Uncovering Multiple domains in Social Movement

——investigating the #Metoo movement with an NLP approach

Jielu He

05/27/2020

We have come a long way from the times when women were classified as militants and arrested for demanding voting right. Today, in just over two years, the #Metoo movement has made us question the sexual engagement and encouraged us to become aware of other's experience. This movement, initially voiced by sexual harassment survivor and activist Tarana Burke, is an international movement against sexual harassment and sexual assault. It went viral as a hashtag on social media, following the exposure of the widespread sexual-abuse allegations against Harvey Weinstein in October 2017. After the rapid expansion of this movement, the scope of #Metoo has become somewhat broader with this expansion, it has been more recently referred as an international movement for justice for marginalized people in marginalized communities.

The goal of this paper is to answer these questions: Can a social movement such as #MeToo be analyzed using Twitter-based User Generated Content(UGC), and would such analysis reveal active agents that influence its development? Is it possible to analyze Twitter users' feelings related to and expressed in relation to #MeToo? I use a text-mining methodology that extracts tweet data from twitter to analyze cultural background and sentimental information behind the Metoo movement. I then explore the twitter data using Natural Language Analyses with NLTK and classify the test data into one the three categories (positive, neutral, and negative). During my exploration of the tweet data, I find that the movement has extended into other semantic domains outside of sexual assault such as "white supremacy" and "police violence". I used colonial theory to explain colonial ideologies and Eurocentrism and looked at literature about violence against Indigenous women and social media.

1 Data

My total sample data combines two datasets: one dataset contains approximately 390,000 tweets (downloaded from data.world) and another dataset contains 13,487 tweets (crawled from twitter API). Only two columns are kept in the datasets: 'text' and 'dataoftweet'.

I first extracted a total of 13,487 tweets and merge the dataset with the existed dataset; after preprocessing, the final dataset contains 89,109 tweets. These tweets were in English and had #MeToo as the hashtag.

After data collection, I filtered and cleaned the data using NLTK package in Python. The initial sample was then cleaned to remove repeated tweets and retweets. Tweets with pure numbers were also deleted. Of note, tweets can contain either a hashtag (#) or a reference to a user profile (@). The replies were analyzed as independent tweets, considering the expressions related to the same subject under the same hashtag. Images, videos, and multimedia content were not analyzed. Instead, I focused exclusively on text and analyzed it using natural language processing (NLP). For filtering, I eliminated URLs that contained the tweets, as well as emoticons and special symbols that did not include capitalization or punctuation.

Table 1 Description of cleaned data

	dateoftweet	text
0	03/05/2020	suck everyday don t security people infiltrate...
1	03/05/2020	s democrat front runner tonight metoo activist...
2	03/05/2020	housing training pay aggrieved female resident...
3	03/05/2020	gaining traction metoo movement victim patel s...
4	03/05/2020	onair chill take midol democrat pollster kil

2 Methods

2.1 Topic modeling with LDA

Topic Modeling is a technique to extract the hidden topics from large volumes of text. In order to identify the topics, I used the Latent Dirichlet Allocation (LDA). LDA is a popular algorithm for topic modeling with excellent implementations in the Python's Gensim package. It processes many "documents" and outputs the most relevant words for each topic. Using the LDA model, in the first step, my objective was to identify keywords in the database. Each word was encoded in a separate document. Then, using the identified keywords, I performed an identification of the topics.

2.2 Text analysis

After topic identification, I use TF-IDF to identify the most frequently used terms in the downloaded tweets.TF-IDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. The tf–idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general. Based on tf-idf value, I then drew a bar chart and a word cloud to show the most salient phrases in Metoo movement.

2.3 Sentiment analysis

Finally, I used an algorithm developed in Python to classify tweets according to sentiments expressed in them. The Sentiment Analysis (SA) approach, also known as Opinion Mining (OM), relies on using the systems that can detect the feelings expressed in a text or textual content—in our case, the tweets with the hashtag #MeToo.

3 Results

3.1 Identification of topics with the LDA model

After the analysis with the LDA model, five topics linked to the #MeToo movement were identified (see Table 2). During this process, the keywords corresponding to each of the sentiments in topics were categorized. On finding 10–20 keywords, these words were used to formulate the topics. In this process, to the goal was to formulate a phrase that contained 10 most repeated words and try to order them to make sense. Thus, the name of the topic was the result of the combination of these phrases and the formulated content.

The identified topics provide important insights about user behavior and their perceptions of the #MeToo movement. In particular, I find #BlackLivesMatter, #MMIW and #notinvisible in the topic description. The # MMIW and #notinvisible are specifically created for Indigenous women but still receive almost no attention, despite the high rates of violence against Indigenous women. While #Metoo is now generally raise awareness of sexual harassment, it also raised a problem related to race inequity.

Table 2 Description of topics in #Metoo tweets

Topic name	Topic Description
Public Figures	Tarana Burke started the movement that was later supported by Alyssa Milano and that reached RealDonaldTrump or Melania Trump against defendants such as Thomas Frieden or Harvey Weinstein.
Sexuality	Survivors of sexual harassment who prosecute their abusers
Female Topics	Women stand together for feminism, females, girlfriend
Media	The stories across the Internet, Hollywood actresses, media and the news
Other Hashtags	The #MeToo movement's relation to #Timesup, #BlackLivesMatter, #MMIW, #GenderEquality,#notinvisible and #WomanEqualityDay

3.2 Text Analysis Results

I calculate the frequency of Bi grams to further illustrate the importance of certain phrases used in tweets with #Metoo hashtag. For example, "police violence" was mentioned 383 times in 15k tweets, while "white supremacy" was mentioned 381 times (See in Fig 1). The #MeToo movement can be interpreted as a feminist movement to protect sexual harassment survivors. Clearly, the scope of this interpretation is limited. According to my exploration about tweet data, people from different gender and race groups shared the fact that they have been the victim of unwanted sexual aggression or police oppression. The scope of this movement has changed and expanded to a broader field.

In the wake of the initial accusations against Hollywood producer Harvey Weinstein, the hashtag #MeToo—the name of a movement launched 10 years earlier by writer and activist Tarana Burke, also a black woman—helped to harness collective outrage into a force that is reshaping workplaces around the country. This movement is the result of the collective labor of women of color who turned private agonies into public battles on behalf of justice.

However, there's an unsettling reality that a movement built largely on the labor of women of color has been co-opted by a discussion that prioritizes the experiences of victims who are white, wealthy, and privileged over those who are not(Angela Onwuachi-Willig,2018).When #MeToo entered the national consciousness, there were questions about whom the movement was really for. Headlines were largely dominated by stories of white, wealthy, straight, cisgender women, even though rates of sexual violence are disproportionally higher for poor women, women of color and LGBTQ people.

Racism and sexism are inextricably intertwined. Navigating a biased system exacts a toll, from lost career opportunities to the energy expended on internal calculations for dealing with inappropriate behavior and self-doubt. Many writers, activists, and scholars have worked tirelessly to highlight the cumulative impact of these experiences in the context of race and gender alike. But the justified outrage around sexual harassment has eclipsed the discussion on race while borrowing its language.

This kind of white supremacy can be explained by Stoler's theory about gender, race and colonial politics (Stoler,1991). The US has a long and sordid history of taking the labor of people of color for granted. When we attempt to explain the behavior of separating gender from race and class in MeToo movement, it will be helpful if we can understand how European culture and class politics resonated in colonial settings, how class and gender discriminations were transposed into racial distinctions.

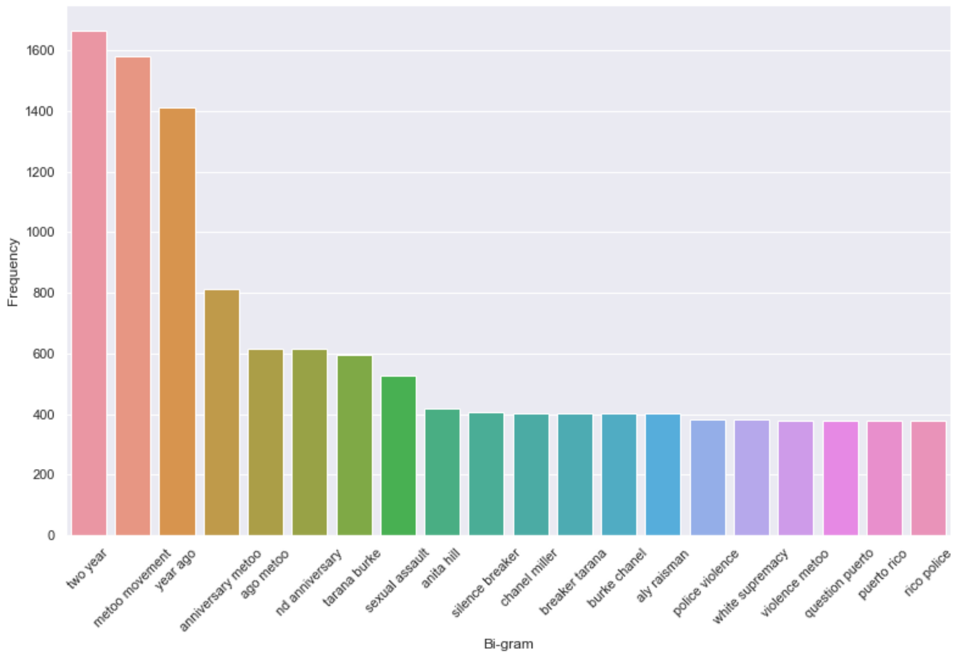

Fig 1 Most salient phrases in #Metoo Tweets

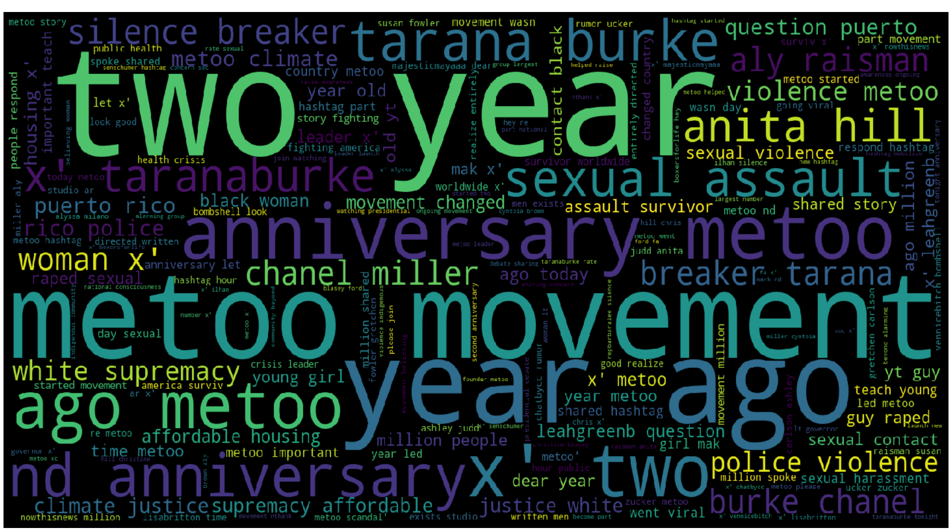

Fig 2 Word cloud #Metoo Tweets

3.3 Sentiment Analysis Results

On identification of the sentiments in the complete sample of tweets, Table 3 shows the date of tweet, corresponding words, sentiment score, and sentiment. The next step will be calculating the sentiment score for each identified topics and salient words.

Table 3 sentiment score of #Metoo tweets

	dateoftweet	text	sentiment	sentiment_score
0	03/05/2020	suck everyday don t security people infiltrate...	negative	-0.1280
1	03/05/2020	s democrat front runner tonight metoo activist...	positive	0.4404
2	03/05/2020	housing training pay aggrieved female resident...	negative	-0.1027
3	03/05/2020	gaining traction metoo movement victim patel s...	negative	-0.4939
4	03/05/2020	onair chill take midol democrat pollster kil	neutral	0.0000

Bibliography

Bekafigo, M. & McBride A. (2012). Who Tweets About Politics? Political Participation of Twitter Users During the 2011Gubernatorial Elections. Social Science Computer Review, 31(5), 625-643.

Blaut, J. M 1993. The Colonizer's Model of the World. New York, NY: Oxford University Press. Ch. 1 (pp. 1-49).

Bubar, R., & Thurman, P. J. (2004). Violence Against Native Women. *Social Justice*, *31*(4 (98)), 70–86.

Cole, N.L. (2019, April 25). Defining Racism Beyond its Dictionary Meaning. Retrieved from https://www.thoughtco.com/racism-definition-3026511

Cram, L., Llewellyn, C., Hill, R., Magdy, W. (2017). UK General Election 2017: A Twitter Analysis. Cornell University.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social order analysis. Behavior Research Methods 49(3), pp. 803-821.

Deer, S., Clairmont, B., Martell, C.A., White Eagle, M.L. (2007). *Sharing our stories of survival: Native women surviving violence.* Lanham, MD: Altamira Press.

Everett, A. (2004). On Cyberfeminism and Cyberwomanism: High-Tech Mediations of *Internet* Feminism's Discontents. *Signs*, *30*(1), 1278–1286.

Gilpin, L. (2016, June 6). Native American Women still have the highest rates of rape and assault. *High Country News*

Kasana, M. (2014). Feminisms and the Social Media Sphere. *Women's Studies Quarterly*, *42*(3/4), 236–249.

Keller, J. (2016). Making Activism Accessible: Exploring Girls & Blogs as Sites of Contemporary Feminist Activism. In Girlhood and the Politics of Place (pp. 261– 278).

Stoler, Ann Laura. "Carnal knowledge and imperial power." Gender at the Crossroads of Knowledge (1991): 51-101.