# Geolocated Twitter Panels to Study the Impact of Events

**Han Zhang**
Department of Sociology, Princeton University
Princeton, NJ, 08544
hz2@prineton.edu

**Shawndra Hill** and **David Rothschild**
Microsoft Research
641 Avenue of the Americas, New York, NY, 10011
shawndra,davidmr@microsoft.com

## Abstract

Data from Twitter have been employed in prior research to study the impact of events. Historically, researchers have relied on keyword-based samples of tweets to create a panel of Twitter users that mention event-related keywords during and/or after an event. There are limitations to the *keyword-based panel* approach. First, the technique suffers from selection bias since users who discuss an event are already more likely to discuss event-related topics beforehand; it is unclear whether observed impacts are merely driven by a set of users who are intrinsically more interested in events. Second, there are no viable groups for comparison to a keyword-based sample of tweeters. We propose an alternative sampling approach to studying response to events on Twitter that addresses the aforementioned two issues. We construct panels of users defined by their geolocation. These panels are exogenous to the keywords in users' tweets, resulting in less selection bias than the *keyword-based panel* method. *Geolocated Twitter panels* allow us to follow within-person changes over time and also enable the creation of comparison groups. We evaluate our panel selection approach in two real-world settings: response to mass shootings and TV advertising. We illustrate how our approach limits selection bias introduced by the keyword-based approach, how discussion among the panel of users shifts before and after an event, and how geography can provide meaningful comparison groups regarding the impact of these events. We believe that we are the first to provide a clear empirical example of how a better panel selection design, based on an exogenous variable like geography, both reduces selection bias compared to the current state-of-the-art and increases the value of Twitter research for studying events.