Mining Twitter Data to Identify Topics of Discussion by Indian Feminist Activists

Nabamallika Dehingia, Anita Raj
Center on Gender Equity and Health, University of California San Diego, USA

Background

Over the past decade, social media platforms have become ubiquitous, serving as a democratic space for activism and providing new opportunities for social movements.[1,2] Twitter has emerged as a popular tool used by feminist activists for spreading awareness and organizing.[3,4] Research examining feminist movements on social media have highlighted the role of Twitter in emphasizing issues related to gender-based violence (GBV) victimization including the #MeToo movement,[5-8] as well as calling out male privilege and regressive gender norms.[9,10] Scholars have examined the high levels of engagement in Twitter discussions and debates by grassroots feminists, as well as the effect of this activity on advancing the feminist agenda in the digital space and amplifying minority voices.[11,12]

Studying Twitter conversations of feminist activists can help identify gender issues that need attention but are underprioritized politically. This brief presents findings from our analysis of a corpus of tweets by 59 Indian feminist activists, tweeted between March and August 2020. The analysis examines how the feminist community in India has used Twitter as a tool for activism during the COVID-19 pandemic. In addition to providing insights related to mainstream gender issues in India, this analysis hopes to contribute to methodological advancement in gender research.

Key Insights

- Tweets by feminist activists in India during March–August 2020 focused broadly on issues related to gender, with a significant emphasis on gender-based violence during the pandemic.

- Frequency of tweets on COVID-19 remained consistent over time. Although, other topics related to women's rights and political rights saw striking peaks and drops, with each peak corresponding to a related real-world event.

- Most observed peaks were also associated with the use of specific hashtags, highlighting the power of effective social media campaigns in spreading awareness around issues of importance.

Our Approach

Sampling and data extraction: We first asked three India-based gender researchers to provide a list of Indian feminist activists who are active on Twitter. These accounts were further screened for our inclusion criteria: a) minimum 1000 followers, b) user accounts from India, and c) an indication of commitment to feminist goals in their Twitter account description. Next, we screened the Twitter accounts followed by each of the eligible users and identified those that satisfied our inclusion criteria. Overall, we included a total of 59 Indian feminist activist user accounts in our analysis. Using the official Twitter API, we collected from these identified user accounts 69,198 tweets between March 25, 2020 and August 14, 2020.

Analysis: We pre-processed the unstructured Twitter data and deleted uninformative parts such as HTML tags, tags of other user accounts, special characters, punctuations, stop words, and emoticons.

To identify distinct topics of discussion from the dataset of tweets, we used Latent Dirichlet Allocation, a common method for topic-modeling.[13] In this case, each individual tweet was represented by a probability distribution of topics, and each topic as a probability distribution of words. For example, one tweet can be classified as belonging to two topics A and B, with Topic A's contribution to the tweet being 80% and Topic B's contribution being 20%. Topic A here is the "dominant topic." In our analysis, each tweet was categorized as belonging to the "dominant topic."

By examining the words corresponding to each topic, we can interpret what the topic signifies. For example, a topic represented by the words "basketball," "soccer," and "football" would refer to sports. The qualitative annotation

of the topics identified from our LDA model was done through a process of discussion and deliberation among the co-authors, based on the top 30 most prevalent words for each topic. One co-author also studied the top tweets corresponding to each topic to get a better contextual understanding of the words. The temporal trend of topics was examined by assessing the number of tweets per day for each topic.

Results: The most tweeted hashtag was #COVID, followed by #India. Multiple hashtags on gender-based violence were identified: #DomesticViolence, #ViolenceAgainstWomen, and #AkhirKyon. The hashtag #AkhirKyon is a Hindi expression which translates to "But why?" and was used by Twitter users to generate awareness about domestic violence in India. Other hashtags related to gender issues included #TheSexEdChallenge, #WomenInSTEM, #BodyImage, #MeToo, #SexGenTech. The hashtag #SexGenTech was used in discussions on gender inclusivity of current technologies.

Figure 1: Word cloud of hashtags in the corpus of tweets

Hashtags related to Caste (the marker of social class in India) — #TheyInspireMe, #DalitHistoryMonth, and #DalitWomenFight — were also among the most tweeted hashtags. Dalits are the most marginalized caste group in India based on traditional caste hierarchy. The hashtag #TheyInspireMe was used by feminist activists to tweet inspiring stories about Dalit women in India.
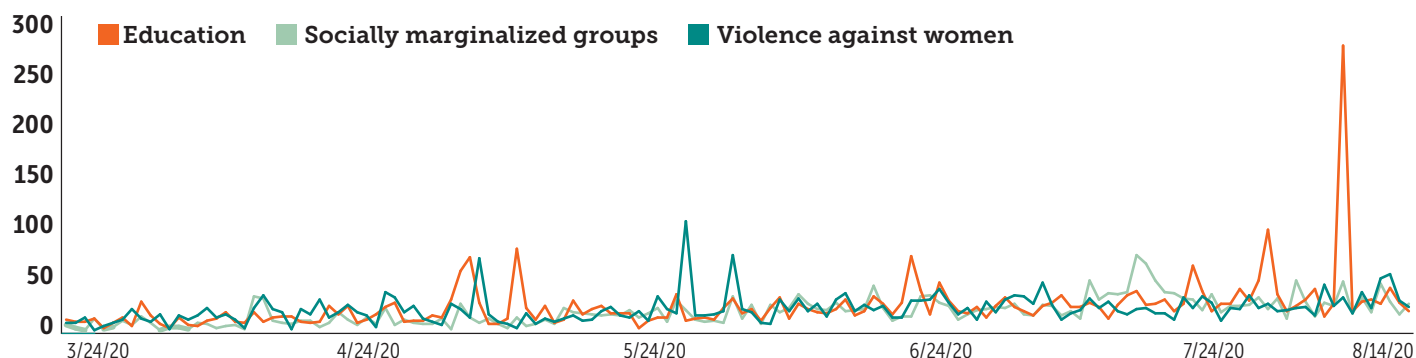
The number of topics to be extracted using the LDA model was decided based on the Coherence score of models.[14] Topic coherence score measures the degree of semantic similarity between high scoring (high probability of occurrence) words in the topic. The score helps distinguish between topics that are semantically interpretable topics, and topics that are artifacts of statistical inference. The higher the Coherence score, the higher the similarity in the words within each topic. Our final model specified 18 topics across our sample of tweets. The topic covering the highest number of tweets was on migrant worker crises during the COVID-19 lockdown. We labeled three other topics as being related to COVID-19: tweets about COVID-19 related health services, COVID-19 prevalence, and the COVID-19 vaccine.

Three topics relevant to gender issues were identified: 1) violence against women, which included tweets on the rising cases of domestic violence during the pandemic, sexual violence, as well as religion and caste-based violence; 2) education, which focused on sex education, and gender gaps in education and technology; and 3) feminism in the context of socially marginalized groups (minority groups based on sexual orientation, gender identity, and caste). Tweets corresponding to the topic violence against women included informational posts with helpline numbers for domestic violence victims, as well as posts meant to raise awareness about the issue of gender-based violence in India. The topic on education covered posts that aimed to raise awareness about the importance of sex education and an understanding of consent among young boys. It also included posts on challenges faced by students with regards to digital education during the COVID-19 lockdown. Our sample also identified multiple topics on political issues as well as on personal situations and feelings/emotions.

Figure 2: Number of tweets by the 18 topics identified through topic modelling

Topic	Number of tweets
Migrant worker crises during COVID-19 lockdown	7022
Posts expressing feelings	5240
Stories and inspirational entertainment	4122
Personal situations	3852
Use of social media	3847
Webinars and events related to gender	3780
COVID-19 related health services	3750
Protest against political arrests	3710
Education	3687
Civil unrest	3532
Violence against women	3320
Response to political environment	3102
COVID-19 vaccine	3079
Socially marginalized groups	2979
Generic posts	2689
Feminism and data	2679
Economy	2596
COVID-19 prevalence, spread, and deaths	2392

Figure 3: Number of tweets per day by topics related to gender

On examining the temporal trends of topics, we observed that the number of tweets for COVID-19 related topics remained consistent overall. However, we observed striking peaks over time for most other topics (temporal trends for gender related topics shown in Figure 3). The topic on violence against women peaked on the first week of May. A qualitative analysis of the tweets corresponding to this topic showed that this surge was due to an increase in tweets related to domestic violence, perhaps as a result of growing concerns globally about increasing rates of domestic violence during the pandemic. A social media campaign initiated by feminist organizations, using the hashtag #AkhirKyon, aimed to highlight the issue of domestic violence and sexual violence. The topic on education peaked significantly around late July, which again was due to a social media campaign on the importance of sex education for children in India, using the hashtag #TheSexEdChallenge. The first noted increase in this topic was observed around early May, coinciding with an incident in India where a group of young male school students leaked obscene images of their female classmates online. This sparked Twitter conversations on the lack of understanding of consent, gender norms, and sexuality among young boys in India. The topic on feminism in the context of socially marginalized groups included tweets related to transgender and non-binary people; it peaked around late June due to several tweets related to Pride Month, which is celebrated in June. The peaks for other topics related to civil unrest, political environment, and inspirational stories and events were also observed to be associated with specific hashtags.

Implications

Twitter data mining has been shown to be an effective method for understanding public perceptions regarding different issues through a gendered lens. Our study contributes to this literature, with a focus on feminist activist Twitter users in India. We find that Twitter is being used in India as a platform for real-time activism; popular topics of discussion in Twitter coincided with key offline events of importance in our analysis. For example, as a response to the increasing rates of domestic violence during the pandemic in India as well as globally, we observed large number of posts/tweets on violence against women by Indian feminist activists, using hashtags #AkhirKyon and #ViolenceAgainstWomen. The arrests of certain activists also led to a Twitter storm demanding the release of these prisoners; with most tweets including similar hashtags (#FreeKalyani, #FreeTanmay). However, the topics of discussion were not limited to issues related to feminism and human rights alone. It included topics on personal experiences and situations as well. Our analysis also identified topics that emphasized minority issues, such as the marginalization of transgender people and Dalits in India. In addition to tweets that discussed rights of the socially marginalized groups, our analysis also identified a topic that celebrated the history and achievements of the groups; we observed large number of tweets with the hashtag #TheyInspireUs, which covered inspiring stories about Dalit women in India. Our findings show that social media, particularly Twitter, can act as a democratic platform, highlighting minority voices.

We focused on Twitter posts during the pandemic, hence tweets related to COVID-19 and the lockdown saw greater representation. While the frequency of posts on COVID-related topics remained consistent over time, significant peaks and drops were observed for other topics, mostly as a reaction to real-world events. Observed peaks for the topics were associated with the use of specific hashtags, such as #ViolenceAgainstWomen, #WomenInSTEM, #TheSexEdChallenge, and #SexGenTech. These were some of the most tweeted hashtags identified in our analysis and were used to underscore critical gender issues such as gender-based violence, women's lacking representation in science and technology, importance of sex education, and the existing gender bias in technology. Our findings highlight the power of effective social media campaigns in spreading awareness around issues of importance.

References

1. George JJ, Leidner DE. From clicktivism to hacktivism: Understanding digital activism. Information and Organization 2019; 29(3): 100249.

2. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. Journal of Public Health 2020.

3. Losh E. Hashtag feminism and Twitter activism in India. Social Epistemology Review and Reply Collective 2014; 3(3): 11-22.

4. Rodriguez S, Allende-Cid H, Gonzalez C, et al. Analyzing# LasTesis Feminist Movement in Twitter Using Topic Models. 2020: Springer. p. 624-35.

5. Manikonda L, Beigi G, Liu H, Kambhampati S. Twitter for sparking a movement, reddit for sharing the moment:# metoo through the lens of social media. arXiv preprint arXiv:180308022 2018.

6. Li M, Turki N, Izaguirre CR, DeMahy C, Thibodeaux BL, Gage T. Twitter as a tool for social movement: An analysis of feminist activism on social media communities. Journal of community psychology 2020.

7. Gill R, Orgad S. The shifting terrain of sex and power: From the 'sexualization of culture'to# MeToo. Sexualities 2018; 21(8): 1313-24.

8. Keller J, Mendes K, Ringrose J. Speaking 'unspeakable things': Documenting digital feminist responses to rape culture. Journal of Gender Studies 2018; 27(1): 22-36.

9. Kim J. # iamafeminist as the "mother tag": feminist identification and activism against misogyny on Twitter in South Korea. Feminist media studies 2017; 17(5): 804-20.

10. Retallack H, Ringrose J, Lawrence E. "Fuck your body image": Teen girls' Twitter and Instagram feminism in and around school. Learning bodies: Springer; 2016: 85-103.

11. Barnard Center For Research On Women. #FEMFUTURE: ONLINE REVOLUTION. 2013.

12. Zimmerman T. # Intersectionality: The Fourth Wave Feminist Twitter Community. Atlantis: Critical Studies in Gender, Culture & Social Justice 2017; 38(1): 54-70.

13. Tong Z, Zhang H. A text mining research based on LDA topic modelling. 2016. p. 21-2.

14. Syed S, Spruit M. Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. 2017: IEEE. p. 165-74.

The complete code and detailed methodology can be found here: https://github.com/mdehingia/TwitterGenderTopics