

Technology and Collective Action Event Size: Lessons for India

William O'Brochta¹

Published in *Studies in Indian Politics*

An event is a geographically and temporally focused 'nonroutine, collective, and public' action aimed at making broad social or political claims (Olzak, 1989, p. 124). Event attendance matters for how the event is perceived and reported on by citizens, the media, and political leaders (Biggs, 2018; McPhail & McCarthy, 2004). Scholars of Indian politics have measured event size with proxies: using deaths, injuries, or the number of First Information Reports (FIRs) filed to quantify riot size and relying on newspaper-based estimates of the size of other collective action events (Bhavnani & Lacina, 2015; Varshney & Wilkinson, 1996). Is there a way to improve our event size estimates in order to more effectively study this critical indicator of event success and impact?

Political scientists have recently developed methods for detecting event size using data from technology-based sources, including geolocation, social media activity, and social media photos. In some contexts, these methods comport well with existing newspaper-based measures of event size, and their automated nature reduces cost and eases implementation compared to traditional newspaper searches (Botta, Moat, & Preis, 2015).

In this note, I apply technology-based event size measures to Indian events. I argue that the way that events develop in India makes using technology-based data to estimate event size

¹ Department of Political Science, Washington University in St. Louis, St. Louis, Missouri 63130. obrochtawj@wustl.edu, williamobrochta.net.

particularly difficult. I take as my case the 2019 and 2020 protests surrounding the Citizenship Amendment Act (CAA) and related causes in Delhi. These protests were highly publicized, politicized, and contested, with public campaigns waged on social media. Drawing on fieldwork at key protest sites, a review of social media data, and media reports, I find that strategic choices on behalf of both protesters and the government made estimating event size using technological data quite challenging. I then discuss some ways that event size measurement techniques can be adapted for the Indian context.

New Approaches to Measuring Event Size

Conventional approaches to estimating event size start with scholars defining the scope of events whose size should be measured. These guidelines help to determine the appropriate empirical research method to employ in order to identify all relevant events and to produce an event size estimate. The most accurate approaches rely on data collected by individuals observing the event, either in real time or using historical records. Common methods include stationing researchers at the event to count participants, using airplanes or drones to capture pictures of the event, or using government data based on police actions during the event (e.g., Schweingruber & McPhail, 1999). The first two methods are expensive and resource intensive, while the third method provides a proxy for event attendance by counting arrests, injuries, deaths, or police reports. Three technology-based methods have emerged that purport to directly measure event size like the first two conventional approaches, but with ease of collection similar to using government data. These methods are: geolocating mobile phone positions, quantifying social media posts, and counting attendees using social media images.

Mobile phone geolocation works by triangulating the user's position relative to mobile phone towers in the area (Mamei & Colonna, 2016). Such data is not available to scholars studying India because the Indian Unified Access Service License agreed to by each licensed telecom provides the government and only the government with location details of all mobile phone customers.² What is more, using mobile phone geolocation is necessarily inductive: scholars must know that an event is happening in order to attempt to measure the event's size, as not all large crowds signify the presence of an event.

Social media data can be used to quantify event attendance, and Twitter is the most accessible platform for collecting such data.³ Twitter adoption is widespread in many countries, though there are only an estimated 11.45 million users in India (Hootsuite & We Are Social, 2020). Twitter is thought to have facilitated the organization of a number of recent Indian collective events (Ahmed & Jaidka, 2013). Beyond identifying the existence of an event using Twitter data (Zhang & Pan, 2019), scholars have estimated event size by counting the number of unique Twitter accounts in a given area utilizing event related keywords and hashtags (Larson, Nagler, Ronen, & Tucker, 2019; Sobolev, Chen, Joo, & Steinert-Threlkeld, 2020).⁴

Twitter images may also provide important information about event attendance (Steinert-Threlkeld, 2019). While the number of Twitter posts serves as a proxy for event attendance, Twitter images can be used to count individual event attendees. Researchers can stitch together Twitter images from the event, estimate the crowd size in each image, and sum the estimates to

² Unified Access Service License (2013), Chapter 8, Section 8.2-8.5.

³ WhatsApp messages are encrypted and are not publicly available, meaning that using WhatsApp as a data source is impractical in this context.

⁴ Tweet geolocation was removed in 2019.

estimate the size of the event. Assuming that all relevant images can be collected, this method works if the images on Twitter are all encompassing, showing many perspectives of the entire event without focusing only on certain types of people or scenes (Sobolev, Chen, Joo, & Steinert-Threlkeld, 2020).

Social Media as Event Size Data in India

How well should we expect social media data to capture event size estimates in India? I argue that strategic decisions by both event attendees and the government make it difficult to estimate event size using social media data alone (Gohdes, 2018; Zeitzoff, 2017). First, the Indian government can employ Internet shutdowns, presenting a coordination problem that encourages others not at the event to post online messages in solidarity and for those at the event to delay posting until they return to an area with Internet access (Gohdes, 2015). Additionally, the police response --- including plainclothes and rapid reaction forces, as well as transportation service shutdowns --- prevent many event attendees from reaching a centrally planned location and make it nearly impossible to determine who is actually attending the event and who is just in the area of the event (Jain & Radhika, 2019). From the citizen perspective, Indians are facing increasing uncertainty about the nature of civil liberties (Varshney, 2019). This may encourage individuals to use private social media and to make strategic decisions about what they report on public social media (Zhong, Wang, & Huang, 2017).

To examine social media event data and to assess its usefulness in the Indian context, I rely on fieldwork conducted in December 2019 and January 2020 in Delhi, along with social media posts and media reports from that time. Fieldwork consisted of visiting CAA protest sites before and after events took place daily from December 15 to December 25. I monitored Twitter

activity related to the Delhi protests from December 15 through January 17. My fieldwork objective was to determine the extent to which Twitter posts and images represented the on-the-ground situation at collective action events and, thus, how well Twitter data can be used to estimate event size.

Three main challenges emerged from linking my field observations to social media posts and images. First, Twitter posts and images were not representative of event attendance. From their profile pictures, those who were most active on Twitter during events tended to be young, which is in line with overall Twitter usage trends. However, the demographics of protest attendees were quite diverse, meaning that Twitter posts did not effectively capture the diversity present at protest events. This is important because different types of people are likely to use technology differently, meaning that different event demographics will result in different event size estimates (e.g., Mellon and Prosser, 2017).

It was difficult to ascertain whether a Twitter user was at a protest event or at home discussing the event because multiple protest events were occurring in Delhi at the same time, and users rarely clearly identified that they were physically at a protest. Users Tweeting from home and from elsewhere in India employed the same language, hashtags, and references to Delhi-based locations as users at events, so Twitter users' posts could not be separated based on whether they attended the protest. Indeed, much like recent protests against police brutality in the United States (e.g., Bonilla and Rosa, 2015), Twitter became a space for digital protest for those who were not able to attend in person protest events. Even if we knew who was protesting, protesters used each other's hashtags and images to promote the overall protest movement. For example, protesters located at Red Fort often used the hashtag #JamiaProtest and protesters at Jamia Millia Islamia discussed protests at Red Fort. Since these were distinct protest events, we

need a way of disentangling which Twitter users were located at which protest --- doing so is difficult when reading through Twitter posts.

Further, Tweets that included images of events did so to tell a particular story, not to present a representative picture of the crowd at the event. Images focused on protest leaders, violence, and police action. Based on my observations at protest sites, I knew the approximate locations of police barricades that were used to corral protesters and could therefore conclude that Twitter images did not show the full extent of the protest site. It did not help that most images were close-up shots of a few individuals, not a picture of the full protest crowd. For example, two images showed similarly large crowds, but one claimed that there were hundreds of event attendees while the other claimed that there were thousands of event attendees. It is impossible to know which estimate is correct because images do not systematically capture event crowds. From a technical standpoint, identifying the number of people in images was made more difficult by the fact that many protests took place at night, so image quality and lighting were both poor.

Second, the police response impacted how, where, and when event attendees were able to Tweet about the event. Police used a mix of containment and detainment tactics depending on whether the event was authorized, the nature of the protest, and whether the police perceived protesters as a threat. In some cases, protesters were allowed to continue the protest unimpeded by police. This provided some freedom for attendees to post Tweets and images from the protest. Additionally, police sometimes worked to separate event attendees from passersby by closing and blocking streets and forming human barricades around protesters, making it easier to accurately count event attendees using images from Twitter. In other cases, police reacted quickly and detained protesters as they arrived, put them on busses, and transported them to

various parts of the city. When coupled with shutting down public transportation stations and installing checkpoints on some highways, this meant that people trying to gather at a centrally coordinated location were unable to do so. Event attendance was dynamic: some protest sites saw a constant stream of people coming and going, with alternate protest locations popping up where people were stuck at transportation hubs and where protesters were taken when they were detained. Because of the fluid nature of protest attendance, images never captured the full set of people who attended the protest, and it was difficult to isolate the specific areas where protests were occurring at any given time in order to estimate attendance based on Twitter posts.

Third, the Internet shutdown that occurred on December 19, 2019 made estimating event attendance using social media data essentially impossible. Protesters did provide information to friends who posted it on Twitter, but few Tweets from protest sites were posted until after the shutdown ended or until people got out of the shutdown area. This period saw a spike in the use of peer-to-peer chat apps that work over Bluetooth, so protesters were communicating with each other, they just were not communicating live on Twitter (Akolawala, 2019). The Internet shutdown also seemed to drive protester self-censorship, making people less likely to openly admit to being at a protest event or to post an original image from that event and more likely to post a viral picture or video from the protest.

Adapting Social Media Data for the Indian Context

Despite the limitations outlined above, social media data can play an important role in estimating the size of Indian events. The challenge for scholars of India is to carefully contextualize social media data, using field observations and traditional media sources to provide a more complete picture of an event. As such, social media data helps researchers gather many different

perspectives from the event that can then be used to develop a narrative about how the event unfolded and to estimate the size of the event. I describe my approach to using social media data in this way below.

During the CAA protests, I searched Twitter hashtags for upcoming protest events. Often, examining Tweets mentioning frequent protest sites and the topic of potential upcoming protests will reveal hashtags that organizers use to disseminate event information. Event details change, so reviewing Tweets as they are posted allows researchers to quickly get a sense of the number of events being planned, their location and timing, who plans to attend, and the purpose of the event. Next, I visited event sites in person prior to the event and took photographs and notes. The purpose of this field observation was to help inform how I interpreted social media posts during the event. I recorded information about the geography of the space, key landmarks, and how people might enter and exit the event location. I found field observations to be critically important, even if the event location was familiar.

A few hours before an event started, I began watching the event on Twitter, checking event hashtags, taking notes, and downloading images. The advantage of using social media data is that it provides many snapshots of what people are seeing and doing during the event, more perspectives that a single researcher could obtain by attending the event. After the event concluded, I brought together information from my field observations, social media data, and news broadcasts about the event that are often archived on social media sites like YouTube. The goal now is to piece together information about the size of the crowd over the course of the event. News broadcasts --- especially videos looking down on the crowd from a high vantage point --- are useful for determining the boundaries within which the event took place and for providing a rough idea about whether crowd density was uniform. Social media data is then

helpful for estimating crowd density at different locations. By using a variety of data sources, I was able to arrive at a relatively complete understanding of how the event unfolded and the associated crowd.

The simplicity of this approach has several practical benefits. No special software or techniques are needed to conduct the analysis. Incorporating field observations into the approach limits the number of researchers who can conduct this work, but field observations help to ensure that researchers have the proper context with which to interpret social media data. Information about social media users is not stored and is aggregated, reducing privacy and ethical concerns (Williams, Burnap, and Sloan 2017). Most importantly, resulting estimates of crowd size contain appropriate uncertainty and useful information about the event timeline.

In India, where significant collective action events draw large crowds, precise estimates of crowd size are less critical than are other details about the event. Estimating crowd size is one part of understanding collective action events. Indeed, examining social media data can provide information about event organizers' strategies, how the government and police respond, and ways in which event participants and the media establish narratives about the event. All of these pieces of information help researchers to develop a fuller picture of the event and to contextualize the resulting crowd size estimates beyond field observations or social media data alone. Therefore, our study of Indian events improves when utilizing multiple methodological approaches and focusing on how event context may influence the data we collect.

Acknowledgements

I thank Sunita Parikh, Divya Vaid, and the Reviewers for helpful comments.

Funding

The Weidenbaum Center on the Economy, Government, and Public Policy provided funding for this project.

References

- Akolawala, T. (2019). *Delhi Internet shutdown: Bridgefy offline messaging app gains popularity amid CAA protests*. NDTV. <https://gadgets.ndtv.com/apps/news/internet-shutdown-delhi-bridgefy-offline-messaging-app-caa-protests-google-play-store-app-store-2151430>
- Ahmed, S., & Jaidka, K. (2013). Protests Against #delhigangrape on Twitter: Analyzing India's Arab Spring. *eJournal of eDemocracy and Open Government*, 5(1), 28-58.
- Bhavnani, R., & Lacina, B. (2015). The effects of weather-induced migration on sons of the soil riots in India. *World Politics*, 67(4), 760-794.
- Biggs, M. (2018). Size matters: Quantifying protest by counting participants. *Sociological Methods & Research*, 47(3), 351-383.
- Bonilla, Y., & Rosa, J. (2015). #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist*, 42(1), 4-17.
- Botta, F., Moat, H.S., & Preis, T. (2015). Quantifying crowd size with mobile phone and *Twitter* data. *Royal Society Open Science*, 2(5), 150162.
- Gohdes, A. (2015). Pulling the plug: Network disruptions and violence in civil conflict. *Journal of Peace Research*, 52(3), 352-367.
- Gohdes, A. (2018). Studying the Internet and violent conflict. *Conflict Management and Peace Science*, 35(1), 89-106.
- Hootsuite & We Are Social (2020). Digital 2020: India. <https://datareportal.com/reports/digital-2020-india>
- Jain, A., & Radhika, J. (2019). *Status of policing in India report 2019*. Common Cause and Centre for the Study of Developing Societies.

- Larson, J., Nagler, J., Ronen, J., & Tucker, J. (2019). Social networks and protest participation: Evidence from 130 million Twitter users. *American Journal of Political Science*, 63(3), 690-705.
- Mamei, M., & Colonna, M. (2016). Estimating attendance from cellular network data. *International Journal of Geographical Information Science*, 30(7), 1281-1301.
- McPhail, C., & McCarthy, J. (2004). Who counts and how: Estimating the size of protests. *Contexts*, 3(3), 12-18.
- Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of social media users. *Research & Politics*, 4(3), 1-9.
- Olzak, S. (1989). Analysis of events in the study of collective action. *Annual Review of Sociology*, 15: 119-141.
- Schweingruber, D., & McPhail, C. (1999). A method for systematically observing and recording collective action. *Sociological Methods & Research*, 27(4), 451-498.
- Sobolev, A., Chen, K., Joo, J., & Steinert-Threlkeld, Z. (2020). News and geolocated social media accurately measure protest size variation. *American Political Science Review*, 114(4), 1343-1351.
- Steinert-Threlkeld, Z. (2019). The Future of Event Data Is Images. *Sociological Methodology*, 49(1), 68-75.
- Varshney, A. (2019). Modi consolidates power: Electoral vibrancy, mounting liberal deficits. *Journal of Democracy*, 30(4), 63-77.
- Varshney, A., & Wilkinson, S. (1996). *Hindu-Muslim riots 1960-93: New findings, possible remedies*. Rajiv Gandhi Institute for Contemporary Studies.

- Williams, M., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data: Taking into account users' views, online context, and algorithmic estimation. *Sociology*, *51*(6), 1149-1168.
- Zeitsoff, T. (2017). How social media is changing conflict. *Journal of Conflict Resolution*, *61*(9), 1970-1991.
- Zhang, H., & Pan, J. (2019). Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, *49*(1), 1-57.
- Zhong, Z., Wang, T., & Huang, M. (2017). Does the Great Fire Wall cause self-censorship? The effects of perceived internet regulation and the justification of regulation. *Internet Research*, *27*(4), 974-990.