

1. How many Retweets are created on the topic of VAW?
2. Are there any differences regarding numbers of Retweets across different topics and subject matter?

RQ3:

1. How many users are there per conversation thread on the topic of VAW?
2. Are there any differences regarding numbers of users across different topics and subject matter?

RQ4:

1. What is the conversational thread depth on the topic of VAW?
2. Are there any differences regarding conversational thread depth across different topics and subject matter?

RQ5:

1. What is the conversation thread degree on the topic of VAW?
2. Are there any differences regarding conversation thread degree across different topics and subject matter?

4 | METHODS

4.1 | Data collection and sample

We collected a dataset consisting of more than 2.5 million Tweets posted across an 8-year period, from 2007 to 2015, by using the Twitter’s Application Program Interface (API). We gathered the Tweets using a Breadth-First Search (BFS)-based crawling technique (Hauffa, Koster, Hartl, Kollhofer, & Groh, 2016; Kwak, Lee, Park, & Moon, 2010; Macropol, Bogdanov, Singh, Petzold, & Yan, 2013; Russell, 2013). The technique, which is like a snowball sampling technique, starts with an initial small set of randomly chosen users. We obtained all available tweets for each user and preceded to his/her followers, and then from each of the followers, continuing to repeat the process. The computer science literature reports that social network graphs, such as Facebook and Twitter, exhibit small-world network characteristics (Bakhshandeh, Samadi, Azimifar, & Schaeffer, 2011; Ugander, Karrer, Backstrom, & Marlow, 2011). According to Ugander et al. (2011), the average distance between pairs of Facebook users was 4.7. This means that on average, just five other people separate any two people on Facebook. A similar conclusion was drawn about the Twitter network (Bakhshandeh et al., 2011). Therefore, by starting from a small set of randomly chosen Twitter users, the BFS-based crawling is able to provide a reasonable snapshot of a subsection of the Twitter network, and therefore it is a commonly used technique to crawl Twitter data.

Figure 1 illustrates a Twitter follower graph (Gabielkov & Legout, 2012) generated by applying the BFS algorithm. In the graph, each square represents a Twitter user account, and the size of the graph is only for illustration purpose. As a result of the crawling, we were able to obtain 2.5 million tweets posted by a total of 37,126 users.

In addition, if any given post has a retweet or reply, we collected the metadata relating to “reply-to” and “retweet” structure for each Tweet. Users’ retweets frequently are broadcast messages, passing along information, news, and messages to large numbers of Twitter followers. Replies, on the other hand, are more private, allowing targeted conversations to occur between users—each reply is made in response to a specific single parent Tweet.

4.2 | Comparison topic selection

In the study, we created eight general classes of topics along with the topic of VAW (Table 1). The eight topics were selected based on previous research on Twitter trending topics, such as food, fashion, and sports (Lee et al., 2011). We wanted the comparison topics to have breadth—many postings—and depth, numerous replies, and retweets. We presumed that some of popular trending topics such as politics
(Bastos, Puschmann, & Travitzki, 2013; Lee et al., 2011), entertainment (Bastos et al., 2013), education/school, and family life (e.g., relationship and kids; Zhao & Jiang, 2011), overlap with the issue of VAW.

4.3 | Topic classification

For topic classification, we classified all Tweets into the nine general topic areas, presented in Table 1: VAW and eight comparison topics.

Twitter "hashtags" are a means of grouping large numbers of Tweets (Bruns & Burgess, 2011; She & Chen, 2014). As a function of Twitter, hashtags represent short keywords, prefixed with the hash symbol "#" (e.g., #obama, #rape, #mental health). Hashtags are not created by Twitter but are user-defined concepts in a free style (Chang & Iyer, 2012). Users who use the same hashtags to define certain topics have their Tweets posted under the hashtag together with all other Tweets that employ the hashtag (Bruns & Burgess, 2011). Hashtags, as clickable links, facilitate the search of Tweets containing the same hashtags and are employed by Twitter users to create and follow conversation threads (Kwak et al., 2010; Suh, Hong, Pirolli, & Chi, 2010).

We utilized a hashtag-based approach to classify Tweets into the nine categories, which is similar to the methods used in previous studies that examine Twitter content (Macropol et al., 2013; Wang, Wei, Liu, Zhou, & Zhang, 2011). From the initial collected Tweets, we built a Naive Bayes classifier (Rish, 2001) using the word frequencies within each Tweet as features for next-step topic classification. We used a Naive Bayes classifier to classify Tweets' threads into the following nine topic categories: politics, entertainment, sports, VAW, relationships, fashion, kids, school, and food. For example, we classified Tweets containing hashtags such as "#politics," "#obama," "#democrat," and "#republican" into the topic of Politics.

All data are divided into a training set and testing set. For each topic, a list of relevant hashtags was manually generated to build a training set used to train the Naive Bayes classifier to classify Tweets into different topics. The class of topic of each thread in the training set is labeled according to its hashtags. Using word frequencies of a thread as attributes, the probabilities of each thread being in a certain category are calculated. The resulting probabilities are then used to place threads with high probabilities into their corresponding category(s). The classifier was then used to discover further Tweets for each category. "Gold Standard Tweets" (randomly chosen and manually categorized) were used to verify the classifier, with an overall precision of over 83% obtained.

4.4 | Twitter conversation structure

A Twitter conversation is also known as a "thread," "Twitter thread," or "conversation thread," terms that are used interchangeably. In order to answer our research questions, we measured the conversation structure using several concepts: (a) the total number of Tweets per conversation thread; (b) the total number of users per conversation; (c) the conversation thread depth; (d) the conversation thread degree; and (e) the number of Tweets of all the conversations under each topic.

We offer Figure 2 as an example of Twitter conversation structure. Figure 2 shows an example of Twitter conversation thread consisting of five retweets and seven users.

We show all terms and their explanations here.

Node: Within each Twitter conversation structure, each tweet, reply, or retweet is defined as a "node." Examples: For example, in Figure 2, "User1: Tweet1," "User2: reply1," "User3: Retweet1" are all "nodes."
Table 1: Dataset—topic classifications by hashtags

<table>
<thead>
<tr>
<th>Topic</th>
<th>Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violence against women</td>
<td>#rape, #domesticviolence, #domestic_violence,</td>
</tr>
<tr>
<td></td>
<td>#womanabuse, #femaleabuse, #wifeabuse, #spousalabuse,</td>
</tr>
<tr>
<td></td>
<td>#womanviolence, #womenviolence, #femaleviolence, #violenceagainstwomen,</td>
</tr>
<tr>
<td></td>
<td>#girlfriendabuse, #daterape, #rapeculture, #raped,</td>
</tr>
<tr>
<td></td>
<td>#indiasdaughter, #sexualviolence</td>
</tr>
<tr>
<td>Politics</td>
<td>#politics, #obama, #obama2012, #obama2008, #obamaforpresident, #obama,</td>
</tr>
<tr>
<td></td>
<td>#romney, #republican, #democrat, #libertarian,</td>
</tr>
<tr>
<td></td>
<td>#barackobama, #tcot, #whitehouse, #voting, #senate, #congress, #teaparty,</td>
</tr>
<tr>
<td></td>
<td>#vote2014, #vote2013, #vote2012, #vote2011, #vote2010, #vote2009, #vote2008, #republicans, #democrats, #impeachobama</td>
</tr>
<tr>
<td>Entertainment</td>
<td>#entertainment, #movies, #blockbuster, #television, #tv, #music, #singer,</td>
</tr>
<tr>
<td></td>
<td>#theater, #movie, #hbo, #gameofthrones, #thejinx, #filmtailers,</td>
</tr>
<tr>
<td></td>
<td>#moviereview, #filmpreview, #seasonfinale, #arrow, #bigbangtheory,</td>
</tr>
<tr>
<td></td>
<td>#supernatural, #movietheater, #imax, #cinema, #concert</td>
</tr>
<tr>
<td>Sports</td>
<td>#sport, #baseball, #football, #soccer, #nba, #fifa, #worldcup,</td>
</tr>
<tr>
<td></td>
<td>#nfl, #nhl, #superbowl, #nflplayoffs, #patriots, #49ers, #worldcup,</td>
</tr>
<tr>
<td></td>
<td>#worldcup2014, #worldcup2013, #worldcup2012, #worldcup2011, #worldcup2010, #worldcup2009, #worldcup2008, #marchmadness, #wwwe, #sochi2014, #usmnt, #worldcupfinal, #nbafinals, #basketball</td>
</tr>
<tr>
<td>Relationships</td>
<td>#girlfriend, #relationship, #marriage, #breakup, #singles, #relationshipgoals, #datingcoach, #single, #breakupandmove, #singlelife, #singleagain, #datingadvice, #boyfriendadvice, #girlfriendadvice, #relationshipadvice, #relationshipspendbecause, #firstdates, #lovemygirlfriend, #lovemyfriend</td>
</tr>
<tr>
<td>Fashion</td>
<td>#fashion, #clothes, #outfit, #shoes, #style, #womenswear, #menswear,</td>
</tr>
<tr>
<td></td>
<td>#accessories, #makeup, #purse, #beautytip, #clothing, #jewelry</td>
</tr>
<tr>
<td>Kids</td>
<td>#kids, #children, #parenting, #babies, #parentingtips, #forkids, #infants,</td>
</tr>
<tr>
<td></td>
<td>#babycare, #babyshower, #diaper, #toddlers, #pediatrics</td>
</tr>
<tr>
<td>School</td>
<td>#school, #college, #university, #highschool, #elementaryschool, #classes,</td>
</tr>
<tr>
<td></td>
<td>#grades, #finals, #collegehighschool, #school, #teachers, #students, #recess, #education, #sat, #finalsweek</td>
</tr>
<tr>
<td>Food</td>
<td>#food, #cooking, #recipe, #recipes, #homecooking, #healthfood, #diet, #culinary</td>
</tr>
</tbody>
</table>

Link (edge): In a Twitter conversation structure, the arrow connecting two nodes is defined as a link (edge). The link represents the “reply-to” or “retweet” relationship between nodes. For example, in Figure 2, the node “User1: Tweet1” has three incoming links and one outgoing link.

Node degree: In a Twitter conversation structure, the number of links pointing to a node is defined as the node degree. For example, in Figure 2, the node “User1: Tweet1” has a degree of 3.

Thread degree: In a Twitter conversation structure, the thread degree is defined as largest node degree within the structure (i.e., the most replies a single Tweet obtained). For example, in Figure 2, the thread degree of the conversation in Figure 2 is 3.

Average thread degree: The average thread degree refers to as the average of thread degree in all Twitter conversation threads under the same topic.

Thread depth/height: In a Twitter conversation structure, the thread depth is defined as the number of links on the longest path from the top (root) node of the thread to the node at the bottom level. For example, in Figure 2, the thread depth starts with the root at level 0, and in this case has a depth of 4.

Average thread depth: The average thread depth refers to as the average of thread depth in all conversation threads under the same topic.

5 | RESULTS

5.1 | Descriptive statistics

First and foremost, we found that there are postings on Twitter on the topic of VAW. Table 2 presents the total number of Tweets, total number of threads, and total number of users under each of the nine topics. The topic “politics” has the highest number of tweets, threads, and users compared to all other topics, followed by the topics of “entertainment” and “sports.” The topic “violence against women” ranked the fourth regarding the total number of tweets. On the other hand, “violence against women” involves the smallest number of users (n = 908) and total conversation threads (n = 730).

In summary, there are a large number of Tweets in the category of “violence against women,” generated by a small number of Twitter users. The size of the conversation threads is the smallest compared to other trending topics.

5.2 | Conversation structure

Classified conversation threads were processed, and for each topic, we collected and analyzed threads in terms of four measures of conversational structure: (a) the number of tweets per conversation thread; (b) the number of users per conversation thread; (c) the average thread depth per conversation thread; and (c) the average thread degree per conversation thread.

Table 3 presents the results of above four measures of the conversation structure for all nine topics, including average tweets per thread, average users per thread, average thread depth, and average thread degree. Results reveal that the topic of “violence against women” is the most-engaged topic compared to other topics, indicating that people are posting and replying to Tweets.
Figures 3 and 4 show the average number of tweets and the average number of users per conversation thread. For both, the topic of "violence against women" has higher values. The results indicate increased tweeting and responding activities on the topic of VAW. Even though the total numbers of users involved in conversations on the topic of "violence against women" is the smallest compared to the number of users posting on the other eight topics, the average number of users per conversation thread on "violence against women" is greater—albeit, the difference is small—a little more than one person. The few Twitter users who are talking about VAW on Twitter, post twice as many tweets per thread compared to posters on the other eight topics—again the actual difference is small—only one more Tweet per user.

Additionally, Figures 5 and 6 show the results of the average thread degree and average thread depth per topic. For both, the topic of VAW had values over three times greater than that of the topic "politics." The results show that the topic "violence against women" has more conversational responses than the other eight trending topics. In addition, the average length of responses is longer for "violence against women" compared with the other eight trending topics. The caution remains that in actual numbers the differences are small.

Consistent with results from Figures 3 and 4, we found that the topic "violence against women" is made up of an engaged community.

### Table 2: Twitter dataset and thread information

<table>
<thead>
<tr>
<th>Topic</th>
<th>Total tweets</th>
<th>Percentage in the sample (N = 2,500,000) (%)</th>
<th>Total threads</th>
<th>Total users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>25,351</td>
<td>1</td>
<td>17,960</td>
<td>20,477</td>
</tr>
<tr>
<td>Entertainment</td>
<td>6,402</td>
<td>0.26</td>
<td>5,523</td>
<td>5,958</td>
</tr>
<tr>
<td>Sports</td>
<td>3,538</td>
<td>0.14</td>
<td>2,889</td>
<td>3,171</td>
</tr>
<tr>
<td>Violence against women</td>
<td>1,701</td>
<td>0.068</td>
<td>730</td>
<td>908</td>
</tr>
<tr>
<td>Relationships</td>
<td>1,518</td>
<td>0.061</td>
<td>1,371</td>
<td>1,433</td>
</tr>
<tr>
<td>Fashion</td>
<td>1,496</td>
<td>0.0598</td>
<td>1,263</td>
<td>1,381</td>
</tr>
<tr>
<td>Kids</td>
<td>1,468</td>
<td>0.059</td>
<td>1,236</td>
<td>1,325</td>
</tr>
<tr>
<td>School</td>
<td>1,356</td>
<td>0.054</td>
<td>1,197</td>
<td>1,295</td>
</tr>
<tr>
<td>Food</td>
<td>1,209</td>
<td>0.048</td>
<td>1,124</td>
<td>1,178</td>
</tr>
</tbody>
</table>

Note: The rest of tweets in the sample (N = 2,500,000) are related to other topics that are not discussed in this study.

### Table 3: Twitter conversation thread

<table>
<thead>
<tr>
<th>Topics</th>
<th>Average thread depth</th>
<th>Average thread degree</th>
<th>Average users per thread</th>
<th>Average tweets per thread</th>
<th>Average Retweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violence against women</td>
<td>0.933</td>
<td>1.015</td>
<td>1.244</td>
<td>2.33</td>
<td>40</td>
</tr>
<tr>
<td>Politics</td>
<td>0.306</td>
<td>0.318</td>
<td>1.14</td>
<td>1.412</td>
<td>50</td>
</tr>
<tr>
<td>Entertainment</td>
<td>0.126</td>
<td>0.135</td>
<td>1.079</td>
<td>1.159</td>
<td>86.22</td>
</tr>
<tr>
<td>Sports</td>
<td>0.205</td>
<td>0.196</td>
<td>1.098</td>
<td>1.225</td>
<td>231</td>
</tr>
<tr>
<td>Relationships</td>
<td>0.099</td>
<td>0.097</td>
<td>1.045</td>
<td>1.107</td>
<td>473</td>
</tr>
<tr>
<td>Fashion</td>
<td>0.157</td>
<td>0.163</td>
<td>1.093</td>
<td>1.184</td>
<td>448</td>
</tr>
<tr>
<td>Kids</td>
<td>0.133</td>
<td>0.159</td>
<td>1.072</td>
<td>1.188</td>
<td>8.57</td>
</tr>
<tr>
<td>School</td>
<td>0.125</td>
<td>0.123</td>
<td>1.082</td>
<td>1.133</td>
<td>32.2</td>
</tr>
<tr>
<td>Food</td>
<td>0.073</td>
<td>0.073</td>
<td>1.048</td>
<td>1.076</td>
<td>12</td>
</tr>
</tbody>
</table>
on Twitter consisting of active users who respond to each other more frequently on topic of “violence against women” compared the users who tweet on other trending topics.

Lastly, we present numbers of retweets under each topic in Table 3. Figure 7 presents the retweet counts per topic.

The average number of retweets/broadcast messages on the topic of “violence against women” is quite small and is substantially smaller than other topics. The fact that the topic of “violence against women” has fewer Retweets than that of other topics shows that those Tweeting one another in conversations about “violence against women,” choose not to diffuse the conversations/messages.

6 | DISCUSSION AND CONCLUSION

We explore and compare the volume and the lengths of VAW-related conversations with those of other trending topics on Twitter. Results show that there are differences regarding the numbers of tweets, numbers of retweets, number of users, conversation thread depth, and conversation thread degree across these trending topics and subject matter. It provides insights regarding how the topic of VAW is
presented in the 140 characters world of Twitter, and have implications for future research, social media advocacy, and public policy.

These exploratory results suggest that there are sufficient contents on Twitter that merit additional research on the topic of VAW. One limitation of the present study is that the current paper does not present the content of Twitter posts on the topic of VAW. For example, it does not further investigate whether the tweets in the sample are denouncing VAW or demeaning to women. Thus, we suggest future research to investigate the content of posts related to VAW by using content analysis.

The study also has implications for methodology by providing readers with a roadmap for how to employ Twitter Big Data to study the nature of VAW in the digital era. The findings have implications for future research by providing a replicable guidebook for gathering and analyzing Big Data from Twitter for social justice research.

We discover a small but active engaged community with users who post on Twitter on the topic of VAW. These users post and reply to one another and produce more conversation threads compared to those posting on other trending topics. In addition, the “violence against women online community” actively engages with one another in a conversational pattern, rather than a broadcast pattern. The “violence against women” Twitter community is intensively involved in each conversation thread on the topic of “violence against women,” but they are not retweeting these conversations to broader audiences. A study examining domestic violence relevant posts on Twitter finds that tweets are grouped under two themes of “victimization” and “high-profile cases” while advocacy is not a salient topic that is neither intensively nor extensively discussed (Xue, Chen, & Gelles, 2019). The findings in the present study inform us that Twitter users are using Twitter as a platform to actively discuss VAW-related contents, which form an actively engaged online community. However, the identified online community might not be involved in conversations of advocacy which refers to the “support and service that helps victims who have an experienced or at risk of domestic violence” (Xue et al., 2019). A further step investigating the demographics information and contents of tweets of this online community on Twitter will inform whether social media, and in particular Twitter could be used for advocacy and prevention by providing services related contents in the Tweets.

ENDNOTES

1 The concept of “thread” is originally defined from email network (people copied in your emails); conversational threads also appear in forums, bulletin board system (BBS, Computerized system used to exchange public messages or files. http://www.britannica.com/technology/bulletin-board-system), and blogs, referred to “reply relationships” (Cao et al., 2012).

2 “Thread depth” refers to the number of links on the longest path from the top (root) node of the thread to the node at the bottom level. A detailed definition is provided later in Section 4.

3 “Thread degree” refers to largest node degree within the structure. Detailed definition is provided later in Section 4.

4 A classifier that computes the posterior of classification variable given a set of attributes by using the Bayes rule under the conditional independence assumption, from http://www.igi-global.com/dictionary/na%2C%2FAfve-bayes-classifier/19838.

REFERENCES


ORCID

Jia Xue ♂ https://orcid.org/0000-0002-1668-2531


Tingshao Zhu, is a professor in CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences. He obtained his PhD in computer science from University of Alberta, Canada. Until now, he has published 60+ papers as first and corresponding author. His study interests include machine learning, techniques of online/offline behavioral big data mining, as well as its applications on mental health, social service, management, and so on.

Richard J. Gelles, Professor, is a former dean at the University of Pennsylvania and holds the Joanne and Raymond Welsh Chair of Child Welfare and Family Violence in the School of Social Policy & Practice at the University of Pennsylvania. He is the director for the Center for Research on Youth & Social Policy and director of the Field Center for Children's Policy Practice & Research. Gelles is an internationally acclaimed expert in domestic violence and child welfare. He was influential in the passage of the Adoption and Safe Families Act of 1997. His book, The Violent Home was the first systematic empirical investigation of family violence and continues to be highly influential. He is the author or coauthor of 26 books and more than 200 articles and chapters on family violence.

### EXAMPLES OF COLLECTED TWEETS

<table>
<thead>
<tr>
<th>No.</th>
<th>Tweets examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RT@ABC: This woman was raped 300 times in her sleep by her husband; <a href="http://t.co/rqJ065tWV">http://t.co/rqJ065tWV</a></td>
</tr>
<tr>
<td>2</td>
<td>By voting for #Labour the evil bastards who covered it up... blamed the raped girls? Great idea. @ABC @ABC</td>
</tr>
<tr>
<td>3</td>
<td>@ABC @ABC Sex education for minors is limited to the basic biological functions.</td>
</tr>
<tr>
<td>4</td>
<td>#IndiasDaughter BANNED in India by powerful men. A rape documentary on Indian women...</td>
</tr>
<tr>
<td>5</td>
<td>Horrifically raped and murdered while serving in military, but her death was ruled a suicide...</td>
</tr>
<tr>
<td>6</td>
<td>An 11-year-old reported being raped twice... <a href="http://t.co/DACDJCqr3z">http://t.co/DACDJCqr3z</a> @ABC</td>
</tr>
<tr>
<td>7</td>
<td>Today is international day to end violence against women... #ViolenceAgainstWomen</td>
</tr>
<tr>
<td>8</td>
<td>@ABC the Muslims are not upset about raped nuns, burned churches... SILENT.</td>
</tr>
<tr>
<td>9</td>
<td>12 years ago, I've been raped, and did not tell because I thought my parents would be mad.</td>
</tr>
<tr>
<td>10</td>
<td>...numbers prove #DomesticViolence is an American epidemic <a href="http://t.co/7LruxKXZhZ">http://t.co/7LruxKXZhZ</a></td>
</tr>
</tbody>
</table>

*The tweets are anonymous here by removing their identifications and key words in the tweets.*