Received: 2 April 2019 | Revised: 18 June 2019 | Accepted: 21 June 2019
DOI: 10.1002/hbe2.160

EMPIRICAL ARTICLE

WILEY

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

Jia Xue[1] | Kathy Macropol[2] | Yanxia Jia[2] | Tingshao Zhu[3] | Richard J. Gelles[4]

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

Correspondence
Jia Xue, 246 Bloor Street West, Toronto, ON, Canada.
Email: jia.xue@utoronto.ca

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 positing, 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."[1] 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 communities. 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[2] 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[3] 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

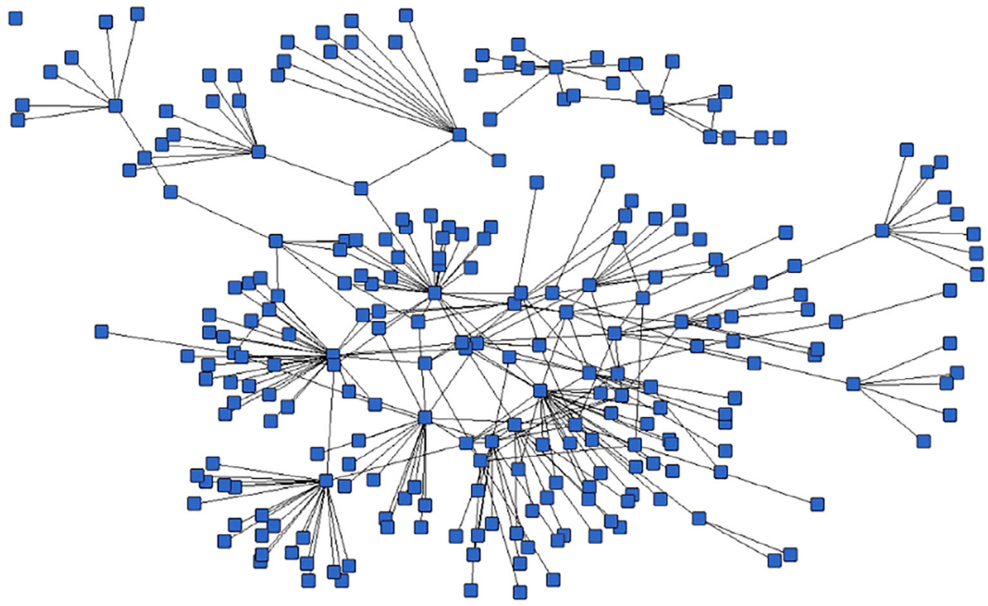

FIGURE 1 An example of part of a twitter following graph (http://stanford-ppl.github.io/Delite/optigraph/index.html)

(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 Naïve Bayes classifier (Rish, 2001) using the word frequencies within each Tweet as features for next-step topic classification. We used a Naïve 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 "#SpousalAbuse," "#DateRape," "#ViolenceAgainstWomen," and "#DomesticViolence" into the topic of VAW. 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[4] 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

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

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 "User2: reply 1" 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.

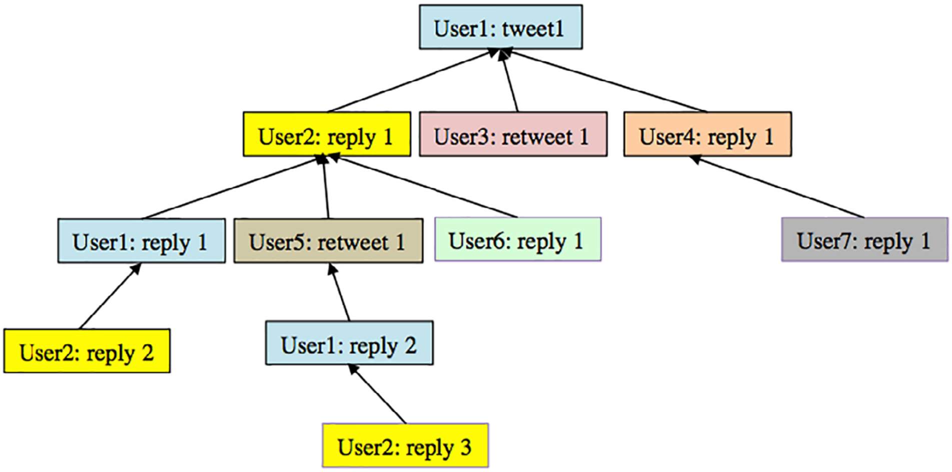

FIGURE 2 An example Twitter conversation thread

TABLE 2 Twitter dataset and thread information

Topic	Total tweets	Percentage in the sample (*N* = 2.5 million) (%)	Total threads	Total users
Politics	25,351	1	17,960	20,477
Entertainment	6,402	0.26	5,523	5,958
Sports	3,538	0.14	2,889	3,171
Violence against women	1,701	0.068	730	908
Relationships	1,518	0.061	1,371	1,433
Fashion	1,496	0.0598	1,263	1,381
Kids	1,468	0.059	1,236	1,325
School	1,356	0.054	1,197	1,295
Food	1,209	0.048	1,124	1,178

Note: The rest of tweets in the sample (*n* = 2.5 million) are related to other topics that are not discussed in this study.

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 3 Twitter conversation thread

Topics	Average thread depth	Average thread degree	Average users per thread	Average tweets per thread	Average Retweets
Violence against women	0.933	1.015	1.244	2.33	40
Politics	0.306	0.318	1.14	1.412	50
Entertainment	0.126	0.135	1.079	1.159	86.22
Sports	0.205	0.196	1.098	1.225	231
Relationships	0.099	0.097	1.045	1.107	473
Fashion	0.157	0.163	1.093	1.184	448
Kids	0.133	0.159	1.072	1.188	8.57
School	0.125	0.123	1.082	1.133	32.2
Food	0.073	0.073	1.048	1.076	12

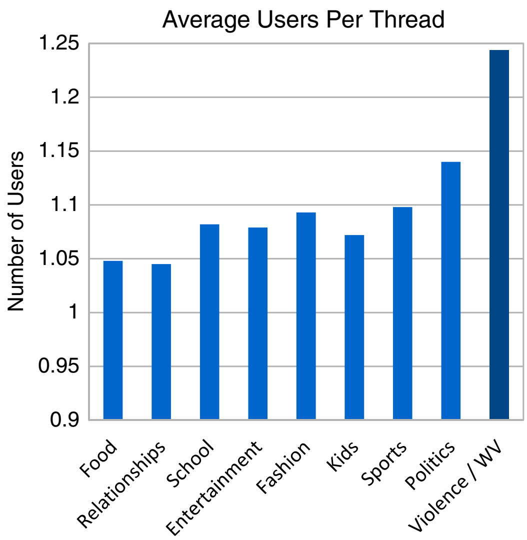

FIGURE 3 Average users per thread

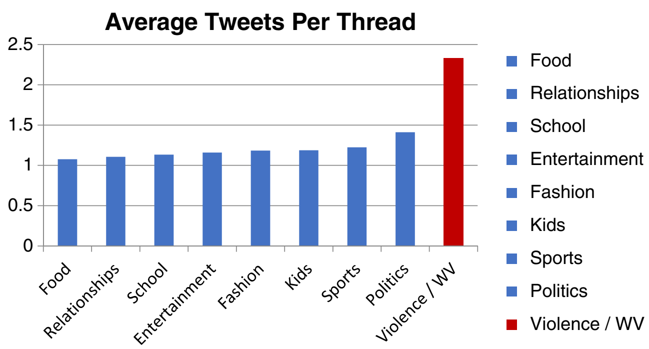

FIGURE 4 Average tweets per thread

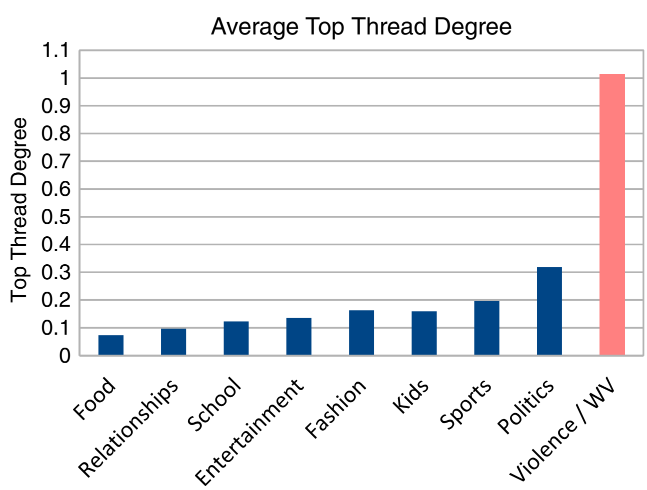

FIGURE 5 Average thread degree

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.

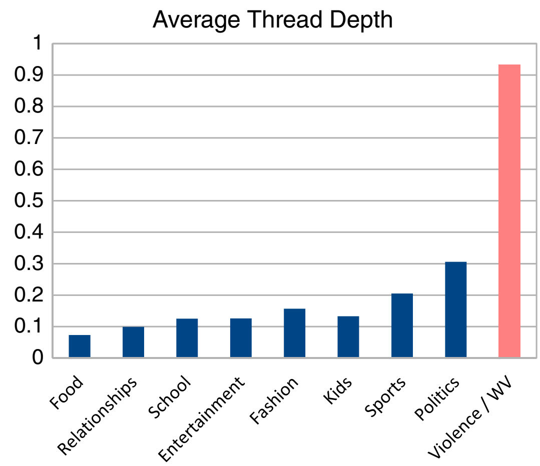

FIGURE 6 Average thread depth

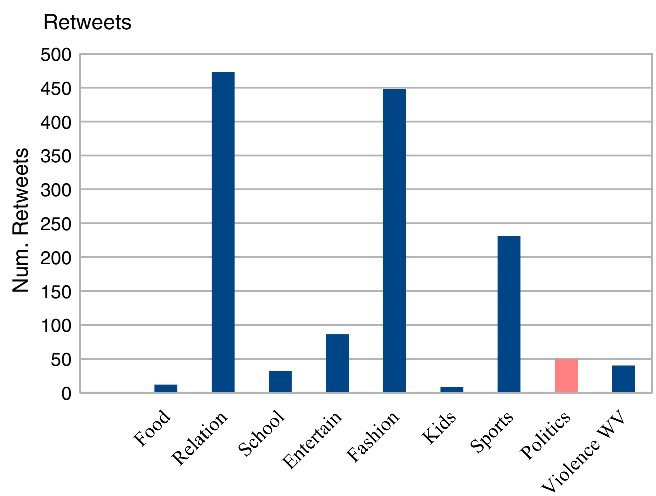

FIGURE 7 Retweet counts

5.3 | Retweets

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%C3%AFve-bayes-classifier/19838.

ORCID

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

REFERENCES

About Twitter. (2016). Retrieved from https://about.twitter.com/company

Bakhshandeh, R., Samadi, M., Azimifar, Z., & Schaeffer, J. (2011). Degrees of separation in social networks. *Proceedings of the 4th International Symposium on Combinatorial Search*, 18–23.

Bastos, M. T., Puschmann, C., & Travitzki, R. (2013). *Tweeting across hashtags: Overlapping users and the importance of language, topics, and politics*. Proceedings of the 24th ACM conference on hypertext and social media, 164–168.

Borau, K., Ullrich, C., Feng, J., & Shen, R. (2009). Microblogging for language learning: Using twitter to train communicative and cultural competence. In *Advances in web based learning–ICWL 2009* (pp. 78–87). Berlin, Heidelberg: Springer.

Bruns, A., & Burgess, J. E. (2011). *The use of twitter hashtags in the formation of ad hoc publics*. Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011.

Cao, P., Liu, S., Gao, J., Shen, H., Li, J., Liu, Y., & Cheng, X. (2012). Finding topic-related tweets using conversational thread. In *Intelligent information processing VI* (pp. 258–267). Berlin, Heidelberg: Springer.

Chang, H., & Iyer, H. (2012). Trends in twitter hashtag applications: Design features for value-added dimensions to future library catalogues. *Library Trends*, *61*(1), 248–258.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One*, *5*(11), e14118.

Culotta, A. (2010). *Towards detecting influenza epidemics by analyzing twitter messages*. Proceedings of the First Workshop on Social Media Analytics, 115–122.

Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using twitter to enhance social presence. *Journal of Information Systems Education*, *20*(2), 129.

Ebner, M., & Reinhardt, W. (2009). Social networking in scientific conferences–Twitter as tool for strengthen a scientific community. *Proceedings of the 1st International Workshop on Science*, *2*, 1–8.

Fox, S. (2011). *The social life of health information* (p. 2011). Washington, DC: Pew Internet & American Life Project.

Gabielkov, Maksym & Legout, Arnaud. (2012). *The complete picture of the Twitter social graph*. Proceedings of the 2012 ACM conference on CoNEXT student workshop (CoNEXT Student '12). 19 – 20.

Gelles, R. J. (2000). Estimating the incidence and prevalence of violence against women national data systems and sources. *Violence Against Women*, *6*(7), 784–804.

Greene, J. A., Choudhry, N. K., Kilabuk, E., & Shrank, W. H. (2011). Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine*, *26*(3), 287–292.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining twitter as an imagined community. *American Behavioral Scientist*, *55*(10), 1294–1318.

Hauffa, J., Koster, B., Hartl, F., Kollhofer, V., & Groh, G. (2016). *Mining Twitter for an Explanatory Model of Social Influence*. Proceedings of the 2nd International Workshop on Social Influence Analysis (SocInf 2016), 3–14.

Himelboim, I., Gleave, E., & Smith, M. (2009). Discussion catalysts in online political discussions: Content importers and conversation starters. *Journal of Computer-Mediated Communication*, *14*(4), 771–789.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is twitter, a social network or a news media?* Proceedings of the 19th International Conference on World Wide Web, 591–600.

Lampos, V., & Cristianini, N. (2010). Tracking the flu pandemic by monitoring the social web. In *2010 2nd International Workshop on Corgnitive Information Processing (CIP)* (pp. 411–416). Elba Island, Italy: IEEE.

Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011). *Twitter trending topic classification.* Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, 251–258.

Macropol, K., Bogdanov, P., Singh, A. K., Petzold, L., & Yan, X. (2013). I act, therefore I judge: Network sentiment dynamics based on user activity change. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 396–402). Niagara Falls, Canada: ACM.

Marwick, A. E. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13* (1), 114–133.

Meeder, B., Tam, J., Kelley, P. G., & Cranor, L. F. (2010). RT@ IWantPrivacy: Widespread violation of privacy settings in the twitter social network. Proceedings of the Web, 2 1.2.

Mischaud, E. (2007). *Twitter: Expressions of the whole self. An investigation into user appropriation of a web-based communications platform.* London, England: Media@ Lse, London School of Economics and Political Science (LSE).

Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhao, Y., Zhang, Z., & Gao, Y. (2013). *FluMapper: An interactive CyberGIS environment for massive location-based social media data analysis.* Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery, 33.

Paul, M. J., & Dredze, M. (2011). You are what you tweet: Analyzing twitter for public health. *Icrwsm*, *20*, 265–272.

Rish, I. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, *3*(22), 41–46.

Russell, M. (2013). *Mining the social web* (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc.

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, *38*(3), 182–188.

She, J., & Chen, L. (2014). *Tomoha: Topic model-based hashtag recommendation on twitter.* Proceedings of the companion publication of the 23rd international conference on world wide web companion, 371–372.

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS One*, *6*(5), e19467.

Sorenson, S. B., Shi, R., Zhang, J., & Xue, J. (2014). Self-presentation on the web: Agencies serving abused and assaulted women. *American Journal of Public Health*, *104*(4), 702–707.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). *Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network.* Social Computing (Socialcom), 2010 Ieee Second International Conference on, 177–184.

Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503.

Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on information and knowledge management* (pp. 1031–1040). Glasgow, Scotland: ACM.

Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social media as information source: Recency of updates and credibility of information. *Journal of Computer-Mediated Communication*, *19*(2), 171–183.

World Health Organization. (2015). *Violence against women.* Retrieved from http://www.who.int/mediacentre/factsheets/fs239/en/

Xue, J., Chen, J., & Gelles, R. (2019). Using data mining techniques to examine domestic violence topics on Twitter. In *Violence and gender* (Vol 6). New Rochelle, NY: Mary Ann Liebert, Inc. http://doi.org/10.1089/vio.2017.0066

Zhao, X., & Jiang, J. (2011). *An empirical comparison of topics in twitter and traditional media.* Singapore Management University School of Information Systems Technical paper series.

AUTHOR BIOGRAPHIES

Jia Xue, is an Assistant Professor in the Factor-Inwentash Faculty of Social Work and the Faculty of Information at the University of Toronto. She received her PhD in Social Welfare from School of Social Policy and Practice, along with a dual master's degree in Statistics from Wharton School, University of Pennsylvania. She took her research fellowship at Harvard Kennedy School of Government. Jia applies computational and big data approaches to examine various facets of intimate partner violence and sexual assault. Jia's research agenda extends to the international context. Her article “Rape Myths and the Cross-Cultural Adaptation of the Illinois Rape Myth Acceptance Scale in China,” published in the *Journal of Interpersonal Violence*, provides the first psychometric measure to assess attitudes about rape in China. Her research also has been published in scholarly journals such as *Child Abuse & Neglect*, *American Journal of Public Health*, and *British Journal of Social Work*.

Kathy Macropol, is an Associate Professor of Computer Science at Arcadia University. Her research focuses upon Data Mining and Bioinformatics, and especially the discovery of new techniques to model and mine large, heterogeneous, time-evolving graphs.

Yanxia Jia, is an Associate Professor of Computer Science in the Department of Computer Science & Math at Arcadia University. She earned a Bachelor's of Science degree in Computer Science from Harbin Engineering University, China, a master's degree in Computer Science from North China Institute of Computing Technology, and a PhD in Computer Science from the University of Alberta, Canada. Dr. Jia's research interests include computer networks, mobile computing, and machine learning. During her time at Arcadia University, she has taught various Computer Science courses and mentored undergraduate research projects in multiple areas, such as wireless sensor networks, mobile health, and social media.

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. Gellee 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.

How to cite this article: Xue J, Macropol K, Jia Y, Zhu T, Gelles RJ. Harnessing big data for social justice: An exploration of violence against women-related conversations on Twitter. *Hum Behav & Emerg Tech*. 2019;1:269–279. https://doi.org/10.1002/hbe2.160

APPENDIX

EXAMPLES OF COLLECTED TWEETS

No.	Tweets examples[a]
1	RT@ABC: This woman was raped 300 times in her sleep by her husband; http://t.co/rqJ065IzWV
2	By voting for #Labour the evil bastards who covered it up … blamed the raped girls? Great idea. @ABC @ABC
3	@ABC @ABC Sex education for minors is limited to the basic biological functions.
4	#IndiasDaughter BANNED in India by powerful men. A rape documentary on Indian women…
5	Horrifically raped and murdered while serving in military, but her death was ruled a suicide…
6	An 11-year-old reported being raped twice … http://t.co/DACDJCqr3z @ABC
7	Today is international day to end violence against women… #ViolenceAgainstWomen
8	@ABC the Muslims are not upset about raped nuns, burned churches … SILENT.
9	12 years ago, I've been raped, and did not tell because I thought my parents would be mad.
10	…numbers prove #DomesticViolence is an American epidemic http://t.co/7LruvKXZhZ

[a]The tweets are anonymous here by removing their identifications and key words in the tweets.