Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter

Jia Xue, PhD, Junxiang Chen, PhD, and Richard Gelles, PhD

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” (Harrald 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, 2015.
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 on issues 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 Twitter posts 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; Heavilin 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) or contiguous sequences of n words (n-gram), 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.
DOMESTIC VIOLENCE TOPICS ON TWITTER

constructing an association between the topics and aspects. Combining LDA for clustering large amounts of documents 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 Twitter 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").

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 function provided in the scikit-learn package.

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 class provided in the scikit-learn package), and compute the rate of perplexity change (RPC) as introduced by Zhao and 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

![FIG. 1. RPC against the number of topics.](image)


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 are 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,” “awareness month,” “victims domestic,” “stop domestic,” and “ronda rousey.” 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 reweet, 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 @WeNeedFeminism: Domestic violence hotline: 1-800-799-7233 #StopDomesticViolence https://t.co/USSenate passed resolution supporting the goals and ideals of National Domestic Violence Awareness Month: https://t.co/YSB1wij8cJ #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, as shown in Table 3. For

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

7Rhonda 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.

8We presented one or two examples under the identified topics.
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 individuals 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 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. 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, 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%.

---

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

10Jose 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 1. Top 20 Popular BiGrams (Pairs of Words)**

<table>
<thead>
<tr>
<th>Popular bigrams</th>
<th>Dataset (%)</th>
<th>Popular bigrams</th>
<th>Dataset (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic violence</td>
<td>10.12</td>
<td>Violence victims</td>
<td>0.29</td>
</tr>
<tr>
<td>Violence awareness</td>
<td>0.88</td>
<td>Jose reyes</td>
<td>0.25</td>
</tr>
<tr>
<td>Greg hardy</td>
<td>0.82</td>
<td>Rt ronda</td>
<td>0.24</td>
</tr>
<tr>
<td>Rt domestic</td>
<td>0.62</td>
<td>Rt stop</td>
<td>0.24</td>
</tr>
<tr>
<td>Awareness month</td>
<td>0.43</td>
<td>William gay</td>
<td>0.21</td>
</tr>
<tr>
<td>Victims domestic</td>
<td>0.40</td>
<td>Support domestic</td>
<td>0.20</td>
</tr>
<tr>
<td>Hardy domestic</td>
<td>0.34</td>
<td>Violence charges</td>
<td>0.19</td>
</tr>
<tr>
<td>Stop domestic</td>
<td>0.32</td>
<td>Arrested domestic</td>
<td>0.18</td>
</tr>
<tr>
<td>Violence incident</td>
<td>0.32</td>
<td>Awareness domestic</td>
<td>0.18</td>
</tr>
<tr>
<td>Ronda rousey</td>
<td>0.31</td>
<td>Alleged domestic</td>
<td>0.18</td>
</tr>
<tr>
<td>Topic</td>
<td>Topic components bigram</td>
<td>Distribution (%)</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Domestic violence, like domestic, rape domestic, 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</td>
<td>2.68</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>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 awareness, violence survivors</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Domestic violence, rape victims, rape domestic, relatives births, relatives deaths, violence rape, violence victim, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>7.50</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Domestic violence, violence charges, rapists domestic, rapists victims, rape victims, rape domestic, relatives births, relatives deaths, violence rape, violence victim, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>6.68</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Domestic violence, violence rape, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths, violence rape, violence victim, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>5.77</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Domestic violence, purple cleats, act domestic, bring attention, mother killed, cleats bring, gay power, awareness, violence rape, 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, rap tweet, feminists mentions</td>
<td>4.98</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>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</td>
<td>5.04</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Domestic violence, violence awareness, hazem el, el masri, alleged domestic, victims 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</td>
<td>5.35</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Domestic violence, violence incident, alleged domestic, relatives births, relatives deaths, violence rape, violence victim, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>5.25</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>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</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Domestic violence, purple cleats, act domestic, bring attention, mother killed, cleats bring, gay power, awareness, violence rape, 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, rap tweet, feminists mentions</td>
<td>4.64</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Domestic violence, violence awareness, awareness month, violence awareness, awareness month domestic, violence awareness, raise awareness, violence rape, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>4.39</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Domestic violence, men domestic, women men, violence join, stand women, participate purplethursday, join participate, issues domestic, people upset, violence horrific, horrific crimes, nfl issues, issue cam, cam likes, likes dance, crimes people, dance thing, women domestic, violence affects, end domestic</td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Domestic violence, william gay, purple cleats, act domestic, bring attention, mother killed, cleats bring, gay power, awareness, violence rape, raise awareness, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>3.87</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Domestic violence, violence awareness, raise awareness, 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, rap tweet, feminists mentions</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Domestic violence, victims domestic, manziel nfl, johnny manziel, nfl contacts, contacts police, experience domestic, inspired help, violence inspired, women experience, child abuse, music video, violence don, anti domestic</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>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</td>
<td>3.01</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Domestic violence, violence charges, rape domestic, relatives births, relatives deaths, violence rape, violence victim, violence rape victims, rape victims, rape domestic, relatives births, relatives deaths</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>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</td>
<td>5.04</td>
<td></td>
</tr>
</tbody>
</table>

**Total:** 100%
between male athletes and their assault toward wom-"..."
### Table 4. BiGrams Distributions Under Topics (Top 3 Presented)

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

**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).
that may produce different topics and themes. Our study suggests
Future studies that cover Tweets for a longer period of time
advocacy-relevant Tweets than other months of the year.
ence Awareness month, in which we expect to see more
period of 3 months. October is the National Domestic Vio-
the data collection lasted from October to December for a
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 pro-
grams, 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 mes-
ages/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 prom-
ising 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 in-
vestigate 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.,
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.
Partner and Sexual Violence Survey. (Centers for Disease Control
and Prevention, Atlanta, GA.)
Brenner J, Smith A. 72% of online adults are social networking site
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
Chew, C., and Eysenbach, G. (2010). Pandemics in the age of Twitter:
content analysis of Tweets during the 2009 H1N1 outbreak. PhoS
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.
language on twitter predicts county-level heart disease mortality.
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, 
Proceedings of the 22nd International Conference on World Wide Web
(ACM), pp. 593–596.
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.
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.


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