



City Research Online

City, University of London Institutional Repository

Citation: Bastos, M. T., Recuero, R. & Zago, G. (2014). Taking tweets to the streets: A spatial analysis of the Vinegar Protests in Brazil. *First Monday*, 19(3), doi: 10.5210/fm.v19i3.5227

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/16381/>

Link to published version: <https://doi.org/10.5210/fm.v19i3.5227>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



Taking tweets to the streets:
A spatial analysis of the Vinegar Protests in Brazil
by Marco Bastos, Raquel Recuero, and Gabriela Zago

Abstract

In this paper we investigate the relationship between the geographic location of protestors attending demonstrations in the 2013 Vinegar protests in Brazil and the geographic location of users that tweeted the protests. We explored the overlap between different sources of geographic information from Twitter — namely geocode, hashtag, and user profile — provided by multiple samples drawn from a population of three million tweets related to the events and compared the data to the location of protestors attending the street demonstrations. We adjusted the data for the uneven distribution of the population and performed geospatial and spatial clustering analysis over sets of spatial locations. We found evidence for the hypotheses that users tweeting the protests are geographically distant from the street protests and that users from geographically isolated areas rely on Twitter hashtags to remotely engage in the demonstrations.

Contents

- [1. Introduction](#)
 - [2. Previous work](#)
 - [3. Objectives](#)
 - [4. Data](#)
 - [5. Methods](#)
 - [6. Results](#)
 - [7. Discussion and conclusion](#)
-

1. Introduction

The “Vinegar Protests” in Brazil were initially organized to oppose bus and underground fare rises in June 2013. Demonstrations were later expanded to protests against the running costs of infrastructure projects associated with international sport events, such as the Confederations Cup, World Cup, and Summer Olympics (Sandy, 2013). Protestors’ demands included conflicting political agendas that encompassed better public services, less taxes, and adequate welfare benefits. The first large protest was held on 6 June in São Paulo, and on 17 June an estimated quarter million protestors took to the streets of major cities across the country (*O Globo*, 2013). Protest marches turned violent and urban riots were observed in a number of Brazilian cities. The demonstrations were subsequently dubbed “Vinegar Protests” in reference to 60 protestors arrested for carrying vinegar as an antidote to tear gas and pepper spray used by police.

Facebook and Twitter reportedly played an important role in the organization of public outcries, facilitating communication between protestors and live streaming the demonstrations (BBC Brasil, 2013). Geographic information was essential to provide context to readers and activists as protests were taking place in several Brazilian cities simultaneously. During this period, social media aggregation companies started offering Twitter data with location information from Twitter profiles along with geocoded data extracted from GPS-enabled devices (Cairns, 2013). Although the matching of a location in users profile to a location in the real world is often inaccurate, this resource provides another source of geographic information related to where users tweeted their messages.

After 16 June, Twitter users started to tag messages indicating the location where protests were taking place. This unconventional development provided social media messages with a description of the location where protesting activity was happening around the country. Users relied on location-based hashtags to join the protesting activity whether they were physically present at the demonstrations or tweeting from the comfort of their residences. After 17 June, the volume of hashtagged messages in our dataset becomes one order of magnitude higher than non-hashtagged messages and remains higher until the end of the period. Because of that, the Vinegar protests in Brazil offers an opportunity to aggregate different sources of geographic data that capture the relationship between online and onsite activities in the context of political protests.

In this paper we investigate the relationship between the geographic location where protests took place in Brazil and the geographic location of users that tweeted the protests. We compare the location of protestors attending demonstrations in Brazil with different sources of geographic information — namely geocode, hashtag, and user profile — provided by multiple samples drawn from a population of over three million tweets related to the events. By exploring the spatial distances between the location of events and the locations of users (identified by content of user's tweets, profiles, and devices), the results of this study offer an assessment of the interplay between onsite and online political activity. In the next sections we review the literature, detail the objectives of this study, and explain the data collection and the process of data analysis. In the last two sections of the paper we present the results and discuss our findings.



2. Previous work

Research focused on geographic information of Web and social media data has boomed in the past five years. Backstrom, *et al.* (2008) investigated the association of Web search keywords with the geographic location of IP addresses and modeled the spatial variation manifested in search queries. The availability of geographic information of Web search data was also explored by Gan, *et al.* (2008) and Ginsberg, *et al.* (2008), who successfully traced the spread of flu in the United States by correlating the number of visits to a local doctor with flu-based search terms on Google. While earlier literature has focused on modelling information of Web search data, recent research has focused on user-generated geographic information by mining social media information streams.

This latter line of research was made possible by the growing availability of geographic information generated by social media users and GPS-enabled devices. Cheng, *et al.* (2010) used a probabilistic model based on hundreds of tweets to estimate the likelihood of users living in a particular city within a 100-mile radius, while Sakaki, *et al.* (2010) investigated the real-time interaction between onsite events and Twitter stream to monitor tweets and detect events in geographic locations. Noulas, *et al.* (2011) studied urban mobility patterns in several metropolitan areas by analyzing a large set of Foursquare users, and Gao, *et al.* (2012) offered a sociohistorical model to explore user's behavior on location-based social networks.

The contrast between geography and the topology of social networks has been more thoroughly explored in recent years. Volkovich, *et al.* (2012) investigated the interaction between users and spatial distance and found that ties in highly connected social groups tend to span shorter distances than connections bridging separated portions of the network. Cranshaw, *et al.* (2012) measured the social dynamics of a city based on Foursquare check-ins and compared with boundaries of traditional municipal organizational units such as neighborhoods. The clustered check-in areas generated by social media users were considerably different than traditionally defined neighbors.

The emergence of Twitter hashtags in specific locations and the role it plays in public conversations has also been investigated. boyd, *et al.* (2010) described the topic function of hashtags, Huang, *et al.* (2010) explored the conversational nature of Twitter tags, and Bruns and Liang (2012) discussed how Twitter was used by the public to exchange information related to natural disasters in Australia. Although most literature on hashtags focuses on the context rather than the geographic information provided, previous investigations have explored the overlap between hashtag and geographic location. Sloan, *et al.* (2013) used a sample of one million tweets with geographic information retrieved from user profile, geotagged tweets, and the content of the messages to estimate demographic information from the messages.

There is also a growing body of research focused on social and geographic information retrieved from user profiles. Hecht, *et al.* (2011) found that 34 percent of users did not provide real location information, and that when the information was available it almost never specified it at a scale more detailed than the city. Quercia, *et al.* (2012) explored a large sample of Twitter profiles to test whether real-life geography and topic associations hold true on Twitter. Leetaru, *et al.* (2013) retrieved geographic data from tweets based on geocode, profile, and messages and found that geographic proximity played a minimal role both in who users communicate with and what they communicate about, providing preliminary evidence that geographic location is not paramount to the exchange of information in social media.

On the other hand, there is a large body of work stressing the importance of geography to Twitter network. Kulshrestha, *et al.* (2012) investigated the participation of Twitter users and network connectivity and reported that geography had a substantial impact on the interaction between users. Similarly, Takhteyev, *et al.* (2012) examined the influence of geographic distance and national boundaries in the formation of social ties on Twitter and found that a substantial share of ties lies within the same region. Yardi and boyd (2010) investigated the tweets related to two local events and found that the geographic location of tweets and users is important in creating context, providing real-time information, and offering eyewitness accounts to the events.

The literature on social movements has devoted considerable attention to the relationship between social media usage and physical protests (Bennett, *et al.*, 2014; Castells, 2012), the articulation between platforms of self-publication and contentious communication (Castells, 1997, 2009; Diani, 2000; Tarrow, 2005), and the increase in speed and scale of political networks (Bennett, *et al.*, 2008; Bennett and Segerberg, 2013). The literature also explored how social media facilitated organization and horizontal logistical coordination (Theocharis, 2013) and provided a positive setting for the construction of elective social affinities (Papacharissi and Oliveira, 2012).

Previous work have explored the interplay between onsite and online activity and explored the awareness and political participation of social media users (Bekafigo and McBride, 2013; Dimitrova and Bystrom, 2013; Gustafsson, 2012); commented on the role played by Twitter on protest communication (Earl, *et al.*, 2013); and assessed the effect of live-streaming events on public conversations (Hawthorne, *et al.*, 2013; Shamma, *et al.*, 2009).

Researchers have explored how social network sites are instrumental to the rapid formation of a geographically interconnected, networked counter public that we investigated in this study, particularly in movements such as the Indignados in Spain (Vallina-Rodriguez, *et al.*, 2012), the Occupy in the U.S. (Penney and Dadas, 2014), the Kony 2012 campaign (Harsin, 2013), and the political unrest in the countries of the so-called Arab Spring (Lim, 2013).

Although the expansion of mobile communications, global satellite, and the Internet has intensified the preoccupation with the geographic centers of political activism, research on the complex relationships between social media use and geographic location is still forthcoming. In this study we take a deeper look into the relationship between onsite and online political activity by exploring how multiple streams of geographic information reported on social media relate to the actual geography of protests. By exploring the spatial synchrony and asynchrony between protestors and Twitter users, we can empirically test hypotheses about the spatial location of protestors and users tweeting about the protests.



3. Objectives

In this paper we compare different sources of geographic information from Twitter related to the protests in Brazil with the location of protestors attending the demonstrations onsite. The source of geographic information for Twitter messages varies considerably in terms of reliability and precision, and the data gathered for this study includes: a) location identified by GPS coordinates; b) location identified by information on user profile; and, c) location identified by information in the text message. We relied on these sources of geographic information to compare the number and the distribution of protestors online with the number and the distribution of protestors attending demonstrations onsite. Data related to the number of protestors onsite was retrieved from press reports and aggregated by number of protestors per location (see [Annex I](#)) in order to be cross-comparable with the data retrieved from Twitter.

The primary objectives of this study are twofold. Firstly, we hypothesize that there is great

spatial heterogeneity between the locations where users tweet their messages and the location they addressed their communication. The underlying assumption being tested is that Twitter users direct their political communication to locations that are relatively remote from where they are physically placed. Secondly, we hypothesize that the distribution of hashtagged tweets is similar to that of street protestors, the underlying assumption being that the geography of hashtagged messages is similar to the political activity onsite. To this end, we tested the following hypotheses:

H1: The geographic distribution of political communication is concentrated in politically influential regions of the country.

H2: The geographic distribution of protestors attending demonstrations is closer to the distribution of hashtag messages than to profile and geocode messages.

H3: The hashtagged location referred to in the messages is relatively remote from the geographic location where users tweeted the message.

H4: The geographic distribution of users tweeting the protests is broader, less clustered, and relatively remote from the geographic distribution of street protestors.

Prior to testing the hypotheses, we collated the sources of geographic information online and onsite and normalized the data based on socio-economic indicators of Brazilian society. The georeferencing of the data was only possible due to unusual features of the information streams. Firstly, the Vinegar protests took place across most of Brazil, thus providing cross-country data about the same political event in a relatively short time frame. Secondly, Twitter messages were hashtagged following a city and/or state location-based method, so that messages related to protest in Rio de Janeiro and other federative units can be easily identified regardless of whether they include geocode information. Lastly, the combination of multiple streams of political activity provides an opportunity to understand how Twitter users engage in political movements from where they presently are; where they are coming from; and to what location they are addressing their communications.



4. Data

We consulted press reports about the location and the number of protestors in Brazil during the second half of 2013 and monitored 35 Twitter hashtags and keywords associated with the protests (see [Annex I](#)) via Twitter Search and Streaming APIs (O'Brien III, 2010). Data collection also relied on keywords to include tweets that otherwise would not have been monitored due to the lack of hashtags in the body of the text. We expect the combination of 35 hashtags and keywords associated with the protests in Brazil to have rendered a representative, if biased, sample of the full dataset (Morstatter, *et al.*, 2013), as the requested data is well below the one percent threshold of the entire public stream allowed by Twitter Streaming API. Although the data collection spans a period of six months, the dataset analyzed in this study covers 19 days of protesting activity, starting on 11 June and ending on 30 June 2013. This is the period when demonstrations filled the streets of Brazilian cities with over two million protestors.

The geographic location of protestors attending demonstrations retrieved from press reports was subsequently geocoded to match the database of Twitter messages. We retrieved the geographic information about the messages using the following three-step process: 1. Reverse geocoding the messages that included geocode information (two percent of the dataset); 2. Extracting the location of messages based on the self-reported geographic location retrieved from user profiles (31 percent of the dataset); and, 3. Identifying geographic locations based on explicit references made in the text of the message (nine percent of the dataset). Tweets were identified as coming from or referring to 3,268 Brazilian cities across the 27 federative units. [Figure 1](#) shows the geographic distribution of messages across the country and the number of protesting messages per federative unity in the period.

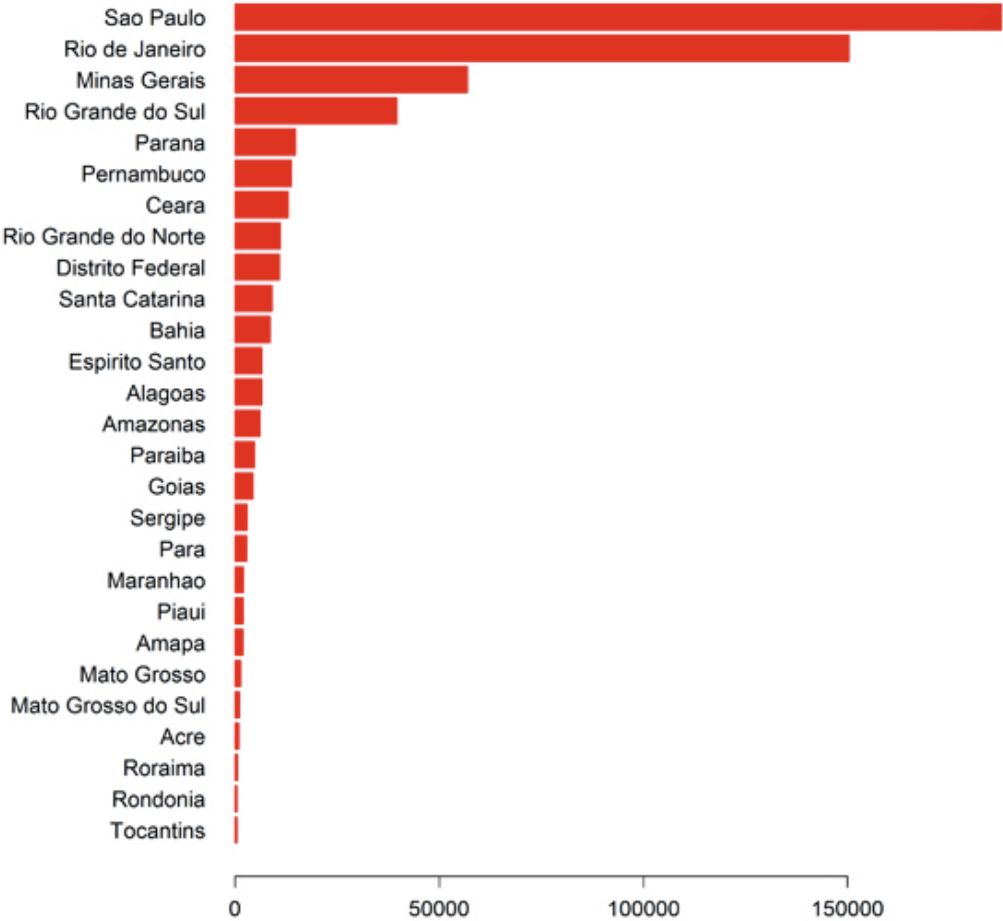


Figure 1: Distribution of messages across the country (top) and rank of states by number of messages posted by users in the period (bottom). A larger version of the top portion of this figure can be found at <http://www.uic.edu/~ejv/img/Figure1a.png>. A larger version of the lower portion of this figure can be found at <http://www.uic.edu/~ejv/img/Figure1b.png>.

The aggregated data presents the geographic coordinates of individuals participating or tweeting the protests in Brazil. Population density in Brazil varies considerable, ranging from three persons per square kilometer in the Amazon region to 30 persons in the Northeast and 150 in the state of São Paulo. The population-dependent data was normalized using Brazilian census (Censo, 2010) by calculating the rate of individuals engaged in political protests per Brazilian federative unit. We computed the proportion of individuals tweeting messages related to political demonstrations to the population of each state (in thousands). The normalized data shows which states presented higher percentages of protestors and/or Twitter messages across the country.

Figure 2 shows that the absolute number of tweets is concentrated in the richer, more densely populated states in the Southeast region, and the data is further explored in the spatial analysis reported in this paper. We also adjusted the data for the uneven geographical distribution of GDP, and Figure 2 shows that richer states are still overrepresented in terms of tweets by GDP per capita, particularly in São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), Rio Grande do Sul (RS), and Distrito Federal (DF). The bars in red show that poorer states in the Northeast region presented higher output of protesting tweets relative to the population, particularly Rio Grande do Norte (RN), Amapá (AP), and Alagoas (AL), with seven percent, five percent, and four percent respectively. Rio de Janeiro stands out with 19 percent of the protesting messages relative to the local population.

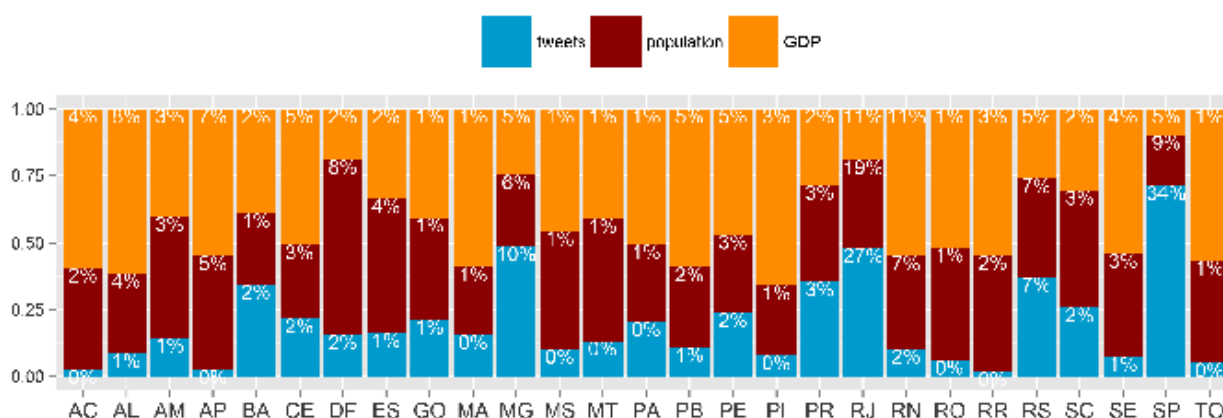


Figure 2: Absolute number of Twitter messages (blue) and volume of messages adjusted for the uneven geographical distribution of population (red) and GDP (yellow). Bars of the same color sum up to 100 percent. A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure2.png>.

4.1. Data cleansing

Tweets retrieved without the use of hashtags present a much higher signal-to-noise ratio than tweets archived using hashtags. We addressed this problem by geocoding and reverse-geocoding the messages and removing tweets associated with other instances of political protest (*i.e.*, protests in Turkey that overlapped with protests in Brazil). We geocoded tweets with geographic information based on hashtags and profiles by interpolating spatial locations from cities' main locations. Geocoded information retrieved from Twitter was reverse-geocoded to the city location in order to analyze descriptive statistics. We identified the geographic location of just under 50 percent the dataset (1.4M tweets) and removed messages not tweeted within the Brazilian territory.

4.2. Data sampling

The analyses reported in this paper were performed over a normalized dataset adjusted to the variations in population density and volume of tweets across Brazil. For the reasons detailed in the [Methods](#) section, we sampled the 1.4M messages in three randomized samples of 49,611 unique events based on protest, geocode, hashtag, and profile information, thus totaling 198,444 events retrieved from a stratified random sample with unequal sampling rates of protest, geocode, hashtag, and profile information streams. A number of functions in the spatstat library (Baddeley and Turner, 2005) require unique coordinates, so prior to sampling we added a random value ranging from 0.0001 to 0.001 to each geographic observation in order to avoid problems with duplicates.

The randomized samples were created by subsequently resampling the data to groups of 10K events to match the nationwide distribution of protests in Brazilian cities, thus producing four subsets of 10K events related to onsite (number of protestors) and online (geocode, hashtag, and profile) protesting activity. For specific analysis, we resampled the dataset to groups of 1K events of onsite and online political streams. Therefore, most of the analyses reported in this paper rely on marked planar point patterns with 40,000 and 4,000 points, respectively, with average intensities of 52.7 and 5.27 points per square unit. Coordinates are given to six decimal places and proportion is equally distributed between geocode, hashtag, profile, and protest information streams.



5. Methods

The analyses reported in this paper were performed using R (R Development Core Team, 2013) for statistical computing and the spatstat package (Baddeley and Turner, 2005) for space-time point pattern analysis. Spatstat supports point pattern data consisting of many different types in the same dataset, so we merged the four streams of political unrest into one marked planar point pattern contained by a window area of 758.974 square units (or 8,514,877 square kilometers using a conversion factor of $9.525e-3$ from geographic coordinates to metric system). Events are thus labeled according to the type they belong and the four categories are merged together into one point pattern. The advantage of this approach is that it allows for analyzing multitype point patterns such as the dataset used in this study.

We projected the x-y coordinates of locales of political unrest in a shape file and simulated a Poisson process conditional to the events and the deviation from CSR. The aim of this method is to indicate the number of protests in each region and project the spread of political upheaval for neighboring space points. To this end, we first removed all shared borders in the polygons (Brazilian map) to avoid problems with self-intersection and geometrical artifacts in the map. We also computed the distance between point patterns and calculated the optimal point matching between multiple streams of political unrest. The specialized primal-dual algorithm implementation in C by Illian, *et al.* (2008) can handle only patterns with a few hundreds of points, so we resampled the planar points to 1K points for each stream of political protest (geocode, hashtag, profile, and protest).

Analyses of the spatial locations of political protests were performed using point processes to provide a probability distribution of objects in a finite set of spatial locations. We defined the set of political protest locations as points in space in order to approach the qualitative emergence of protests as a spatial statistics problem (Barthelmé, *et al.*, 2012) and defined the window of observation as the Brazilian territory which comprises 26 states and one federal district. We relied on the Kulldorff and Nagarwalla (1995) model that supports data with exact geographic coordinates for each individual. The model also provides a method of detection and inference for spatial clusters and alternative hypotheses (Kulldorff, *et al.*, 1998).

The Cartesian coordinates of tweets and street protests were used to create point pattern data (Illian, *et al.*, 2008) that allowed for performing pair correlation function of point process using kernel methods and to determine the dependence between points in the spatial point process. We detected spatial and space-time clusters of political unrest and tested for random distribution over space (Kulldorff, 1997; 2001). We have not found political activity across all information streams in three states of the north and one state in the central-west region of Brazil, so the model was fed with information from 23 hotspot locations. We explored the intensity function $\lambda(s)$ to define where events are likely to happen in the area A

as the integral of the intensity function over A .

5.1. Limitations of the method

Limitations of the methods include the lack of a proper way to address cross-event correlation and cross-location correlation with space-time statistics. Moreover, complex visualization for pattern detection and hypothesis formulation is still forthcoming. There are also several limitations with the method due to limitations with the data. First, we managed to identify the location of only half of all users that tweeted messages related to the protests in Brazil. Second, the location of users was retrieved using sources that vary considerably in terms of reliability and precision. We expect considerable differences in the geographic information provided by geocoded tweets, user profiles, and Twitter messages, not only in terms of varied levels of accuracy, but also due to different locations from where users are talking; to whom users are talking; and, to which location users send their messages.

6. Results

We normalized the data to include an equal number of protestors and messages with location defined by geocode, profile, and hashtag and found major differences at the state level. [Figure 3](#) compares the distribution of protestors onsite (black), tweets (light blue), tweets with location informed by geo coordinates (blue), user profile (dark blue), and hashtags (red) across the states of Brazil. As we hypothesized in $H1$, the chart shows that the states in the wealthier southeast — particularly São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) — are the object of most hashtagged messages in sharp contrast with the distribution of messages tweeted at these locations. In the states of Rio de Janeiro and São Paulo, the ratio of messages tweeted about the protests in the state (34 percent and 17 percent, respectively) is about four times as high as the relative number of protestors at these locations (nine percent and four percent, respectively).

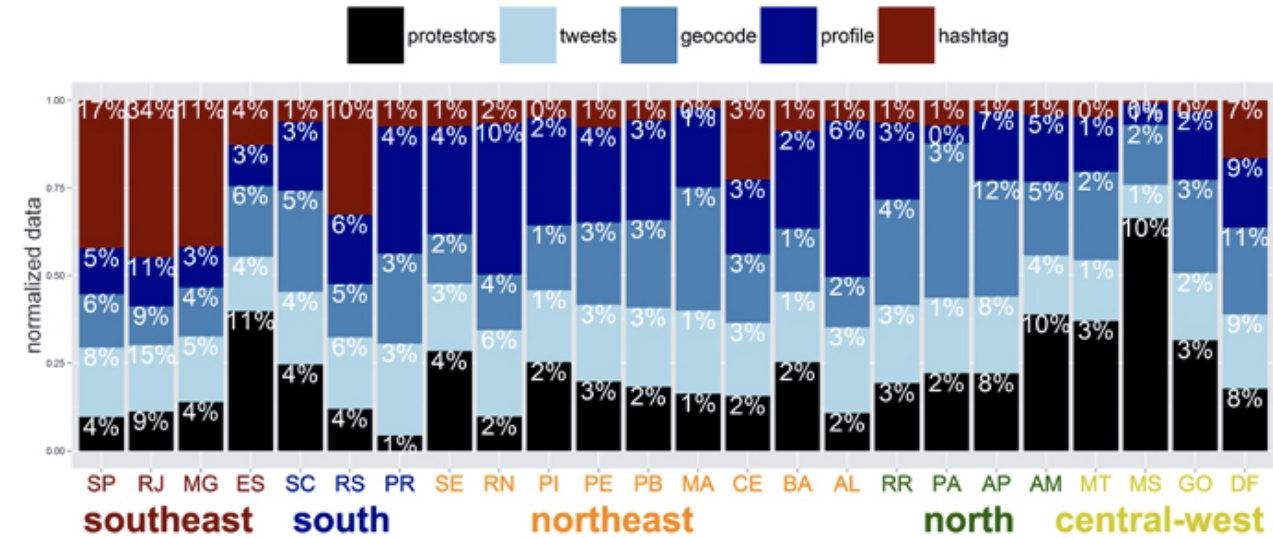


Figure 3: Volume of protestors attending demonstrations and tweets with geocoded, hashtagged, and self-reported location of users during the Brazilian Vinegar protests. Bars of the

same color add up to 100 percent. A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure3.png>.

A direct comparison between the adjusted distribution of protestors and tweets also shows that states in the wealthier southeast region, particularly Rio de Janeiro and São Paulo, present nearly twice as many tweets (15 percent and eight percent, respectively) as protestors (nine percent and four percent, respectively). On the other hand, less-connected, isolated regions of Brazil reported the inversed trend: a higher ratio of protestors attending demonstrations at these locations and a lower ratio of messages covering such protests — particularly the states of Amazonas (AM), Mato Grosso (MT), and Mato Grosso do Sul (MS). The state of Espírito Santo in the southeast region is the exception that confirms the rule. One extreme case is the state of Mato Grosso do Sul (MS), with 10 percent of protestors relative to local population and less than one percent of the ratio of messages tweeted in the period.

The comparison between the overall distribution of Twitter messages (light blue) and the distribution of hashtagged messages (red) sheds considerable light on the divisions of Brazilian society. The differences between the geographic locations from where users tweeted (geocode) and the geographic locations to where users directed their communications (hashtag) follows a socio-economic gradient between the wealthier states in the southeast region of Brazil, which concentrates large portions of the metropolitan public opinion of Brazil, and the peripheries of the country that direct their communication to this geopolitical center. Even though the states in the south and southeast regions of Brazil concentrate nearly 80 percent of all politically charged and hashtagged messages, the local population did not engaged more actively to the street protests in comparison to the remaining areas of Brazil.

In order to further explore *H1*, we plotted the aggregated data across the five regions of Brazil. In [Figure 4](#) we normalized both the source of location (geocode, profile, and hashtag) and the number of tweets according to the population in the regions of Brazil. The north region presents a considerable higher incidence of mobile platforms and geocoded tweets are on average twice as likely to come from the northern part of the country if population distribution was the same across the country. Summarizing the results reported in [Figure 3](#), the southeast region of Brazil takes up two-thirds of the entire debate about political protests in the country measured by the use of hashtags. Moreover, the ratio of messages with geographic information follows the GDP distribution of regions in Brazil: the southeast is followed by the south, which is followed by the northeast, central-west, and north regions. These results are broadly consistent with the hypothesis *H1*.

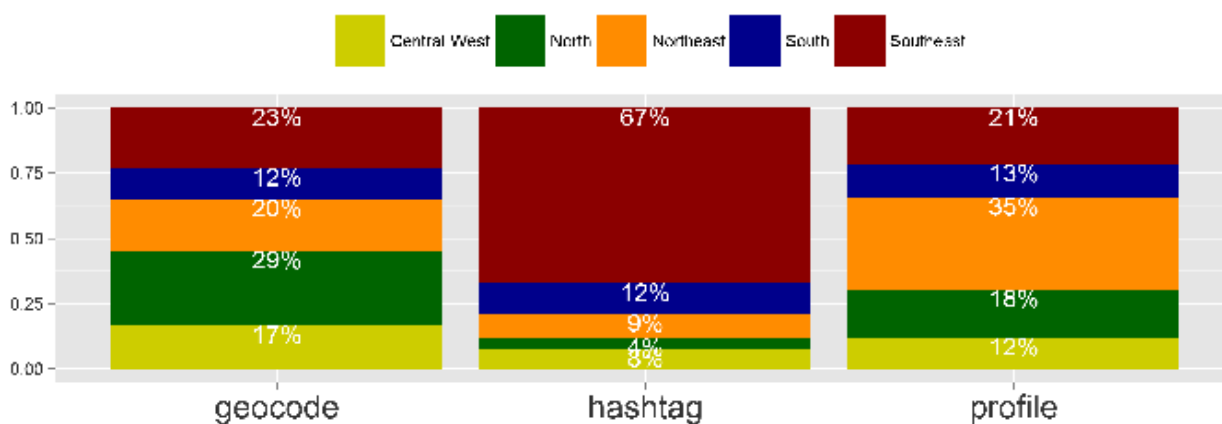


Figure 4: Location source of protesting messages aggregated by regions of Brazil. Population data is adjusted to the variations in population density across the country. A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure4.png>.

Figure 4 also shows remarkable differences across the locations where users reported they live (profile), locations to where users addressed their messages (hashtag), and locations where users posted the tweet (geocode). Hashtagged messages are associated with the formation of *ad hoc* publics (Bruns and Burgess, 2011), so that Twitter tags are organized as modern agoras engineered to provide visibility and press coverage to events. Because of that, users hashtag messages aimed at drawing attention to a particular cause or opinion regardless of whether they are physically present at that location. Similarly, the self-reported geographic location retrieved from user profiles is prone to returning locations users identify themselves with, rather than the place where users live.

The data shows that while the geographic information collected through hashtags is more focused on the southwest, where most protests occurred, the other types of geographic information are more evenly distributed among the remaining areas of Brazil. This suggests that hashtagged messages are more dedicated to the formation of *ad hoc* publics than addressing issues and problems at the local level. These differences indicate important political differences and economic inequalities across the country. First, although the north region of Brazil shows a higher incidence of geotagged tweets relative to population density, it presents a much lower volume of messages directed to that area of Brazil. Location reported in user profiles is seemingly equal across regions of Brazil, except for the northeast area, which is more likely to include this information, and the central-west, which is less likely to include this data.

In view of the concentration of messages in few regions of Brazil, we compared the central point of diffusion of protests to the central point of messages located via geocode, hashtag, and profile information. [Figure 5](#) shows the hotspot locations where messages related to political unrest were posted. Hotspots are defined by an intensity function $\lambda(s)$ in which s is the spatial location and the intensity function defines where events are likely to happen and the expected number of events to happen within the window of observation. The three sources of location retrieved from Twitter present equal centroids around the São Paulo–Rio de Janeiro axis, which form the economic center of the country. However, the intensity function varies greatly across location sources based on geocode, hashtag and profile information. The north region of Brazil presents points of diffusion mostly in the geocode projection plot, and the hashtag projection and the actual location of protests are particularly intense in the southeast region of Brazil.

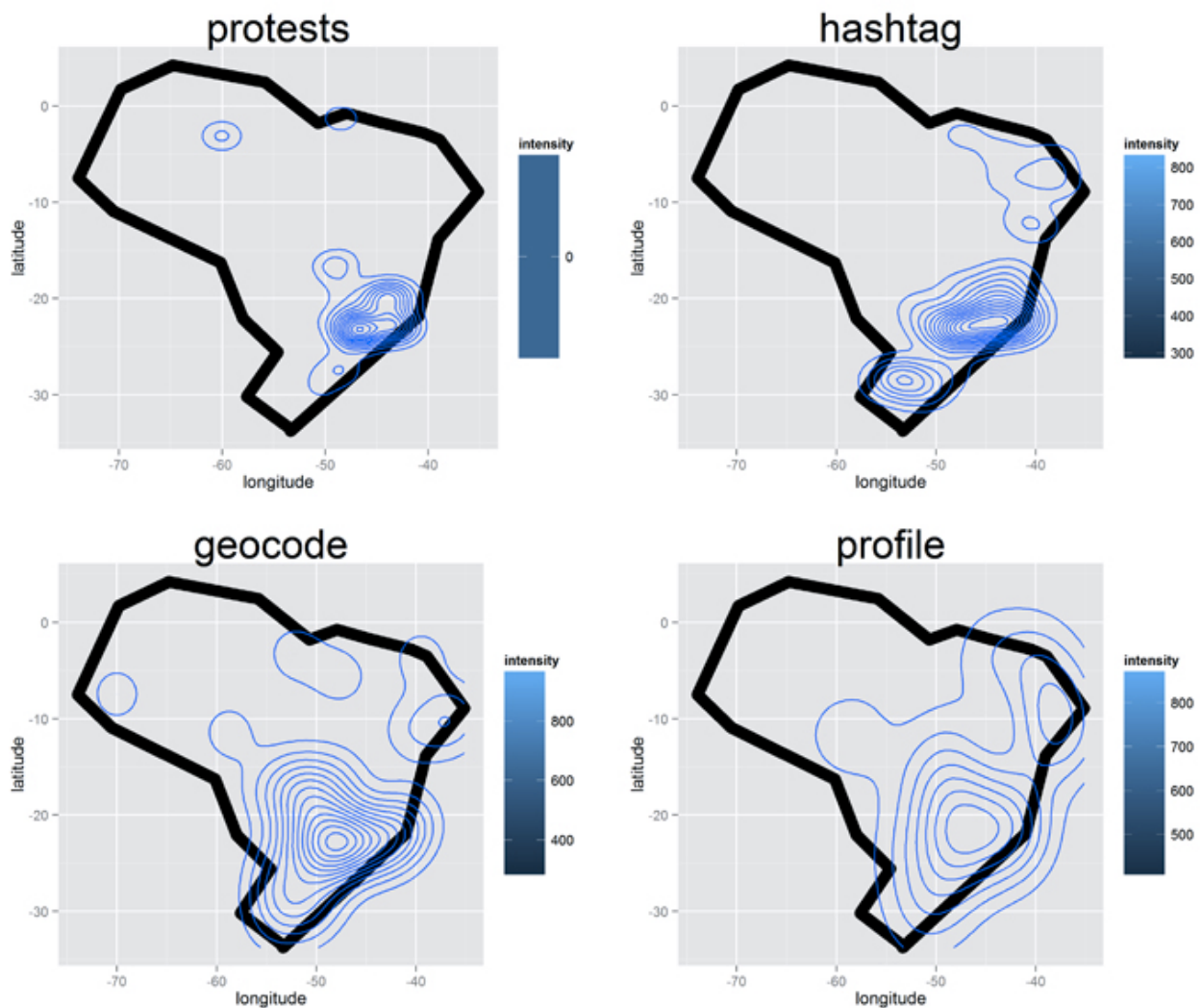


Figure 5: Central point of diffusion of political protests in Brazil and related Twitter messages based on geocode, hashtag, and user profile. A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure5.png>.

The intensity function does not take into account the population density across Brazil, so messages with location based on geocode information include hotspot location in the north part of Brazil, particularly around the cities of Belém and Rio Branco, but also in the northeast and the central-west regions of the country. Consistent with the information described in [Figure 4](#), messages with location based on data retrieved from user profiles present a more balanced distribution across the country, with a higher-than-average occurrence in the northeast part of the country. Finally, messages in which location was coded based on hashtag information are more broadly consistent with the actual location of protests, with density curves concentrated in the southeast of Brazil and smaller hotspots in the southern and northeast part of the country.

[Figure 5](#) also shows that the perimeter of the points of diffusion grows larger as we move away from onsite towards online protesting activity. Conforming hypothesis *H2*, the points of diffusion show that hashtag stream is geographically more similar to the restricted perimeter of the actual protests, while geocode projection much expands the covered zone and profile stream reaches the majority of the Brazilian territory. The results of this analysis are consistent with *H2* and indicate a two-way exchange from online to onsite political protests. On the one hand we observe that the point of diffusion in politically influential locations presents a much shorter geographic perimeter. These areas stem from the actual location of political protest and include the hashtag information stream (upper quadrants). The lower quadrants, on the other hand, show the actual location of users tweeting the protests. Consistent with hypothesis *H3*, this area covers a much larger portion of the country and is demographically more representative and largely different from the core areas where protesting activity thrived. These areas are particularly represented by the geocode and the

profile information streams.

Political activity often displays singular concentrations of intensity and an intensity function might not apply to these cases. Because of that, we estimated the intensity measure non-parametrically counting the numbers of points falling in each quadrat. [Figure 6](#) shows that the occurrences of onsite protesting activity are proportional to the occurrence of online protesting in geocoded and profiled messages, with symmetrical values across the six quadrats analyzed. The remarkable difference lays in hashtagged messages, which present a much higher occurrence of messages in the salmon-colored quadrat (southeast part of the country) and a much lower occurrence of messages in the light green-colored area (central-west part of the country) and the dark-blue quadrat (Amazon region). This confirms hypothesis *H1* again and shows that hashtagged messages are disproportionately driven towards highly populated, urbanized, and wealthier parts of the country.

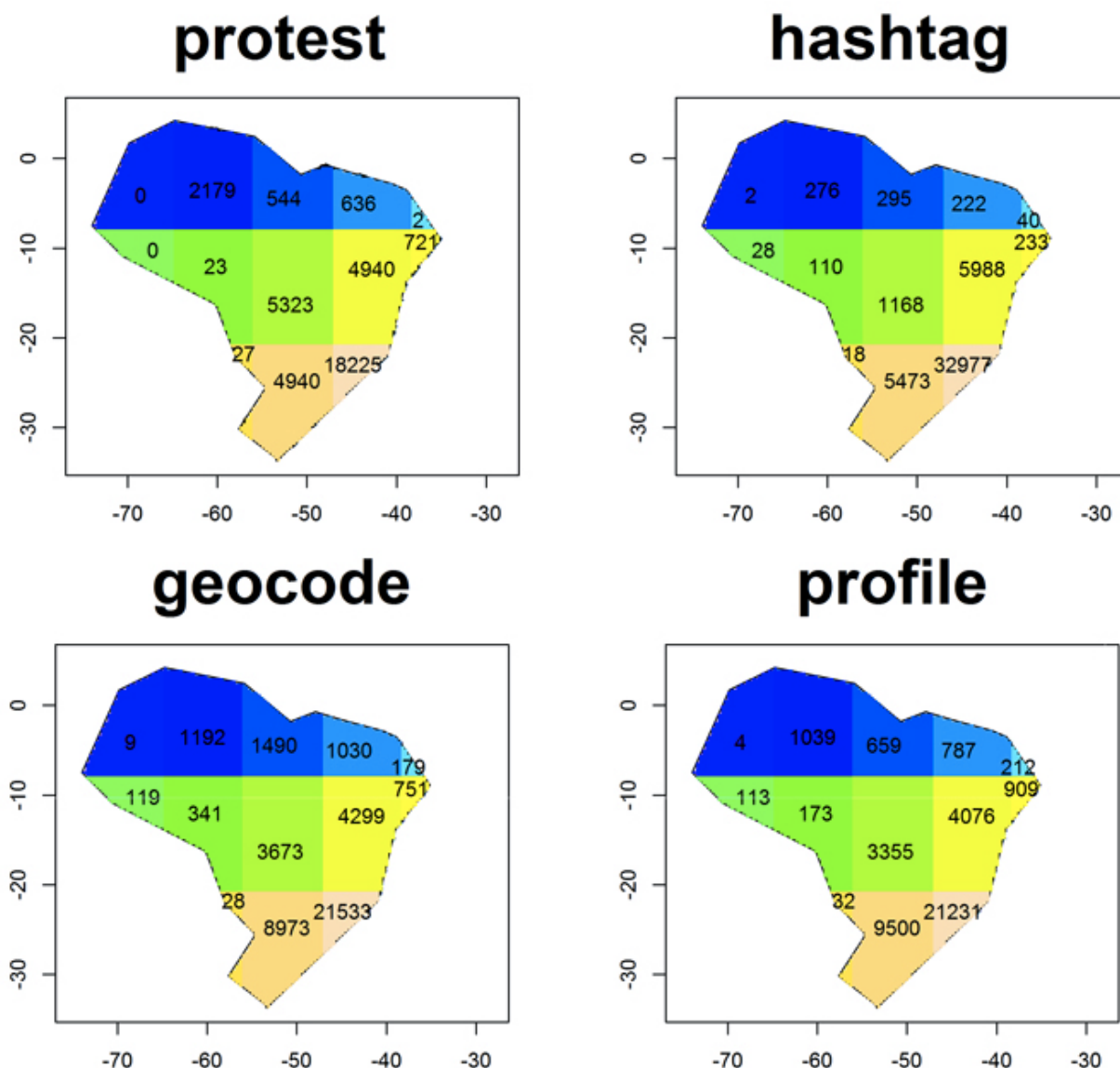


Figure 6: Quadrat counts for the activity streams of protestors (onsite), geocode, hashtag, and profile messages (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure6.png>.

Consistent with hypothesis *H4*, the plot of the chi-squared test using quadrat counts show that the geographic distribution of users that tweeted the protests does not necessarily overlap with the geographic distribution of street protests. The values reported in Figure 6 also confirm hypothesis *H3* by showing that hashtagged messages work as a conduit to bring together users that are sympathetic to the demonstrations but are not physically present to the demonstrations. Figure 6 also shows that the overall increase of protesting activity is consistent across all instances of online participation relative to onsite protesting. There are on average 5,000 more geocode and profile messages in the southeast part of the country, and nearly 1,500 less messages in the central-west part of Brazil relative to the distribution of individuals attending demonstrations in the country.

These differences are dwarfed by the sheer contrast between onsite and online political protest measured by the use of hashtags. Consistent with *H1*, the plot of the chi-squared test using quadrat counts show that there are 15,000 more hashtagged messages for every instance of political protest onsite in the southeast part of the country, and 4,000 less messages in the central-west region on Brazil relative to onsite political activity. The similarity is also noticeable when we look at the geometric center of protests across each information stream — *i.e.*, the mean position of all the points in the coordinate directions. [Figure 7](#) shows the centroids of each information stream — *i.e.*, the mean position of all the points in the coordinate directions. Both protest and hashtag streams are centered in the southeast, although the former is slightly drawn towards the south and the latter towards the northeast.

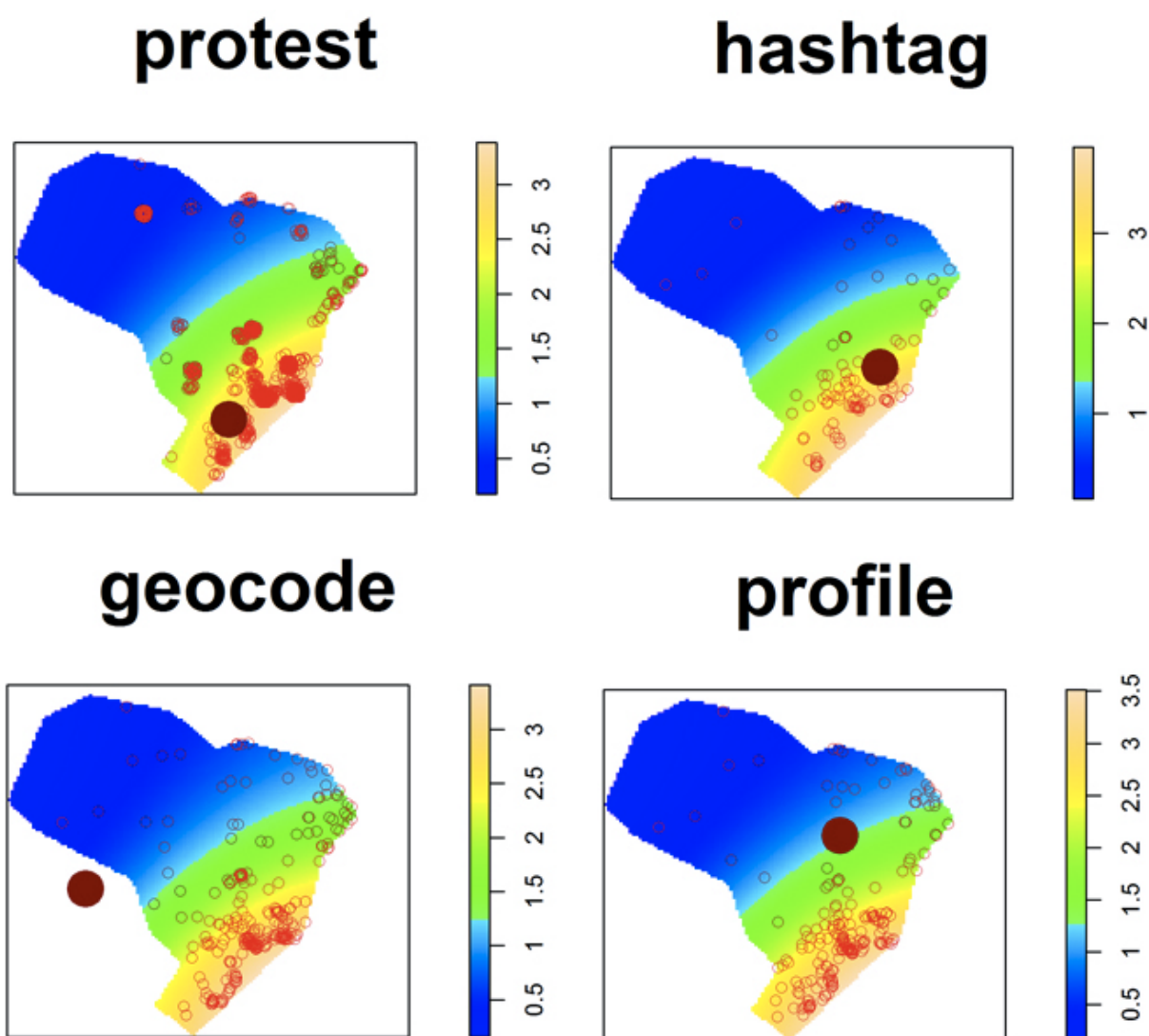


Figure 7: Centroid of the activity streams of protestors (onsite), geocode, hashtag, and profile

messages (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure7.png>.

In order to test whether hashtagged messages are geographically closer to the hotspots location of protests than profile and geocode messages, as asserted in hypothesis *H2*, we calculated the number of marks that are attached to close neighbors of onsite political protests in Brazil. We compiled a contingency table of the marks of geocode, hashtag, and profile points within a 0.1 radius of the geographic location where onsite protests took place. The results confirmed hypothesis *H2* and showed that hashtagged messages are geographically closer to the location of the protest, with a total of 48M (48,440,571) neighboring marks to the location of onsite protests, compared to 22M (22,784,824) and 28M (28,473,160) of geocode and profile, respectively.

In short, hashtagged messages presented nearly twice as many neighboring marks to the actual location of protests in comparison to geocoded and profile messages. Hashtagged messages are also the source of information that is closer to all sources of geographic information. We computed the average diameter of the five closest neighbors of each location registered by the four information streams. We first set the neighborhood of point *X* to consist of all points within a radius distance of 10 units (95,250 square meters), and the results returned 3.1M (3,124,744) neighbors for hashtags, followed by 2.8M (2,807,288) for protests, and 2.7M for geocode and profile (2,751,637 and 2,788,501, respectively). This is the number of neighboring marks for an area the size of 14 football pitches (95,250 square meters).

The difference between the information streams grows bigger as we set a more restrictive perimeter. With radius adjusted to 0.1 (95 square meters), hashtagged messages presented almost the same number of neighbors as all other sources of geographic information combined, with 500K (501,191) for hashtags, 230K for geocode (238,837), 280K for profile (288,421), and only 8K for protest (8,273). Again confirming hypothesis *H2*, the results show that hashtagged messages are not only geographically closer to the actual location of the protests, but also more connected to all sources of geographic information studied in this paper. In short, hashtagged tweets connect users from geographically remote regions to events gravitating toward urban centers and offer a place where users can track the developments on the ground.

Figure 8 shows the average distances between point patterns in onsite and online political activity streams. Although the shortest and the maximum distance are fairly equal across the point patterns considered (0.0019 and 36.29 units, or 0.20 and 3,810 kilometers, respectively), we found that the average distance from the point patterns of onsite activity to hashtag is much shorter at 7.31 units (767 kilometers) than the average distance from onsite activity to profile or geocode activity streams (8.69 and 8.86 units, or 912 and 930 kilometers, respectively).

The comparison between average distances shows that the geographic distribution of hashtags is on average 150 kilometers closer to the geographic location of street protests in comparison to the geographic distribution found in profile and geocode information streams. In fact, the difference between median distances of hashtag to protest and geocode or profile to protest is even more pronounced at 238 kilometers. Therefore, the average distances between sources of political protest show that hashtag activity is geographically closer to the geographic distribution of protestors across Brazil (hypothesis *H2*), although it also indicate that hashtag activity stream is not a good predictor to the actual location of users (hypothesis *H3*).

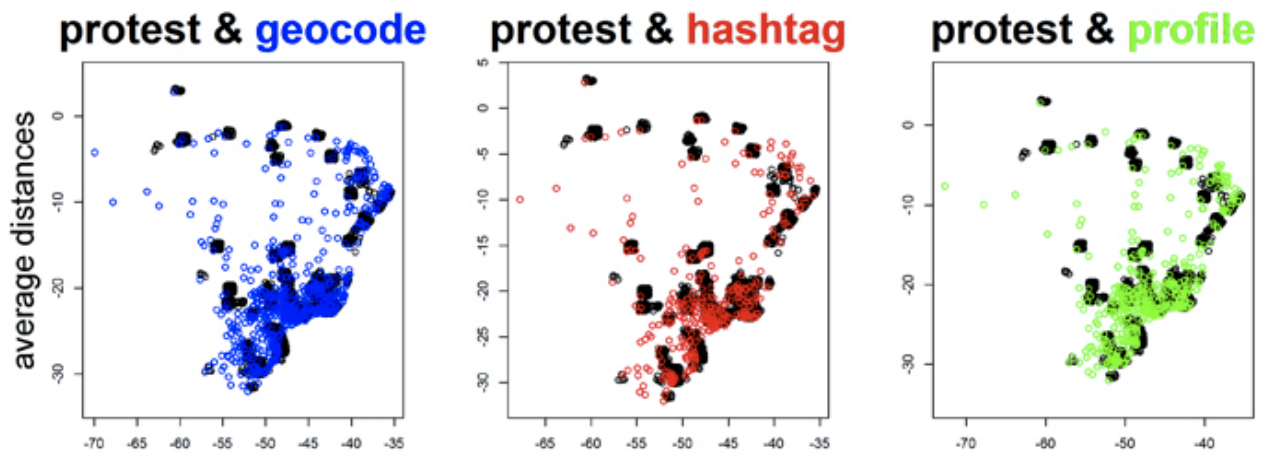


Figure 8: Average distances from point patterns of protestors (onsite) to tweets with geocode, hashtag, and profile (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure8.png>.

The distribution pattern of tweets with location information includes not only observations with very different neighbors (hotspots), but also observations that cluster together because they present very similar neighbors (Schabenberger and Gotway, 2005). We calculated the point matching between the geographic location of protests to the sources of tweets based on geocode, hashtag, and user profile. Figure 9 shows the optimal point matching between two point patterns with the larger cardinality n that is closest to the point pattern with the smaller cardinality m . Geocode data (blue) presented a larger-than-average volume of cardinalities in the northern part of the country, while hashtag data (red) is heavily concentrated in the southeast of Brazil.

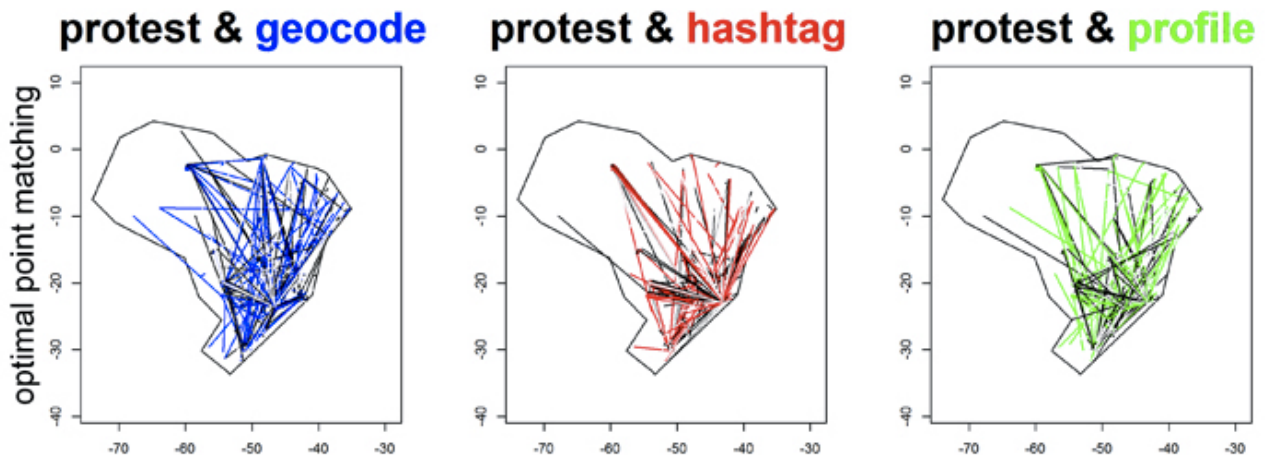


Figure 9: Point matching between the location of protestors (onsite) and the location of tweets based on geocode, hashtag, and profile (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure9.png>.

The matching between the planar point patterns of protest locations and social media location is revealing. The results are based on a bipartite weighted graph in which the vertices are provided by the two point patterns and edges are drawn each time a point of the first point pattern is matched with a point of the second point pattern (*i.e.*, a geographic point of protest matches the geographic point of a tweet). The randomized sample with 1K points matching returned 0.63 matching points between the planar point patterns of onsite political protests and geocoded messages; 0.66 between the planar point patterns of political protests and profile messages; and 0.75 points between the planar point patterns of political protests and hashtagged messages (cutoff = 1). These results confirmed hypothesis $H2$, and we

understand that the higher number of matching points between the geographic locations of hashtags and political protests results from both planar point patterns being heavily driven towards urban, wealthier, and politically influential areas.

In order to test hypothesis *H2* further, we converted the matrices with two-dimensional coordinates into a neighbors list to triangulate the grid points and draw a graph (Zuyev and White, 2013). Gabriel graphs draw a neighborhood only if there are no other points in their line set. [Figure 10](#) shows the Gabriel graphs between protesting activity and related tweets. Triangulations between locations of onsite protests and geocoded tweets are very similar to the triangulations between locations of onsite protests and profile tweets. In other words, the connections between geocode and protest location are fairly similar to the connections between protest and profile tweets. Both graphs presented identical percentage and number of nonzero weights at 0.08705 and 3,482 (and identical average number of links at 1.741). On the other hand, and again confirming hypothesis *H2*, triangulations between locations of onsite protests and hashtagged tweets are very dissimilar to geocode and profile messages, with percentage and number of nonzero weights at 0.0911 and 3,644 (identical average number of links at 1.822).

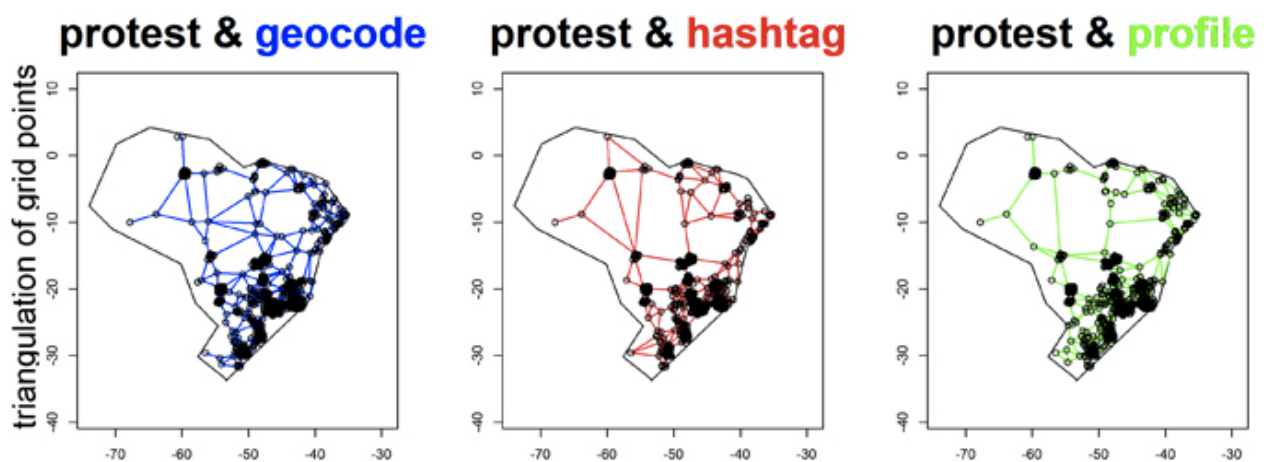


Figure 10: Gabriel graphs connecting the location of protestors (onsite) to the location of tweets based on geocode, hashtag, and profile (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure10.png>.

We also calculated the Gabriel graphs for each information stream and found substantial differences between them. Protest and hashtag graphs are again very similar, with 1.804 and 1.826 average number of links, respectively, and 243 and 233 regions with no links. Similarly, geocode and profile graphs are also symmetric, with 1.632 and 1.646 average number of links, respectively, and 246 and 249 regions with no links. We also found that protest and hashtag graphs presented a single clique as the most connected region with seven links, while the geocode graph presented two most connected regions with seven links and the profile graph included four most connected regions with six links. [Figure 11](#) shows the Delaunay triangulation based on the location of points in a tangent sphere. The graph maximizes the angles of the edges in the triangulation to avoid skinny formations.

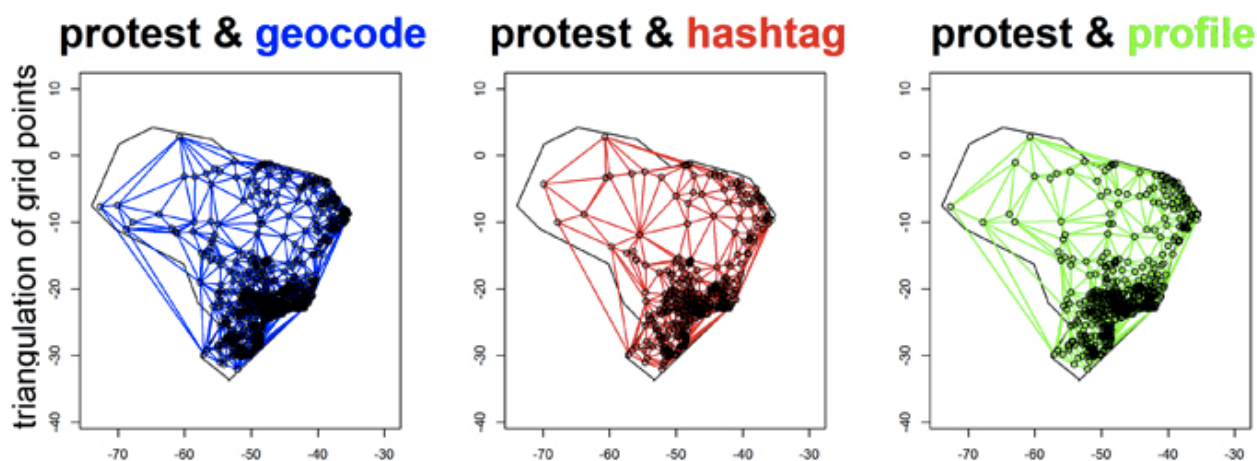


Figure 11: Delaunay graphs connecting the location of protestors (onsite) to the location of tweets based on geocode, hashtag, and profile (online). A larger version of this figure can be found at <http://www.uic.edu/~ejv/img/Figure11.png>.

7. Discussion and conclusion

In this paper we compared the geography of political protest in Brazil with the geographic location of tweets related to the protests. The data shows that Twitter users disproportionately focused their attention on the southeast region of Brazil and that this attention is not consistent with the distribution of population and/or protestors attending demonstrations. The conversational focus on this region is nonetheless consistent with the timeline of events, as protests started in the Brazilian city of São Paulo against increases in bus and metro fare. Protests in the southeast region were also more violent with 71 percent of the arrests and 88 percent of all injuries in the period, and included the majority of protestors (67 percent) engaged in demonstrations (see [Annex I](#)).

We found that users are often not in the actual location of the protests they are tweeting and that hashtagged messages work as a channel to bring together users that are sympathetic to the protests but are not attending the demonstrations. The locations indicated on hashtagged messages are both geographically closer to the actual location of the protests and more connected to all sources of geographic information investigated in this study. Hashtags thus connect regions geographically more isolated to urban centers at the same time they offer a platform that brings social media users closer to street protests.

In the remaining of this paper we summarize the hypotheses tested in this study and discuss the results.

H1: The geographic distribution of political communication is concentrated in politically influential regions of the country.

We confirmed *H1* and found that political communication is channeled from geographically remote areas to politically influential regions of the country. In other words, we found that wealthier and more prominent regions of Brazil are the object of a higher-than-average volume of hashtagged tweets not tweeted at these locations. In fact, the southeast region of Brazil takes up two-thirds of the entire debate about political protests in the country measured by the use of hashtags.

H2: The geographic distribution of protestors attending demonstrations is closer to the distribution of hashtag messages than to profile and geocode messages.

The results confirm *H2* and show that the geographic distribution of hashtagged messages is very similar to the geographic distribution of protestors onsite, with twice as many neighboring marks than geocode and profile streams. The central point of diffusion of

hashtagged messages is similar to that of street protests, both being particularly intense in the southeast region of Brazil. This is indicative that Twitter is used to report ongoing events and that hashtags follow closely the development of events onsite.

H3: The hashtagged location referred to in the messages is relatively remote from the geographic location where users tweeted the message.

We found that *H3* is consistent with our data as the location referred to on hashtagged tweets is relatively remote from the actual location where users tweeted their messages, with hashtagged messages being particularly poor at predicting the actual location of users. We found evidence that hashtagged messages are associated with the formation of *ad hoc* publics and that the geography of Twitter user base differs considerably from the geography of political communication. Our results also show that hashtagged messages are disproportionately driven towards highly populated, urbanized, and wealthier parts of the country whereas the user base is more equally distributed over the Brazilian territory.

H4: The geographic distribution of users tweeting the protests is broader, less clustered, and relatively remote from the geographic distribution of street protestors.


The results confirm *H4* and show that the location of users tweeting the protests is on average considerably distant at 768, 912, and 930 kilometers (7.31, 8.69, and 8.86 units) from the location of hashtag, profile or geocode activity streams, respectively, to the location of street protestors. The central point of diffusion of online activity presents a much larger perimeter in comparison to that of onsite activity. The geographic area from where users tweeted the protests covers a much larger portion of the territory and is demographically more representative of the national population compared to the area where protests occurred.

The comparison between the distribution of hashtagged and non-hashtagged tweets sheds considerable light on the divisions of Brazilian society. The difference suggests a socio-economic gradient between the wealthier states in the southeast region of Brazil, which concentrates large portions of the metropolitan public opinion of Brazil, and the peripheries of the country that direct their communication to this geopolitical center. The density curves of street protests and hashtags are concentrated in the economic center of the country, while messages based on geocode and profile are clustered in relatively remote areas. This confirms the role of hashtags in the formation of *ad hoc* publics (Bruns and Burgess, 2011), mostly framed at the national level against the backdrop of local politics. These results are also consistent with the hypothesis that social media activity is organized into neighborhoods with boundaries (Cranshaw, *et al.*, 2012) that differ from the local geography.

The main findings of this study offer a valuable contribution to the debate on media activism and can be broadly summarized in two findings. Firstly, the geography of street protests is considerably remote from the geography of users tweeting the protests (distance of 768, 912, 930 kilometers from the location of hashtag, profile or geocode activity streams, respectively). In fact, the analyses reported in this study provide empirical evidence that the geographies of online and onsite political activism are to a large extent dissimilar. These results support and extend the earlier findings of Leetaru, *et al.* (2013), who found that geographic proximity had minimal impact on what users communicate. The results also highlight that media places more emphasis on the nationwide political context than the actual locality where users tweeted their messages.

Secondly, and more critically, the results show that users from geographically remote areas engage in political communication as a means for airing one's political views despite the unequal geographic distribution of power. This is indicative that digital communication is instrumental in bypassing the constraints of broadcast media. Instead of having to deal with the high costs of production and distribution (as in print), or the use of scarce and expensive resources such as the electromagnetic spectrum (as in broadcasting), social media enable users to channel their concerns and aspirations to the political center of the country. We found that social media allowed for a larger and geographically more diverse collective of opinions and voices, but we have not identified any fundamental change or decentralization in the geography of power.

In fact, if anything, social media has amplified and consolidated the socio-economic and political divisions within the country. This is perhaps to be expected, as metropolitan areas and/or areas of high population density are likely to affect the political agenda and the public

opinion of less populated, geographically distant locations. As a matter of fact, the southeast region of Brazil is so influential that it includes more tweets than the remaining regions combined. Lastly, and to conclude, the unequal distribution of messages based on geocode, hashtag, and profile shows that identifying geographic location of social media users is a challenging task given the multiple locations users inhabit, occupy, and communicate at any given time. Despite these caveats, we expect the results reported in this study to inform future research focusing the relationship between onsite and online protesting activity. 

About the authors

Marco Toledo Bastos is the NSF EAGER HASTAC postdoctoral fellow at Duke University. A large portion of this research was completed while the author was a postdoctoral fellow at the University of São Paulo.

Direct comments to: marco [at] toledobastos [dot] com

Raquel da Cunha Recuero is an associate professor and researcher at the Department of Applied Linguistics and Social Communication in Universidade Católica de Pelotas (UCPel) in Brazil.

E-mail: raquel [at] raquelrecuero [dot] com

Gabriela da Silva Zago is a professor at the Department of Digital Design at Universidade Federal de Pelotas (UFPel) and a Ph.D. candidate at the Communication and Information Graduate School at Universidade Federal do Rio Grande do Sul (UFRGS) in Brazil.

E-mail: gabrielaz [at] gmail [dot] com

References

Lars Backstrom, Jon Kleinberg, Ravi Kumar, and Jasmine Novak, 2008. "Spatial variation in search engine queries," *WWW '08: Proceedings of the 17th International Conference on World Wide Web*, pp. 357–366.

doi: <http://dx.doi.org/10.1145/1367497.1367546>, accessed 25 February 2014.

Adrian Baddeley and Rolf Turner, 2005. "Spatstat: An R package for analyzing spatial point patterns," *Journal of Statistical Software*, volume 12, number 6, at

<http://www.jstatsoft.org/v12/i06/>, accessed 25 February 2014.

Simon Barthelmé, Hans Trukenbrod, Ralf Engbert, and Felix Wichmann, 2012. "Modelling fixation locations using spatial point processes," *arXiv*, at

<http://arxiv.org/pdf/1207.2370.pdf>, accessed 25 February 2014.

BBC Brasil, 2013. "Brasileiros 'descobrem' mobilização em redes sociais durante protestos" (11 July), at

http://www.bbc.co.uk/portuguese/noticias/2013/07/130628_protestos_redes_personagens_cc.shtml, accessed 25 February 2014.

Marija Anna Bekafigo and Allan McBride, 2013. "Who tweets about politics? Political participation of Twitter users During the 2011 gubernatorial elections," *Social Science Computer Review*, volume 31, number 5, pp. 625–643.

doi: <http://dx.doi.org/10.1177/0894439313490405>, accessed 25 February 2014.

W. Lance Bennett and Alexandra Segerberg, 2013. *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge: Cambridge University Press.

W. Lance Bennett, Alexandra Segerberg, and Shawn Walker, 2014. "Organization in the crowd: Peer production in large-scale networked protests," *Information, Communication & Society*, volume 17, number 2, pp. 232–260.

doi: <http://dx.doi.org/10.1080/1369118X.2013.870379>, accessed 25 February 2014.

W. Lance Bennett, Christian Breunig, and Terri Givens, 2008. "Communication and political mobilization: Digital media and the organization of anti-Iraq War demonstrations in the U.S.," *Political Communication*, volume 25, number 3, pp. 269–289.

doi: <http://dx.doi.org/10.1080/10584600802197434>, accessed 25 February 2014.

danah boyd, Scott Golder, and Gilad Lotan, 2010. "Tweet, tweet, retweet: Conversational

aspects of retweeting on Twitter," *43rd Hawaii International Conference on System Sciences (HICSS)*.

doi: <http://dx.doi.org/10.1109/HICSS.2010.412>, accessed 25 February 2014.

Axel Bruns and Yuxian Eugene Liang, 2012. "Tools and methods for capturing Twitter data during natural disasters," *First Monday*, volume 17, number 4, at

<http://firstmonday.org/article/view/3937/3193>, accessed 25 February 2014.

doi: <http://dx.doi.org/10.5210/fm.v17i4.3937>, accessed 25 February 2014.

Axel Bruns and Jean E. Burgess, 2011. "The use of Twitter hashtags in the formation of ad hoc publics," *6th European Consortium for Political Research General Conference*, at

<http://eprints.qut.edu.au/46515/>, accessed 25 February 2014.

Ian Cairns, 2013. "Get more Twitter geodata from Gnip with our new profile geo enrichment" (22 August), at <http://blog.gnip.com/twitter-geo-data-enrichment/>, accessed 25 February 2014.

Manuel Castells, 2012. *Networks of outrage and hope: Social movements in the Internet age*. Cambridge: Polity Press.

Manuel Castells, 2009. *Communication power*. Oxford: Oxford University Press.

Manuel Castells, 1997. *The power of identity*. Cambridge: Blackwell.

Censo, 2010. "Instituto Brasileiro de Geografia e Estatística," at

<http://censo2010.ibge.gov.br/en/>, accessed 25 February 2014.

Zhiyuan Cheng, James Caverlee, and Kyumin Lee, 2010. "You are where you tweet: A content-based approach to geo-locating Twitter users," *CIKM '10: Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 759–768.

doi: <http://dx.doi.org/10.1145/1871437.1871535>, accessed 25 February 2014.

Justin Cranshaw, Raz Schwartz, Jason Hong, and Norman Sadeh, 2012. "The Livehoods Project: Utilizing social media to understand the dynamics of a city," *6th International AAAI Conference on Weblogs and Social Media*, at

http://justincranshaw.com/papers/cranshaw_livehoods_icwsm12.pdf, accessed 25 February 2014.

Mario Diani, 2000. "Social movement networks virtual and real," *Information, Communication & Society*, volume 3, number 3, pp. 386–401.

doi: <http://dx.doi.org/10.1080/13691180051033333>, accessed 25 February 2014.

Daniela V. Dimitrova and Dianne Bystrom, 2013. "The effects of social media on political participation and candidate image evaluations in the 2012 Iowa caucuses," *American Behavioral Scientist*, volume 57, number 11, pp. 1,568–1,583.

doi: <http://dx.doi.org/10.1177/0002764213489011>, accessed 25 February 2014.

Jennifer Earl, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan, and Ashley Arlotti, 2013. "This protest will be tweeted: Twitter and protest policing during the Pittsburgh G20," *Information, Communication & Society*, volume 16, number 4, pp. 459–478.

doi: <http://dx.doi.org/10.1080/1369118X.2013.777756>, accessed 25 February 2014.

Qingqing Gan, Josh Attenberg, Alexander Markowetz, and Torsten Suel, 2008. "Analysis of geographic queries in a search engine log," *LOCWEB '08: Proceedings of the First International Workshop on Location and the Web*, pp. 49–56.

doi: <http://dx.doi.org/10.1145/1367798.1367806>, accessed 25 February 2014.

Huiji Gao, Jiliang Tang, and Huan Liu, 2012. "Exploring social-historical ties on location-based social networks," *Sixth International AAAI Conference on Weblogs and Social Media*, at

<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4574>, accessed 25 February 2014.

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant, 2008. "Detecting influenza epidemics using search engine query data," *Nature*, volume 457, number 7232 (19 February), pp. 1,012–1,014.

doi: <http://dx.doi.org/10.1038/nature07634>, accessed 25 February 2014.

Nils Gustafsson, 2012. "The subtle nature of Facebook politics: Swedish social network site users and political participation," *New Media & Society*, volume 14, number 7, pp. 1,111–

1,127.

doi: <http://dx.doi.org/10.1177/1461444812439551>, accessed 25 February 2014.

Jayson Harsin, 2013. "WTF was Kony 2012? Considerations for *Communication and Critical/Cultural Studies* (CCCS)," *Communication and Critical/Cultural Studies*, volume 10, numbers 2–3, pp. 265–272.

doi: <http://dx.doi.org/10.1080/14791420.2013.806149>, accessed 25 February 2014.

Joshua Hawthorne, J. Brian Houston, and Mitchell S. McKinney, 2013. "Live-tweeting a Presidential primary debate: Exploring new political conversations," *Social Science Computer Review*, volume 31, number 5, pp. 552–562.

doi: <http://dx.doi.org/10.1177/0894439313490643>, accessed 25 February 2014.

Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H. Chi, 2011. "Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles," *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 237–246.

doi: <http://dx.doi.org/10.1145/1978942.1978976>, accessed 25 February 2014.

Jeff Huang, Katherine M. Thornton, and Efthimis N. Efthimiadis, 2010. "Conversational tagging in Twitter," *HT '10: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pp. 173–178.

doi: <http://dx.doi.org/10.1145/1810617.1810647>, accessed 25 February 2014.

Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan, 2008. *Statistical analysis and modelling of spatial point patterns*. Chichester: Wiley.

Martin Kulldorff, 2001. "Prospective time periodic geographical disease surveillance using a scan statistic," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, volume 164, number 1, pp. 61–72.

doi: <http://dx.doi.org/10.1111/1467-985X.00186>, accessed 25 February 2014.

Martin Kulldorff, 1997. "A spatial scan statistic," *Communications in Statistics — Theory and Methods*, volume 26, number 6, pp. 1,481–1,496.

doi: <http://dx.doi.org/10.1080/03610929708831995>, accessed 25 February 2014.

Martin Kulldorff and Neville Nagarwalla, 1995. "Spatial disease clusters: Detection and inference," *Statistics in Medicine*, volume 14, number 8, pp. 799–810.

doi: <http://dx.doi.org/10.1002/sim.4780140809>, accessed 25 February 2014.

Martin Kulldorff, William F. Athas, Eric J. Feurer, Barry A. Miller, and Charles R. Key, 1998. "Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos, New Mexico," *American Journal of Public Health*, volume 88, number 9, pp. 1,377–1,380.

doi: <http://dx.doi.org/10.2105/AJPH.88.9.1377>, accessed 25 February 2014.

Juhi Kulshrestha, Farshad Kooti, Ashkan Nikraves, and Krishna P. Gummadi, 2012. "Geographic dissection of the Twitter network," *Sixth International AAAI Conference on Weblogs and Social Media*, at

<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4685>, accessed 25 February 2014.

Kalev Leetaru, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook, 2013. "Mapping the global Twitter heartbeat: The geography of Twitter," *First Monday*, volume 18, number 5, at <http://firstmonday.org/article/view/4366/3654>, accessed 25 February 2014.

doi: <http://dx.doi.org/10.5210/fm.v18i5.4366>, accessed 25 February 2014.

Merlyna Lim, 2013. "Framing Bouazizi: 'White lies', hybrid network, and collective/connective action in the 2010–11 Tunisian uprising," *Journalism*, volume 14, number 7, 921–941.

doi: <http://dx.doi.org/10.1177/1464884913478359>, accessed 25 February 2014.

Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley, 2013. "Is the sample good enough? Comparing data from Twitter's streaming api with Twitter's firehose," *Proceedings of the Seventh International Conference on Weblogs and Social Media (ICWSM)*; version at <http://arxiv.org/abs/1306.5204>, accessed 25 February 2014.

Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil, 2011. "An empirical study of geographic user activity patterns in Foursquare," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 570–573, and at <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2831>, accessed 25 February 2014.

- John O'Brien III, 2010. "yourTwapperKeeper," at <https://groups.google.com/forum/#!forum/yourtwapperkeeper>, accessed 25 February 2014.
- O Globo, 2013. "Protestos mobilizaram pelo menos 240 mil pessoas em 11 capitais" (17 June), at <http://oglobo.globo.com/pais/protestos-mobilizaram-pelo-menos-240-mil-pessoas-em-11-capitais-8716155>, accessed 25 February 2014.
- Zizi Papacharissi and Maria de Fatima Oliveira, 2012. "Affective news and networked publics: The rhythms of news storytelling on# Egypt," *Journal of Communication*, volume 62, number 2, pp. 266–282.
doi: <http://dx.doi.org/10.1111/j.1460-2466.2012.01630.x>, accessed 25 February 2014.
- Joel Penney and Caroline Dadas, 2014. "(Re)Tweeting in the service of protest: Digital composition and circulation in the Occupy Wall Street movement," *New Media & Society*, volume 16, number 1, pp. 74–90.
doi: <http://dx.doi.org/10.1177/1461444813479593>, accessed 25 February 2014.
- Daniele Quercia, Licia Capra, and Jon Crowcroft, 2012. "The social world of Twitter: Topics, geography, and emotions," *Sixth International AAAI Conference on Weblogs and Social Media*, at <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4612>, accessed 25 February 2014.
- R Development Core Team, 2013. "R: A language and environment for statistical computing," R Project for Statistical Computing, version 3.0.1, at <http://www.r-project.org/>, accessed 25 February 2014.
- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, 2010. "Earthquake shakes Twitter users: Real-time event detection by social sensors," *WWW '10: Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860.
doi: <http://dx.doi.org/10.1145/1772690.1772777>, accessed 25 February 2014.
- Matt Sandy, 2013. "Brazil kicks off: World Cup excess draws hundreds of thousands to street protests," *Independent* (18 June), at <http://www.independent.co.uk/news/world/americas/brazil-kicks-off-world-cup-excess-draws-hundreds-of-thousands-to-street-protests-8662863.html>, accessed 25 February 2014.
- Oliver Schabenberger and Carol A. Gotway, 2005. *Statistical methods for spatial data analysis* Boca Raton, Fla.: Chapman & Hall/CRC.
- David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill, 2009. "Tweet the debates: Understanding community annotation of uncollected sources," *WSM '09: Proceedings of the First SIGMM Workshop on Social Media*, pp. 3–10.
doi: <http://dx.doi.org/10.1145/1631144.1631148>, accessed 25 February 2014.
- Luke Sloan, Jeffrey Morgan, William Housley, Matthew Williams, Adam Edwards, Pete Burnap, and Omer Rana, 2013. "Knowing the tweeters: Deriving sociologically relevant demographics from Twitter," *Sociological Research Online*, volume 18, number 3, at <http://www.socresonline.org.uk/18/3/7.html>, accessed 25 February 2014.
doi: <http://dx.doi.org/10.5153/sro.3001>, accessed 25 February 2014.
- Yuri Takhteyev, Anatoliy Gruzd, and Barry Wellman, 2012. "Geography of Twitter networks," *Social Networks*, volume 34, number 1, pp. 73–81.
doi: <http://dx.doi.org/10.1016/j.socnet.2011.05.006>, accessed 25 February 2014.
- Sidney Tarrow, 2005. *The new transnational activism*. Cambridge: Cambridge University Press.
- Yannis Theocharis, 2013. "The wealth of (Occupation) networks? Communication patterns and information distribution in a Twitter protest network," *Journal of Information Technology & Politics*, volume 10, number 1, pp. 35–56.
doi: <http://dx.doi.org/10.1080/19331681.2012.701106>, accessed 25 February 2014.
- Narseo Vallina-Rodriguez, Salvatore Scellato, Hamed Haddadi, Carl Forsell, Jon Crowcroft, and Cecilia Mascolo, 2012. "Los Twindignados: The Rise of the Indignados Movement on Twitter," *SOCIALCOM-PASSAT '12: Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, pp. 496–501; version at <http://www.cl.cam.ac.uk/~nv240/papers/twindignados.pdf>, accessed 25 February 2014.

doi: <http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.120>, accessed 25 February 2014.

Yana Volkovich, Salvatore Scellato, David Laniado, Cecilia Mascolo, and Andreas Kaltenbrunner, 2012. "The length of bridge ties: Structural and geographic properties of online social interactions," *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, at <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4670/5002>, accessed 25 February 2014.

Sarita Yardi and danah boyd, 2010. "Tweeting from the town square: Measuring geographic local networks," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, at <http://research.microsoft.com/apps/pubs/default.aspx?id=122433>, accessed 25 February 2014.

Sergei Zuyev and Denis White, 2013. "tripack: Triangulation of irregularly spaced data," R package, 1.3–6 version, at <http://cran.r-project.org/web/packages/tripack/index.html>, accessed 25 February 2014.

ANNEX I

Table 1 List of Twitter Hashtags and Keywords Associated with the Vinegar Protests

	Vinegar	Tweets	Start Date
1	acordabrasil	68581	18 Jun 2013
2	brazilianspring	358	18 Jun 2013
3	catraca_livre	1512	13 Jun 2013
4	catracalivre	1515	13 Jun 2013
5	changebrazil	285385	16 Jun 2013
6	chupadilma	28200	16 Jun 2013
7	contraoamento	7167	13 Jun 2013
8	mudabrasil	103614	18 Jun 2013
9	obrasilacordou	12411	18 Jun 2013
10	ogiganteacordou	181511	18 Jun 2013
11	passe_livre	41810	13 Jun 2013
12	passelivre	37942	13 Jun 2013
13	primaverabrasileira	4460	18 Jun 2013
14	protesto	2210304	13 Jun 2013
15	protestobh	41651	17 Jun 2013
16	protestobr	30683	17 Jun 2013
17	protestoce	10698	17 Jun 2013
18	protestoemvitoria	1899	18 Jun 2013
19	protestopelotas	1572	17 Jun 2013
20	protestopoa	18044	17 Jun 2013
21	protestorj	114557	16 Jun 2013
22	protestosp	133260	13 Jun 2013
23	revogaoamento	2243	17 Jun 2013
24	saladuprising	839	18 Jun 2013
25	sp13j	6366	13 Jun 2013
26	sp17j	953	18 Jun 2013
27	tarifa	455756	13 Jun 2013
28	tarifa_zero	3767	13 Jun 2013
29	tarifazero	2166	13 Jun 2013
30	todarevolucaoomeca	2219	16 Jun 2013
31	todosjuntosporumbr	3578	16 Jun 2013
32	vdevinagre	1776	18 Jun 2013
33	vemprarua	534936	16 Jun 2013
34	verasqueumfilhoteu	603876	16 Jun 2013
35	vforvinegar	179	18 Jun 2013
	Number of Tweets:	4,955,788	

Table 2 Daily-Aggregated Data of Vinegar Protests with Number of Tweets, Protestors, Injuries, and Arrests

Date	Twitter	Protestors	Injuries	Arrests
13 Jun 2013	17728	17200	100	325
14 Jun 2013	48818	NA	NA	NA
15 Jun 2013	25605	NA	NA	NA
16 Jun 2013	39857	NA	NA	NA
17 Jun 2013	265774	275000	4	5
18 Jun 2013	294257	130500	NA	63
19 Jun 2013	180496	152000	NA	3
20 Jun 2013	33365	1569150	78	52
21 Jun 2013	33728	330450	NA	27
22 Jun 2013	20696	52500	NA	5
23 Jun 2013	14594	4000	NA	NA
24 Jun 2013	13036	8650	2	NA
25 Jun 2013	12166	11000	NA	35
26 Jun 2013	17069	88500	NA	1
27 Jun 2013	13342	4150	4	8
28 Jun 2013	10626	9500	NA	NA
29 Jun 2013	8525	NA	NA	NA

Received 12 February 2014; revised 20 February 2014; revised 22 February 2014; accepted 23 February 2014.



This paper is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/).

Taking tweets to the streets: A spatial analysis of the Vinegar Protests in Brazil
by Marco Bastos, Raquel Recuero and Gabriela Zago.

First Monday, Volume 19, Number 3 - 3 March 2014

<http://firstmonday.org/ojs/index.php/fm/rt/prINTERfriendly/5227/3843>

doi: <http://dx.doi.org/10.5210/fm.v19i3>.