VIOLENCE AND GENDER
Volume 6, Number 2, 2019

DOI: 10.1089/vio.2017.0066

Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter

Jia Xue, PhD,[1] Junxiang Chen, PhD,[2] and Richard Gelles, PhD[3]

Abstract

This study aims to discover hidden topics and thematic structures among domestic violence-related texts on Twitter. We collected 322,863 messages using the key term ''domestic violence.'' We used unsupervised machine-learning methodology Latent dirichlet allocation, and found that the most common 20 pairs of words were ''violence awareness,'' ''greg hardy,'' ''awareness month,'' ''victims domestic,'' ''stop domestic,'' and ''ronda rousey.'' We identified 20 topics that appear most frequently, such as Topic 19 with frequent words ''greg hardy,'' ''photos greg,'' ''dallas cowboys,'' ''charges expunged,'' ''hardy girlfriend,'' and also assigned themes (e.g., ''*Greg Hardy domestic violence case*'') for the topics. This study demonstrates the feasibility of using topic-modeling methods for mining gender-based violence data on Twitter.

Keywords: domestic violence, Twitter, topic modeling

Introduction

TWITTER, LAUNCHED IN 2006, is one of the most widely used social media platforms to update personal status information and interact with others across the world. Twitter users increased from 8% of U.S. online adult population in 2010 to 18% in 2013 (Brenner and Smith 2013). Furthermore, Twitter is used as a data analysis source for health research (Prier et al. 2011). Twitter offers larger numbers of participants than any form of survey research and also provides ''open-vocabulary exploratory analysis'' (Schwartz and Ungar 2015). In addition, Twitter is an important channel to reach out to ''traditionally difficult-to-reach populations'' (Harris et al. 2014). Twitter helps eliminate response bias because Twitter offers the venue where the public often posts health-related topics that they withhold from offline friends or families (Kolmes and Taube 2016). Scholars have examined health-related contents on Twitter, such as cancer (Koskan et al. 2014); mention of nonspecific diseases (Weeg et al. 2015); heart disease (Eichstaedt et al. 2015); allergies, obesity, and insomnia (Paul and Dredze 2011); antibiotics usage (Scanfeld et al. 2010); and dental pain (Heaivilin et al. 2011). Researchers also describe dialog-specific content on Twitter, such as lung cancer clinical trials (Sedrak et al. 2016). Scholars also assess the use of Twitter among social work scholars (Greeson et al. 2017). Studies using social media data demonstrate the possible value of using social media to investigate the impact of domestic violence on mental health (Liu et al. 2018).

Domestic violence is the most common form of violence against women, affecting as many as one-third of women worldwide (Black et al. 2011). While there are public health issues posted on Twitter, we know little about the nature and content of domestic violence-related posts on Twitter. Thus, the goal of this study is to identify domestic violence-related content within Twitter's conversational data. The result of our exploratory research may have implications for domestic violence scholars and practitioners by opening up a new source of data and information about domestic violence. The study provides a unique view of domestic violence information on Twitter by linking social science with advanced statistics methods to better understand violence against women in the current social media environment.

Literature Review

Twitter and public health

Twitter is one of the most widely used social media platforms, serving as a public viewing platform for collecting, disseminating, and sharing information. There are an estimated 288 million active Twitter users every month, and there are >500 million Tweets posted *every day* on About Twitter. (2015, October 5). Retrieved October 5,

[1]Faculty of Information, Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, Canada.
[2]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania.
[3]School of Social Policy & Practice, University of Pennsylvania, Philadelphia, Pennsylvania.

2015, from https://about.twitter.com/en_us.html). Twitter users are allowed to send any messages up to a 140-character limit. Besides microblogging function, Twitter users can reply or retweet (RT) others' Tweets. The default function for users' accounts and their Tweets are open and publicly available on Twitter (Marwick, A. E., and Boyd, D. 2011).

Twitter is a community for public health information and data (Liu et al. 2018). 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 (Paul and Dredze 2011; Scanfeld et al. 2010). Twitter users tweet feeds in areas such as influenza, obesity, insomnia, antibiotics, depression, and cancer. 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 (Paul and Dredze 2011). For instance, Tweets have been used to determine the extent of the H1N1 outbreak (Chew and Eysenbach 2010). Culotta (2010) found that monitoring influenza-related Tweets provides cost-effective and quick health status surveillance. Other public health problems are also examined on Twitter to inform public health programs, such as heart disease, obesity, dental pain, and cancer (Eichstaedt et al. 2015; Heaivilin et al. 2011; Paul and Dredze 2011; Sedrak et al. 2016). Researchers systematically reviewed the use of Twitter for health research (Sinnenberg et al. 2017), which showed that public health (23%) and infectious disease (20%) were the most commonly represented topics among those 137 peer-reviewed original studies.

Domestic violence as a public health problem

Domestic violence is a serious social problem worldwide (Xue et al. 2018). It is estimated that one-third of women worldwide have experienced some form of domestic violence by their intimate partner in their lifetime (WHO 2017). *The National Intimate Partner and Sexual Violence Survey* (2011) found that ~35.6% of women report a lifetime rate of intimate partner victimization of some form of violence, such as rape, physical violence, or stalking. Even though women are more likely to be victims of domestic violence, men are also victimized by intimate partners. Nearly 28.5% of men report being the victims of some form of violence by an intimate partner in their lifetime. Same-sex intimate partner violence is also a serious public health issue (Mitchell-Brody et al. 2010). A third of lesbian women (33.5%) and one in four gay men (26%) experience at least one type of domestic violence in their lifetime (Black et al. 2011). Domestic violence is associated with negative consequences for physical health (e.g., injury, chronic pain), mental health (e.g., depression, posttraumatic stress disorder), sexual health (e.g., sexually transmitted diseases), and women's reproductive health (Campbell 2002).

Domestic violence and Twitter

Domestic violence is a global public health problem. For decades, scholars have collected data about the nature of this social problem from interviews with victims, surveys that employ in-person interviews or questionnaires, and by analyzing official and administrative data, such as crime statistics or medical records (Gelles 2000). With the widespread use of social media, Twitter provides a new window into the nature of domestic violence. For example, 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 et al. 2014). Victims of partner violence and sexual assault use information communication technology, including Twitter to seek information (Xue et al. 2018), and/or attempt to build communities that allow them to discuss their personal experience as well as inform the public about the magnitude of the social problem, such as the #Metoo campaign. Given the importance of the social problem of domestic violence and the growing and rather substantial use of Twitter, there is a reasonable argument for exploring the contents regarding what Twitter users are talking about with regard to domestic violence on Twitter. However, thus far, there is no research that examines the topics posted on Twitter. The findings of the study could be a resource for practitioners and advocates to better understand Twitter's possible contribution as a platform of information diffusion to implement violence prevention and intervention.

Advanced statistical methods: latent dirichlet allocation

According to Blei et al. (2003), Latent dirichlet allocation (LDA) is an unsupervised machine-learning method that identifies latent topic information in a document collection. It employs a ''bag of words'' approach; that is, documents are represented using counts of linguistic units, where the linguistic units can be either single words (uni-grams)[1] or contiguous sequences of *n* words (*n*-gram)[2], disregarding grammar and the order of the units. The model assumes that each document consists of a mixture over various latent topics, and each topic is characterized using a distribution over the linguistic units. By applying the model to a document collection, we expect to extract the following information:

(1) The distribution over linguistic units for each latent topic, where the units with high frequency indicate that those units tend to cooccur together. We are able to assign a theme for each latent topic by analyzing the distributions.
(2) The distribution over topics for each document. By observing the distribution, we understand on which topics each document focuses.
(3) The distribution over topics for the whole document collection. The distribution tells us an overview about which topics are more popular and which appear less frequently.

LDA employs unsupervised learning methods and presents the data distributions based on the data themselves, which indicates that LDA can be used in large dialog datasets like Twitter. Prier and colleagues (2011) identify health-related topics on Twitter, in particular Tobacco-related Tweets by applying LDA. The study generated 250

[1]Uni-gram: when an n-gram of size equals 1

[2]When we use bi-gram (N = 2), it means the pairs of consequent words.

topic distributions for single words (uni-grams) and structural units (n-grams), which exhibit sufficient cohesion. Wang and colleagues (2014) applied LDA to website posts and generated 20 topics. LDA gives a topic probability distribution that reveals the probability of a post corresponding to each topic. Godin and colleagues (2013) used LDA model in the context of Tweets hashtag recommendation. They trained the LDA model to cluster Tweets into various topics, and then used the keyword to suggest new Tweets. Zhao and colleagues (2011) used LDA model to discover topics from Twitter and compare them with traditional news media—for example, *The New York Times*. They compared standard LDA, author-topic model, and Twitter-LDA, and proposed that the Twitter-LDA model outperforms the other two models for identifying topics from Twitter. Their Twitter-LDA model is based on the hypothesis that one Tweet expresses one content of a topic. Yamamoto and Satoh (2013) used LDA to extract topics and also propose a two-phase extraction method by combining LDA for clustering large amounts of documents and constructing an association between the topics and aspects.

Purpose of the study

Our goal is to explore the conversations and discussions regarding domestic violence on Twitter. We employ LDA to explore latent topics related to domestic violence in a dataset of Tweets. Specifically, we propose several research questions with regard to Twitter postings that include the term ''domestic violence'':

(1) What are the most popular words in the whole document collection?
(2) What domestic violence-related words tend to cooccur together?
(3) Which domestic violence-related topics appear most frequently?
(4) Which topics does the whole document collection focus on?
(5) What are the themes of the identified latent topics?
(6) For each latent topic, what are the distributions of the linguistic units? Which words appear more frequently with high frequency?

Methodology

Dataset

We collected messages through the Twitters Streaming Application program interface (API). We used the key term ''*domestic violence*'' as the search term to ''fetch'' messages that mention the pair of words ''domestic violence.'' Thus, all collected Tweets contain the words ''domestic violence.'' We collected Twitter messages from October 2015 through January 2016. The total sample and dataset for the study consisted of 322,863 Tweets that included the terms ''domestic violence.'' The sample is a random sample of 1% of the full stream of posts. We downloaded the dataset in the ''CSV'' format and read it through the software Python.

Data analysis

We used Python to analyze the data. We configured LDA to generate 20 latent topic distributions by using structural units bigrams (n-gram, when $n=2$). A bigram is a sequence of two adjacent linguistic elements, such as a pair of words (e.g., ''domestic violence,'' ''violence victims'').

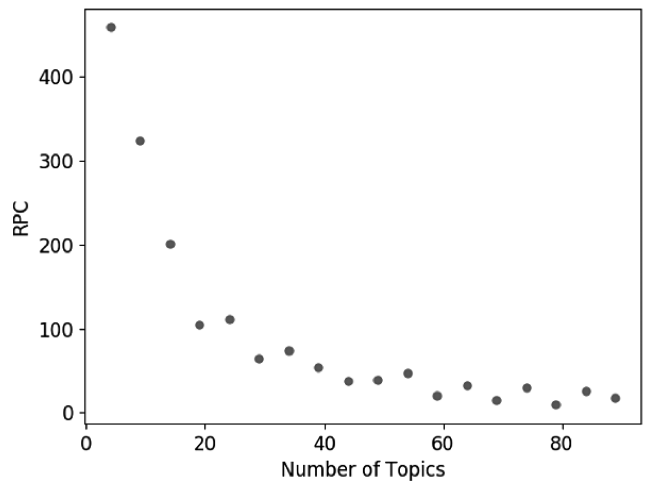

FIG. 1. RPC against the number of topics. RPC, rate of perplexity change.

The process is provided as follows:

(1) We removed the hashtag symbol ''#,'' ''@ users,'' and URLs from the messages because, in our analysis, we did not make use of the author information, and the hashtag symbols or the URLs did not provide topic information. In addition, since we focused our analysis on the messages in English, we removed all non-English characters.
(2) We converted Twitter messages into a document-term matrix, whose element represents the count of each bigram (contiguous sequences of two words, such as ''domestic violence'' or ''human trafficking'') that occurs in each of the messages. This was done by applying the CountVectorizer[3] function provided in the scikit-learn package[4].
(3) We first determine the number of topics, which is a parameter for the LDA model. We achieve this by tentatively changing the number of topics, run the LDA model (by making use of the LDA[5] class provided in the *scikit-learn* package), and compute the rate of perplexity change (RPC) as introduced by Zhaoand colleagues (2015). We plot RPC against the number of topics in Figure 1. We follow the heuristics introduced by Zhao and colleagues (2015), such that we choose the number the first i satisfying RPC(i) <RPC($i+1$). By observing Figure 1, we let the number of topics be 20.
(4) We analyzed the obtained document-term matrix using the LDA model with 20 topics. The computer

[3]Convert a collection of text documents to a matrix of token counts, from http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

[4]Machine Learning in Python, from http://scikit-learn.org/stable/

[5]Latent Dirichlet Allocation with online variational Bayes algorithm, from http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html

program fit the LDA model of the obtained matrix, and returned the distributions of topics in each of the documents and the distributions of terms for each topic. We summarized the results in Tables 1–4.

(5) To better understand what the themes in the latent 20 topics are, we randomly sampled 10 Twitter messages as examples for each topic. These examples constitute ≥90% of the content in each topic; for example, the Tweets example of ''*Dallas Cowboys Rumors: Greg Hardy's Domestic Violence Charges Expunged In Spite Of Common*'' in Topic 18. About 90% of the linguistic units in this Tweet belong to Topic 18. We selected 1–2 of 20 examples in several latent topics and presented them in Table 3.

Results

Popular words relating to domestic violence

In the whole document collection, we identified the most popular words related to domestic violence. In addition to the key search term ''domestic violence,'' the results show that popular bigrams (pairs of words) are ''violence awareness,'' ''greg hardy[6],'' ''awareness month,'' ''victims domestic,'' ''stop domestic,'' and ''ronda rousey[7].'' Note that bigram merely captures two concessive words, regardless of the grammar structure and semantic meaning. Therefore, some bigrams might not be self-explanatory. For instance, popular pairs of words such as ''rt domestic,'' ''hardy domestic,'' and ''rt ronda'' are not long enough to be meaningful. After we investigate other popular bigrams, we identify that they represent the meanings of ''rt domestic violence,'' ''greg hardy domestic violence,'' and ''rt ronda rousey.''

We collected 322,863 Tweets as our document population. Among all collected Tweets, there are 80,868 bigrams (e.g., ''domestic violence,'' ''stop domestic''). We choose the 20 most common words (16.72%) with the highest percentage in all 80,868 bigrams (100%) and present them in Table 1. For instance, ''domestic violence'' constituted 10.12% among all 80,868 bigrams, which means ''domestic violence'' appears, on average, once in every 10 bigrams. We also included ''rt'' in the bigram analysis, for example, ''rt domestic,'' ''rt ronda,'' and ''rt stop.'' As an artifact of the API, rt means retweet, which shows that the message has been reposted. The results of popular bigrams inform us that certain words are popular because not only they have been mentioned frequently but they have also been reposted frequently.

High frequency of cooccurred domestic violence bigrams

We identified the domestic violence-related words that tend to cooccur together and appear most frequently. LDA helps browse words that are frequently found together or share a common topic. Our LDA outputs reveal that many bigrams tend to cooccur together among our sampled domestic violence-related Tweets, such as ''justice4cindy cindy,'' ''live pets,'' ''raise awareness,'' and ''participate purplethursday,'' and celebrity-athlete names, including ''greg hardy,'' ''william gay,'' and ''ronda rousey.'' In addition, the cooccurring words share common topics (we set the number of topics as 20 in this study). All the identified 20 latent topics with high frequency of cooccurrence bigrams are sorted according to their frequency and are presented in Table 2.

Table 2 presents the distributions of all 20 latent topics (sum equals 100%), indicating the most common latent topics that the whole document of collection focuses on. For instance, Topic 19 has the highest distribution (8.33%), ranking the most latent one, among all 20 latent topics. Table 2 also indicates the bigrams that tend to cooccur together among all collected domestic violence-related Tweets in the sample. For instance, within Topic 19, pairs of words ''greg hardy,'' ''violence incident,'' ''photos greg,'' ''hardy girlfriend,'' and ''girlfriend alleged'' have high frequency of cooccurring together. These pairs of words cooccur together to share the same Topic 19.

Topics distributions by date

We also calculated the topic distributions on all 20 latent topics by date. Figure 2 shows the changes of several topics' distributions over time. In Figure 2, we present the topic distributions for Topics 2, 3, 6, 9, 10, 16, 17, 18, and 19 from October 1, 2015 to January 7, 2016, because the distributions of these topics change over time while the changes of other topics do not fluctuate a lot. For each single date, the distributions of total 10 topics sum up to 100%.

In Figure 2, we can see that topics change over time. For example, Topic 10 (dashed line) has three peaks of distribution: 64.1% on December 3rd, 54.2% on December 6th, and 53.1% on December 26th. We found important Tweets examples within Topic 10: ''*RT @WeNeedFeminlsm: Domestic violence hotline: 1-800-799-7233 #StopDomesticViolence https://t.co/ymSMyWHhKV*.'' and ''*US Senate passed resolution supporting the goals and ideals of National Domestic Violence Awareness Month: https://t.co/YSB1wij8zJ#DVAM2015*,'' indicating that Twitter users frequently posted information about hotline and DV awareness month sporadically in December even though October was the Domestic Violence Awareness month. Similarly, Topic 6 (green line) has a distribution of 36.4% on November 11th, which takes one-third of all topics' distributions on that date. In contrast, Topic 6 has steadily low topic distributions on other dates. We noticed that one important Tweet example within Topic 9 is ''*RT Ronda Rousey Domestic Violence: 'Rowdy' Benefits From Double Standard [VIDEO],*'' indicating that Twitter users were frequently broadcasting Ronda Rousey's domestic violence news events on November 11th compared with other days.

Themes of the identified latent topics

We also assigned themes for several identified latent topics after examining the popular words in each identified topic and their relevant examples,[8] as shown in Table 3. For

[6]Greg Hardy was a professional football player. During the time of data collection, he played for the National Football League team, The Dallas Cowboys.

[7]Rhonda Rousey is an American mixed martial artist, judoka, and actress. Rousey was the first U.S. woman to earn an Olympic medal in judo at the 2008 Summer Olympics in Beijing.

[8]We presented one or two examples under the identified topics.

Table 1. Top 20 Popular BiGrams (Pairs of Words)

Popular bigrams	*Dataset (%)*	*Popular bigrams*	*Dataset (%)*
Domestic violence	10.12	Violence victims	0.29
Violence awareness	0.88	Jose reyes[10]	0.25
Greg hardy	0.82	Rt ronda	0.24
Rt domestic	0.62	Rt stop	0.24
Awareness month	0.43	William gay	0.21
Victims domestic	0.40	Support domestic	0.20
Hardy domestic	0.34	Violence charges	0.19
Stop domestic	0.32	Arrested domestic	0.18
Violence incident	0.32	Awareness domestic	0.18
Ronda rousey	0.31	Alleged domestic	0.18

We choose top 20 common words with the highest percentage in all 80,868 bigrams (100%). The rest of the 80,848 bigrams constituted 83.28%.

example, Topics 15, 18, and 19 are assigned as the theme ''Greg Hardy domestic violence case'' because these three topics focus on the news event of the NFL football player Greg Hardy who was arrested for assaulting his ex-girlfriend in November 2015. Topic 19 has a distribution of 8.33% among all identified 10 topics (Topic 18 with 6.68% and Topic 15 with 5.55%), which suggest that the news event of Greg Hardy was a salient news event and discussed widely among Twitter users.

Within Topic 6, topic components involve popular bigrams, including ''ronda rousey,'' ''double standard,'' ''standard video,'' ''benefit double,'' and ''violence accusations.'' After carefully investigating the Tweets examples under Topic 6, we identify that all bigrams under Topic 6 cover news contents about domestic violence and famous people Ronda Rousey. Therefore, we assign Topic 6 with a theme of *Double standard & Ronda Rousey*.

Distribution and frequency of bigrams under each latent topic

Within each identified popular topic, we ran the analyses on the distribution of each bigram. We present the results of top three common bigrams under each latent topic in Table 4.[9] For example, ''greg hardy'' has a distribution of 1.71% within Topic 15, and it also comprises 2.22% under Topic 18 and 2.94% under Topic 19. Even though the percentage is small, it is higher compared with all other bigrams in the datasets ($n = 80{,}868$). The popular bigram ''greg hardy'' ranks at the top of the popular pairs of words that are more likely to cooccur together under three topics, which suggest that the news event Greg Hardy is identified as a high-profile domestic violence news broadcast on Twitter from October 2015 to January 2016.

Discussion and Conclusions

There are a comparatively large number of postings on Twitter that pertain to domestic violence. There may be more if we used other filter terms such as ''*Intimate Partner Violence*,'' ''*Wife Beating*,'' or ''*Wife Abuse*.'' In addition, there are computational social science techniques that allow us to extract and classify information on domestic violence that is posted on Twitter. Topic-modeling techniques produce clusters of words, allowing us to organize large collections of unstructured texts on social media, which offers insights understanding the messages. Third, during the time frame we sampled, with the key word ''domestic violence'' we identified patterns in the postings. The postings can be grouped under the following general themes:

(1) Victimization. We found that the word ''victims'' appears often on social media. The terms include ''victims domestic,'' ''help victims,'' ''violence survivors,'' ''violence victims,'' and ''male victims.'' In contrast, we did not identify terms such as ''abuser,'' ''batterer,'' ''perpetrator,'' ''perp,'' or ''offender.'' Instead, the abusers' names (e.g., Greg Hardy) are directly posted to indicate specific instances of domestic violence. This reveals a trend on social media that online domestic violence-related topics focus on protection and support of victims, rather than intervention against abusers. Research shows that media representation of domestic violence impacts individual behaviors as well as public policy responses because the portrayals influence people's understanding of a social problem, including the causes or consequences of an incident (Sotirovic 2003). Thus, the media depictions of domestic violence are important in terms of creating a social climate to support victims. Our study echoes the current social movement #Metoo with which sexual assault victims post their personal victimization experience of sexual assault and harassment on Twitter. Our study informs policy advocates and practitioners regarding utilizing social media as a venue to empower victims. Future research can conduct content analyses of the Tweets related to victims to develop strategies for how to create a social environment on social media to empower victims.

(2) Discussion of high-profile cases of domestic violence—in particular sports figures who committed domestic violence. Results show that most topics are classified as high-profile sports-related domestic violence topics, including Greg Hardy and his team, the Dallas Cowboy. Other sports figures mentioned in Tweets include William Gay, Jose Reyes, and Ronda Rousey. Research shows that there is an interplay

[9]The bi-gram ''domestic violence'' ranks top 1 for all topics, thus we removed it from analyses.

[10]Jose Reyes is a professional baseball player. During the time of data collection, he was a member of the Major League baseball team, The Colorado Rockies.

Table 2. Topics Relevant to Domestic Violence and Their Components with Distribution

Topic	*Topic components bigram*	*Distribution (%)*
19	Domestic violence, stop domestic, greg hardy, violence incident, alleged domestic, photos greg, hardy girlfriend, girlfriend alleged, incident released, hardy domestic, victims domestic, raising awareness, help victims, cancer domestic, ray rice, breast cancer, violence serbia, serbia donate, awareness domestic, violence women	8.33
10	Domestic violence, 800,799, violence hotline, hotline 800, 7,997,233, 7233 stopdomesticviolenc, violence awareness, violence hormonestheseries3, victims domestic, awareness month, campaign domestic, violence work, national domestic, powerful campaign, breast cancer, domesticabuse vaw, parents domestic, talk powerful, wow talk, gay parents	7.50
18	Domestic violence, hardy domestic, violence charges, greg hardy, charges expunged, violence incident, dallas cowboys, incident published, cowboys rumors, rumors greg, expunged spite, spite common, common htt, photographs hardy, violence sexual, sexual assault, violence crime, killed domestic, awareness domestic, end domestic	6.68
8	Domestic violence, pet friendly, friendly domestic, violence shelters, violence joke, need pet, stand domestic, charged domestic, violence isn, violence victims, hurt person, importance domestic, el masri, glad actually, share importance, actually wants, wants share, person loves, violence shelter, joke hurt	6.08
6	Domestic violence, ronda rousey, rousey domestic, double standard, violence rowdy, standard video, benefits double, rowdy benefits, responds domestic, rousey responds, violence accusations, victims domestic, violence women, violence murder, women health, treatment asylum, asylum victims, humane treatment, extend humane, administration extend	5.77
17	Domestic violence, victims just, just important, important female, violence victims, male domestic, female victims, victims male, rape victims, male rape, female victi, violence sexual, violence police, affected domestic, sexual assault, violence problem, violence shelter, women affected, violence awarenes, violence survivors	5.56
15	Domestic violence, greg hardy, awareness domestic, raise awareness, nfl fined, violence rape, jokes domestic, trying raise, don understand, player trying, fined player, violence related, make jokes, people make, like really, understand people, mentions like, tweet feminists, rape tweet, feminists mentions	5.55
11	Domestic violence, violence awareness, hazem el, el masri, alleged domestic, victims domestic, assault domestic, violence incident, end domestic, sexual assault, victim domestic, anti domestic, wear purple, shit domestic, hardy alleged, violence claims, violence experiment, help domestic, ex wife, national domestic	5.35
16	Domestic violence, violence case, police domestic, end domestic, awareness domestic, experiences domestic, victims domestic, violence victims, manziel nfl, johnny manziel, nfl contacts, contacts police, experience domestic, inspired help, violence inspired, women experience, child abuse, music video, violence don, anti domestic	5.25
20	Domestic violence, victims domestic, violence victims, nfl players, police officers, violence rate, stand domestic, officers domestic, active nfl, 300 higher, rate 300, higher active, men victims, violence abuse, greg hardy, violence cases, gun violence, looks like, end domestic, think domestic	5.04
7	Domestic violence, violence awareness, support domestic, william gay, wearing purple, help support, add twibbon, awareness speak, speak add, gay fined, purple shoes, greg hardy, shoes domestic, awareness greg, fined nfl, hardy make, nfl wearing, purple cleats, fined wearing, violence victims	4.98
2	Domestic violence, victims domestic, johnny manziel, male victims, manziel domestic, violence national, greg hardy, bench johnny, nfl bench, national average, reported domestic, assault domestic, sexual assault, average 10, police reported, homes police, nfl average, 40 homes, 10 nfl, sign petition	4.78
13	Domestic violence, men domestic, women men, violence join, stand women, participate purplethursday, join participate, issues domestic, people upset, violence horrific, horrific crimes, nfl issues, upset cam, cam likes, likes dance, crimes people, dance thing, women domestic, violence affects, end domestic	4.64
12	Violence awareness, awareness month, domestic violence, october domestic, nfl fines, raise domestic, national domestic, cleats raise, player wearing, fines player, william gay, wearing cleats, october national, honor domestic, awareness mont, awareness domestic, fines william, wear purple, speak speak, purple domestic	4.39
14	Domestic violence, william gay, purple cleats, act domestic, bring attention, mother killed, cleats bring, gay mother, attention dv, killed act, worn purple, violence octobers, octobers worn, speak domestic, end domestic, protest domestic, voice speak, love doesn, hope voice, inspire hope	3.90
9	Domestic violence, jose reyes, arrested domestic, reyes arrested, report jose, new domestic, violence policy, baseball new, test baseball, violence laws, shortstop jose, rockies shortstop, violence incident, reyes test, arrest jose, violence victims, alleged domestic, south carolina, granting victims, veterans domestic	3.87
4	Domestic violence, condone domestic, jerry jones, violence victim, violence victims, passed away, victim passed, katy attention, away friends, kathryn domestic, trying katy, friends trying, attention ripkatycatkat, stay safe, violence guess, guess jerry, did condone, said organization, organization did, jones sure	3.44
3	Domestic violence, stand domestic, taking stand, experience domestic, women killed, violence purplethursda, violence 40, homes police, 40 homes, explores domestic, 10 homes, homes experience, police blacklivesmat, killed day, day explores, survivor domestic, fighting domestic, emotional abuse, sexual abuse, violence awareness	3.19
5	Domestic violence, fight domestic, opens door, street harassment, rape street, harassment domestic, violence suicide, violence pedophilia, door entitlement, door rape, entitlement opens, luth consultant, joke domestic, suicide controversy, violence men, opens domestic, consultant photo, controversy death, violence powerful, death luth	3.01
1	Domestic violence, like domestic, papua new, new guinea, violence victims, violence emergency, png domestic, pass laws, laws like, victims need, commit domestic, violence play, need shelters, needs stop, government papua, emergency government, guinea needs, violence workplace, suffering domestic child domestic	2.68
Total: 100%		

Table 3. Tweets Examples and Themes for Several Domestic Violence Topics

Topic	*Tweets example*	*Theme*
15	(1) … @Drudge_Report_: #NFL: #EAGLES put ‘‘extra #mustard’’ on Greg Hardy hits after domestic violence pics… (2) …AllUnitAllDay: So Greg Hardy has be in the news more than Floyd Mayweather has for Domestic Violence and Floyd made half a billion…	Greg Hardy domestic violence case
18	(1) … Dallas Cowboys Rumors: Greg Hardy’s Domestic Violence Charges Expunged In Spite Of Common… https://t.co/… (2) … @LiveMatchInfo: Photographs of Hardy domestic violence incident published - https://t.co/rLZ5 …	
19	(1) …NationNFL: The Cowboys are endorsing domestic violence as long as they employ Greg Hardy … (2) Photos of Greg Hardy’s …girlfriend …domestic violence incident are released.	
6	(1) …Ronda Rousey Domestic Violence: ‘Rowdy’ Benefits From Double Standard … (2) …RT @LiveMatchInfo: Ronda Rousey Domestic Violence: ‘Rowdy’ Benefits From Double Standard [VIDEO] - https://t.co/Tcpqhk6…	Double standard & Ronda Rousey
7	(1) …William Gay’s mother killed … domestic violence… worn purple cleats to bring attention to DV… (2) …RT @… William Gay was fined by the NFL for wearing purple shoes for Domestic Violence Awareness. Greg Hardy will make over …	William Gay fights domestic violence
14	(1) RT @… William Gay’s mother killed in act of domestic violence. Last two Octobers, worn purple cleats to bring attention to …	
9	(1) …iveMatchInfo: Report: Jose Reyes arrested for domestic violence - https://t.co/18L9JnD… (2) #MLB #ROCKIES SHORTSTOP _ Jose Reyes arrested for domestic violence incident that sent wife to ER … via @YahooSports	Jose Reyes domestic violence case
2	(1) NFL: Bench Johnny Manziel for domestic violence - Sign the Petition! https://t.co/WfvA6 …	NFL quarterback Johnny Manziel
16	(1) RT … Johnny Manziel—NFL Contacts Police In Domestic Violence Probe https://t.co/weFFG … (2) Johnny Manziel in at quarterback. What a joke @NFL so much for all that domestic violence big talk earlier #fraud …	
20	(1) … Don’t stop supporting victims and survivors of sexual and domestic violence. #DontCutVOCA (2) Every year, more than 3 million children witness domestic violence in their homes. Let’s #SpreadLoveDC and model healthy love. #NOMore #DVAM	Domestic violence victims and survivors
17	(1) …DontCutVOCA Speak up 4 victims of domestic violence & sex assault! RT Please share! Tweet your senators! https://t.co/3ECxE… (2) RT … male rape victims are just as important as female victims…	
10	(1) … @WeNeedFeminlsm: Domestic violence hotline: 1-800-799-7233 #StopDomesticViolence … (2) 5:30–8:00 @UOGTriton Lecture Hall Support support … Since its domestic violence awareness month	Domestic violence month, awareness and hotline support
12	(1) @TrillKitten: Remember, October is also Domestic Violence Awareness Month … http://t.co/Bt89gt7bFI (2) RT … October is Domestic Violence Awareness Month. Learn more: https://t.co/qtMNVPxuUK or https://t.co/qedov0Gbfa	

For anonymous protection, we deleted several words in the Tweet examples, and replace these words by ‘‘….’’

between male athletes and their assault toward women (Webb 2011). Male athlete such as Ray Rice and his domestic violence incident generated a national conversation about the interplay between domestic violence and sports, and the need for change (Martin 2017). In 2014, Ray Rice’s attack on his fiancé became a widely publicized incident of domestic violence. However, our study suggests that the Rice case was not a prominent topic a year later. Instead of being constructed as an understanding of domestic violence by journalists in traditional media outlets, including newspapers, our findings represent the public understandings and perceptions of domestic violence and sports. Sports-related domestic violence feeds are promoted by real-time events in a timely manner.

There are limitations in the study. First, we only used *domestic violence* as a key term to collect data from Twitter. Data collection using multiple key terms would be expected to present a more complete picture of the topics related to this phenomenon on Twitter. For example, the results show overlaps between the topics, which suggests that they are drawn closer together due to the single filtering term ‘‘domestic violence’’ that we used in the study. Future studies can use more intimate and sexual violence-related filtering

Table 4. BiGrams Distributions Under Topics (Top 3 Presented)

Topic	*Topic components*	*Component distribution (%)*	*Topic*	*Topic components*	*Component distribution (%)*
1	Like domestic	0.90	2	Victims domestic	0.58
	Papua new	0.87		Johnny manziel	0.56
	New guinea	0.85		Male victims	0.42
3	Stand domestic	1.14	4	Condone domestic	1.40
	Taking stand	0.82		Jerry jones	1.28
	Experience domestic	0.55		Violence victim	1.03
5	Fight domestic	0.64	6	Ronda rousey	4.51
	Opens door	0.60		Rousey domestic	2.20
	Street harassment	0.40		Double standard	2.14
7	Violence awareness	4.33	8	Pet friendly	0.61
	Support domestic	3.40		Friendly domestic	0.60
	William gay	2.06		Violence shelters	0.55
9	Jose reyes	5.09	10	800,799	1.68
	Arrested domestic	4.21		Violence hotline	1.62
	Reyes arrested	3.71		Hotline 800	1.49
11	Violence awareness	0.75	12	Violence awareness	11.43
	Hazem el	0.50		Awareness month	8.62
	El masri	0.42		October domestic	3.08
13	Men domestic	3.09	14	William gay	1.82
	Women men	2.99		Purple cleats	1.73
	Violence join	2.90		Act domestic	1.57
15	Greg hardy	1.71	16	Violence case	0.37
	Awareness domestic	1.07		Police domestic	0.37
	Raise awareness	1.03		End domestic	0.36
17	Victims just	0.99	18	Hardy domestic	4.09
	Just important	0.99		Violence charges	2.41
	Important female	0.97		Greg hardy	2.22
19	Stop domestic	3.65	20	Victims domestic	0.93
	Greg hardy	2.94		Violence victims	0.70
	Violence incident	1.76		Nfl players	0.39

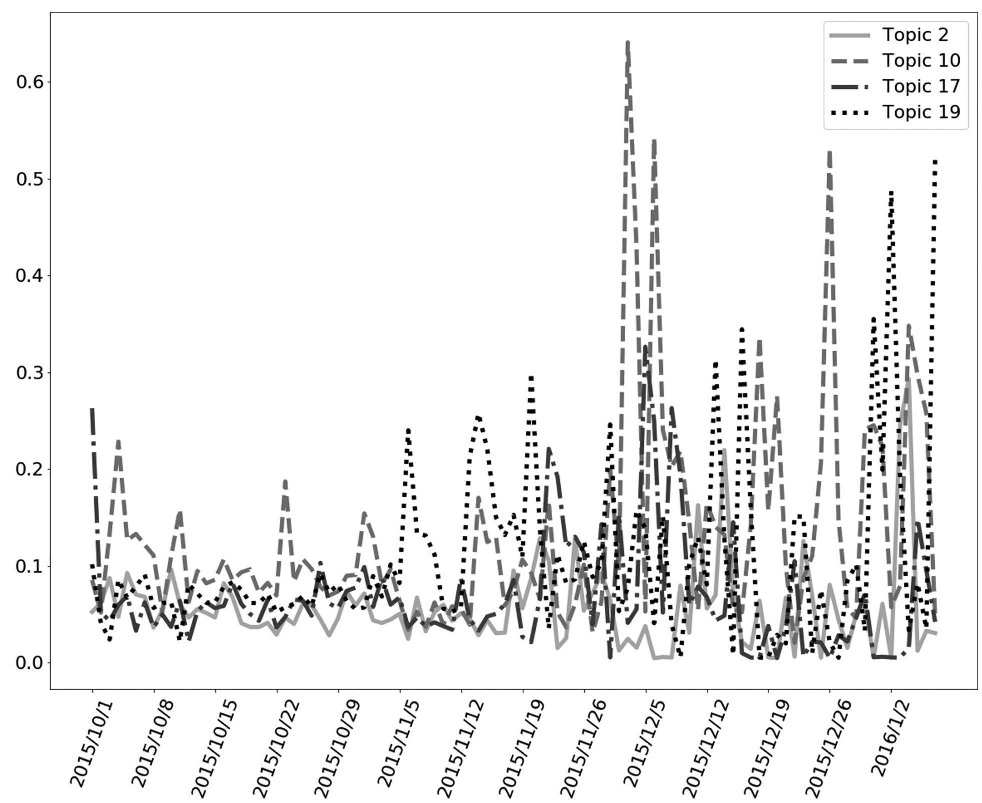

FIG. 2. Topics distributions by date. The *x*-axis shows days from October 1, 2015 to January 7, 2016. The *y*-axis represents the topic distributions (percentage).

terms to access the topics on Twitter. Another limitation of using social media data is that social statistical research is nascent using big data (Williams et al. 2017). We do not have information about the gender or demographic information of the Twitter users, which limits the generalization of our study findings to a general population. However, Twitter still provides us a valuable source to reach a valuable population offline and enable social scientists to analyze real-time social problems in a cost-effective way. Third, the data collection lasted from October to December for a period of 3 months. October is the National Domestic Violence Awareness month, in which we expect to see more advocacy-relevant Tweets than other months of the year. Future studies that cover Tweets for a longer period of time may produce different topics and themes. Our study suggests that *advocacy* was not a salient topic that is neither intensively nor extensively discussed on Twitter even during the National DV Awareness month. We suggest that DV advocacy organizations could better leverage Twitter as a broadcast tool to raise awareness and engage public discussions.

Our study has implications for advocacy and intervention. Our study is the first project that uses topic modeling to explore domestic violence-related topics on social media. Advocacy, in this study, refers to the support and service that help victims who have experienced or at risk of domestic violence. First, our research demonstrates that Twitter is an untapped and potentially valuable data source to explore the public health issue of domestic violence. More specifically, Twitter holds potential for use by advocacy groups to join in and provide context and information to those on Twitter. Those who provide services might be able to add information for those victims seeking assistance for themselves or others. Our study found sports-related high-profile cases are most tweeting or retweeting pairs of words and latent topics on Twitter, but advocacy groups as well as researchers that online communities (e.g., advocacy, public) are talking about cases, but are not messaging about the intervention/preventions information. Our study found that clusters of words focus on the level of problem recognition of the issue of domestic violence, while no salient topics were identified related to existing policy programs, advocates messages, awareness raising, or existing social services. Our findings indicate that the level of public perception of domestic violence on Twitter stays at the level of problem recognition, rather than providing effective messages/information/communication regarding social supporting services online. Here is another opportunity for advocates to provide information and context to the social media discussion about domestic violence.

Second, advocacy and intervention have a large potential audience on Twitter if it can capitalize on the 140-character format. It is possible that 140 characters limit the probability of making advocacy-related words as common ones on Twitter. When people tweet or retweet about a message, the 140-character limit reduces the likelihood of adding more advocacy/victim assistance-related words following a high-profile domestic violence case message. Thus, our findings provide insights for advocacy groups to better use the Tweets messages to promote health communication about violence prevention.

Finally, our study contributes to the research on domestic violence by providing a novel methodology for public health research. Our study reveals that Twitter is a promising venue for exploring how the majority of online Twitter users talk about public health issue of domestic violence. Our study provides insights for researchers and scholars undiscovered health contents that Twitter users focus on. Further studies can employ the same methodology to investigate domestic violence-related contents on social media during other times of the year. Furthermore, our study has implications for studying other health problems on Twitter by offering an innovative methodology in health research.

Author Disclosure Statement

No competing financial interests exist.

References

Black, M.C., Basile, K.C., Breiding, M.J., Smith, S.G., Walters, M.L., Merrick, M.T., and Stevens, M.R. (2011). The national intimate partner and sexual violence survey: 2010 summary report. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 19, 39–40.

Black MC, Basile KC, Breiding MJ, et al. (2011). National Intimate Partner and Sexual Violence Survey. (Centers for Disease Control and Prevention, Atlanta, GA.)

Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation (2003). J Machine Learn Res. 3, 993–1022.

Brenner J, Smith A. 72% of online adults are social networking site users. *Washington, DC: Pew Internet & American Life Project.* August 5th, 2013. www.pewinternet.org/2013/08/05/72-of-online-adults-are-social-networking-site-users/ (accessed May 1, 2016).

Campbell JC. (2002). Health consequences of intimate partner violence. Lancet. 359, 1331–1336.

Chew, C., and Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PloS one, 5, e14118.

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. In Proceedings of the first workshop on social media analytics. (ACM), 115–122.

Eichstaedt JC, Schwartz HA, Kern ML, et al. (2015). Psychological language on twitter predicts county-level heart disease mortality. Psychol Sci. 26, 159–169.

Gelles, R.J. (2000). Estimating the Incidence and Prevalence of Violence Against Women: National Data Systems and Sources. Violence Against Women, 6, 784–804.

Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B., and Van de Walle, R. (2013). Using topic models for twitter hashtag recommendation. In Proceedings of the 22nd International Conference on World Wide Web (ACM), pp. 593–596.

Greeson J, An S, Xue J, et al. (2018). Tweeting social work: How social work faculty use Twitter. Br J Soc Work. 48, 2038–2057.

Harris JK, Moreland-Russell S, Tabak RG, et al. (2014). Communication about childhood obesity on twitter. Am J Public Health. 104, e62–e69.

Heaivilin N, Gerbert B, Page JE, Gibbs JL. (2011). Public health surveillance of dental pain via twitter. J Dent Res. 90, 1047–1051.

Kolmes K, Taube DO. (2016). Client discovery of psychotherapist personal information online. Prof Psychol Res Pract. 47, 147–154.

Koskan A, Klasko L, Davis SN, et al. (2014). Use and taxonomy of social media in cancer-related research: A systematic review. Am J Public Health. 104, e20–e37.

Liu M, Xue J, Zhao N, et al. (2018). Using social media to explore the consequences of domestic violence on mental health. J Interpers Violence. [Epub ahead of print]; DOI:10.1177/0886260518757756

Martin T. (2017). Wake up call: How the ray rice incident opened the public's eyes to domestic violence in professional sports and the need for change. Sports Law J. 24, 183.

Marwick, A.E., and Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New media & society, 13, 114–133.
Mitchell-Brody M, Ritchie AJ, Finney J, et al. (2010). National Coalition of Anti-Violence Programs (New York City Gay & Lesbian Anti-Violence Project, Inc., NewYork, NY).
Paul MJ, Dredze M. (2011). You are what you tweet: Analyzing twitter for public health. Icwsm. 20, 265–272.
Prier KW, Smith MS, Giraud-Carrier C, Hanson CL. (2011). Identifying health-related topics on twitter. In Social Computing, Behavioral-Cultural Modeling and Prediction. J Salerno, SJ Yang, D Nau, SK Chai, eds. (Springer, Berlin, Heidelber), pp. 18–25.
Scanfeld D, Scanfeld V, Larson EL. (2010). Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control. 38, 182–188.
Schwartz HA, Ungar LH. (2015). Data-driven content analysis of social media A systematic overview of automated methods. Ann Am Acad Pol Soc Sci. 659, 78–94.
Sedrak MS, Cohen RB, Merchant RM, Schapira MM. (2016). Cancer communication in the social media age. JAMA Oncol. 2,822–823.
Sinnenberg, L., Buttenheim, A.M., Padrez, K., Mancheno, C., Ungar, L., and Merchant, R.M. (2017). Twitter as a tool for health research: a systematic review. Am J Public Health. 107, e1–e8.
Sorenson SB, Shi R, Zhang J, Xue J. (2014). Self-presentation on the web: Agencies serving abused and assaulted women. Am J Public Health. 104, 702–707.
Sotirovic M. (2003). How individuals explain social problems: The influences of media use. J Commun. 53, 122–137.
Violence against women. (2017, August 7). Retrieved August 7, 2017, from https://www.who.int/news-room/fact-sheets/detail/violence-against-women
Wang X, Zhao K, Street N. (2014). Social support and user engagement in online health communities. In International Conference on Smart Health. X Zheng, D Zeng, H Chen, Y Zhang, C Xing, DB Neill, eds. (Springer, Cham), pp. 97–110.
Webb B. (2011). Unsportsmanlike conduct: Curbing the trend of domestic violence in the national football league and major league baseball. Am UJ Gender Soc Pol'y & L. 20, 741–763.
Weeg C, Schwartz HA, Hill S, et al. (2015). Using twitter to measure public discussion of diseases: A case study. JMIR Public Health Surveill. 1, e6.
Williams ML, Burnap P, Sloan L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. Br J Criminol. 57, 320–340.
Yamamoto S, Satoh T. (2013). Two phase extraction method for extracting real life tweets using lda. In Web Technologies and Applications. Y Ishikawa, J Li, W Wang, R Zhang, W Zhang. (Springer, Berlin, Heidelberg), pp. 340–347.
Xue J, Lin K, Sun IY, Liu J. (2018). Information communication technologies and intimate partner violence in China. Int J Offender Ther Comp Criminol. 62, 4904–4922.
Zhao WX, Jiang J, Weng J, et al. (2011). Comparing Twitter and Traditional Media Using Topic Models. Advances in Information Retrieval. (Springer, Berlin), pp. 338–349.
Zhao W, Chen JJ, Perkins R, et al. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinform. 16 (Suppl 13), S8.

Address correspondence to:
Richard Gelles, PhD
School of Social Policy &Practice
University of Pennsylvania
3815 Walnut Street, Room 201
Philadelphia, PA 19104

E-mail: gelles@sp2.upenn.edu