The Geospatial Characteristics of a Social Movement Communication Network

Michael D. Conover*, Clayton Davis, Emilio Ferrara, Karissa McKelvey, Filippo Menczer, Alessandro Flammini

Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, Indiana, United States of America

Abstract

Social movements rely in large measure on networked communication technologies to organize and disseminate information relating to the movements’ objectives. In this work we seek to understand how the goals and needs of a protest movement are reflected in the geographic patterns of its communication network, and how these patterns differ from those of stable political communication. To this end, we examine an online communication network reconstructed from over 600,000 tweets from a thirty-six week period covering the birth and maturation of the American anticapitalist movement, Occupy Wall Street. We find that, compared to a network of stable domestic political communication, the Occupy Wall Street network exhibits higher levels of locality and a hub and spoke structure, in which the majority of non-local attention is allocated to high-profile locations such as New York, California, and Washington D.C. Moreover, we observe that information flows across state boundaries are more likely to contain framing language and references to the media, while communication among individuals in the same state is more likely to reference protest action and specific places and times. Tying these results to social movement theory, we propose that these features reflect the movement’s efforts to mobilize resources at the local level and to develop narrative frames that reinforce collective purpose at the national level.

Introduction

One of the most prominent American political movements of the past thirty years, Occupy Wall Street (‘Occupy’) is remarkable in the extent to which social media played a central role in its development and organization [1,2]. In this study, we examine how the needs and constraints of social movements are reflected in the geospatial characteristics and information sharing practices of Twitter users engaged in communication about the Occupy movement. Specifically, we focus on the geographic distribution of these users and the ways in which the relationships among them diverge from those of users contributing to the two most popular streams for stable political discourse in the United States, ‘Top Conservatives on Twitter’ and ‘Progressives 2.0.’

The organizing forces underlying successful social movements have been studied extensively by sociologists and political scientists [3–7]. From this body of work common themes have emerged, include the problems of resource mobilization and collective framing, which together constitute two of the core issues any social movement must address in order to effect social or political change. Resource mobilization refers to the process by which a social movement must marshal the financial, material, and human resources required to sustain its activities [8]. Collective framing is a process whereby the constituents of a social movement, through formal or informal processes, come to establish the narratives, language, and imagery that capture the essential features of the movement’s purpose and struggle [9]. Effective framing helps to foster a sense of community and engagement, and can be a powerful response to countervailing social pressures from establishment organizations [10].

Here we study Occupy Wall Street, a social movement focused on issues relating to the uneven distribution of wealth, social inequality, corporate greed, and the regulation of major financial institutions. Since the first protest on September 17th, 2011, a major feature of the movement has been the long-term physical occupation of high-visibility encampments, often found in parks, banks, libraries and foreclosed homes. As a result, the Occupy movement requires substantial supporting infrastructure, including housing and sanitation facilities, as well as access to communication technologies. In spite of this, Occupy has sustained a lasting presence in American cities including New York City, Oakland, Washington, D.C., and Boston, which also represent key loci of decision making and protest activity [1,2]. Under the Occupy model, proposals are brought to a vote before a general assembly, a form of direct democracy in which any participant is free to comment or vote on any proposal under consideration. The most prominent among these organizational structures is the New York City General Assembly, which has been responsible for producing policy and key narrative frames such as the popular protest slogan, “We are the 99%,” which references the disproportionate

Citation: Conover MD, Davis C, Ferrara E, McKelvey K, Menczer F, et al. (2013) The Geospatial Characteristics of a Social Movement Communication Network. PLoS ONE 8(3): e55957. doi:10.1371/journal.pone.0055957

Editor: Yamir Moreno, University of Zaragoza, Spain

Received October 18, 2012; Accepted January 4, 2013; Published March 6, 2013

Copyright: © 2013 Conover et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors gratefully acknowledge support from the National Science Foundation [http://www.nsf.gov/] (grant CCF-1101743), DARPA [http://www.darpa.mil/] (grant W911NF-12-1-0037), and the McDonnell Foundation [http://www.jsmf.org/]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: midconov@indiana.edu
Materials and Methods

Twitter Platform

Twitter is a popular social networking and microblogging site extensively explored in recent literature [15–21]. Among others, it has been used to study influence and credibility [22–26], social structure [27–29] and to monitor users’ sentiment [30–33].

Twitter users can post 140-character messages containing text and hyperlinks, called tweets, and interact with one another in a variety of ways. Communication on Twitter is characterized by directed, non-reciprocal social links that allow users to subscribe to the stream of content produced by another user. The content produced by every user an individual follows is aggregated into a single streaming feed, from which an individual can selectively rebroadcast content to his or her followers by choosing to retweet it. In this way, a retweet serves to broaden the potential audience of a piece of content, and signifies that information has been transmitted between two individuals. Hashtags, short tokens prepended with a pound sign (e.g. #taxes or #obama), constitute another important feature of the platform, and allow the content produced by many individuals to be aggregated into a custom, topic-specific stream including all tweets containing a given token.

Data

The analysis described in this article relies on data collected from the Twitter ‘gardenhose’ streaming API between July 3rd, 2011 and March 12th, 2012—a nine month period including the birth and maturation of the Occupy Wall Street movement. The gardenhose provides an approximately 10% sample of the entire Twitter stream in a machine-readable format. Gardenhose tweets include useful metadata, among them a unique tweet identifier, the timestamp, the username of the account that produced the tweet, a free text ‘location’ string associated with the originating user’s profile, and for retweets, the account names of the other users associated with the tweet. Tweets from geolocation-enabled mobile devices also report latitude/longitude coordinates, however the incidence rate of tweets with this data is not enough to be useful as a feature in general.

To isolate a representative sample of Occupy Wall Street content we flagged for collection any tweet containing hashtags associated with the Occupy movement, including #ows and #occupy{*} (e.g. #occupywalls, #occupyboston, etc.). To provide a baseline against which to compare our observations, we also extracted content originating from the two most popular communication channels associated with stable domestic political communication, #tcot (Top Conservatives on Twitter) and #p2 (Progressives 2.0). In total, this sampling procedure produced 1,522,415 tweets associated with Occupy Wall Street and 825,262 tweets associated with domestic political communication. As this analysis is concerned primarily with information spreading processes we consider only retweet events from this corpus, resulting in 676,369 retweets among 257,657 users associated with Occupy Wall Street, and 259,703 retweets among 68,049 users associated with stable domestic political communication. Henceforth, we consider these corpora to constitute representative samples of retweet interactions among users participating in the streams of content associated with the Occupy Wall Street movement and stable domestic political communication in the United States.

Geocoding

To facilitate a geospatial analysis of communication activity associated with these content streams we require a high quality method to infer individual users’ locations. To accomplish this, we rely on self-reported location strings and the services of a commercial geocoding API. This technique, popularized in work by Omela et al. [34], has been shown to produce high-resolution, high-quality geolocation data in the presence of geographically meaningful input.

A caveat to this technique, however, is that it relies on raw text generated by a broad swath of the Twitter population, and so we find geographically meaningless location descriptors included in the dataset. To address this issue we rely on an extensive hand-curated blacklist of popular non-geographical responses such as ‘everywhere’ and ‘the dance floor’. To produce this list we sorted topic-specific stream including all tweets containing a given token.

The Geography of Social Movement Communication
tailed distribution, 53% of all tweets in the data set are associated with a location among the 1,000 most popular responses, with 27% of all tweets containing one of the top hundred location strings. From this set of one thousand we blacklisted 161 non-location strings, corresponding to 6% of the tweets associated with the 1,000 most popular responses.

To improve recall in the presence of novel input, we used a modified version of the Ratcliff-Obershelf algorithm [35] to detect fuzzy matches between free text location strings and the blacklist of popular non-location responses. As a result, because ‘the dance floor’ is in the set of blacklist responses, strings taking a slightly modified form, such as ‘on the dance floor,’ will also be classified as invalid input. The hand-coded blacklist combined with the Ratcliff-Obershelf fuzzy matching technique resulted in 9% of the free-text location strings being classified as non-location input.

From among the remaining responses we submitted location strings to the Bing.com geocoding API, which returns a best-guess estimate for the corresponding physical coordinates. This output is hierarchically formatted to describe the finest level of geographic resolution available. For example, if a user reports ‘Logan Square, Chicago’ as his or her location, the Bing API will return information about the likely zip code, city, state and country associated with that location. However, if the user reports only ‘USA,’ the information provided by the API describes only a country-level guess as to the user’s location. Owing to decreased coverage at the city-level and the proportionately few users associated with each individual city, we utilize the state-level location estimates for the geospatial components of this analysis.

In total, 68.4% of Occupy Wall Street users reported location strings, and from these we were able to obtain geolocation estimates for 55.7% of these accounts. Among this set of users, 60% of the resulting geolocation estimates included state-level metadata. Response rates were somewhat diminished for users associated with the stream of domestic political communication, with 36% of individuals reporting free-text location strings. Using the procedure described above, we were able to obtain geolocation estimates for 29.3% of all users in the domestic political communication stream, 82.4% of which contained state-level metadata.

Geographic Profile

One of our goals is to establish a coarse-grained geographic profile for communication activity associated with the Occupy Wall Street movement. Formally, for each stream we define an activity distribution across states as, $$A_i = \frac{T_i}{t}$$ where $$T_i$$ is the total number of retweets originating from state $$i$$ and $$|T|$$ is the total number of retweets originating from all states. As outlined above, we focus on retweets as they correspond to attention allocation rather than total content production volume.

In addition to the distribution of activity across individual states we examine the information sharing relationships among users in different locations. To accomplish this, we rely on a network representation to characterize the flow of information on Twitter. Taking users as nodes, we define a weighted directed network in which an edge with weight $$w$$ is drawn from node $$U_1$$ to $$U_2$$ in the event that user $$U_2$$ retweets user $$U_1$$ $$w$$ times. The intuition underlying this approach is that each retweet provides evidence suggesting that information produced by user $$A$$ was evaluated and acted upon by user $$B$$.

Combining the user-level geocode metadata previously described with the network representation defined here we can induce another network describing the volume of communication between users in each state. In this network, nodes represent states, and weighted directed edges are drawn among them. The weight of the edge from $$S_1$$ to $$S_2$$ is defined as the sum of the weights among all edges originating from users in state $$S_1$$ and terminating in state $$S_2$$. We note, however, that this induced network must have geolocation labels for each node in a dyad. In the Occupy Wall Street stream we identify 143,437 tweets for which both the source and target have state-level geolocation data and 78,467 likewise restricted tweets in the stream of stable domestic political communication.

Textual Content

Finally, we wish to investigate whether the content of tweets with different geospatial properties serve distinct communication functions. To accomplish this, we segregate Occupy Wall Street retweets into two classes: interstate retweets connect pairs of users in different states, and intrastate retweets connect users in the same state. We compute the probability of observing a token, $$t$$, in a tweet from a given class, $$x$$, as $$P(t|x)$$. Comparing these probabilities yields a ratio, $$\frac{P(t|x)}{P(t|x_{\text{nominate}})}$$, a value which is large when a token is more common in intrastate traffic than interstate traffic and small under the opposite conditions.

Results

Geographic Concentration

Figure 1, in which states are ordered according to the proportion of stream activity, shows that content in the Occupy stream is substantially more geographically concentrated in a few key states compared to domestic political communication. For example, New York accounts for 30% of the total retweet activity in the Occupy stream, while the most popular source for stable domestic political communication, Washington D.C., accounts for only 10.7% of the stream’s total volume. As these plots make clear, the primary locations for on-the-ground Occupy activity are those places responsible for the majority of widely rebroadcast Occupy content, with California, New York and Washington D.C. acting as the source of 53.8% of total retweets. Figure 2 maps the states where the proportion of activity associated with the Occupy stream deviates the most from that associated with the stream of domestic political communication.

We also study the ratio of content production to content consumption by locale. Figure 3 shows this ratio, defined as the total number of retweets originating from users in that state divided by the total number of tweets retweeted by users in that state. This value serves to highlight the extent to which users in a given location are functioning as content producers or content consumers. Inspecting this plot, we find that in the Occupy stream users from just five states produced more content than they consumed. This stands in contrast to the stream of stable domestic political communication, in which fourteen states exhibit a ratio greater than one.

To highlight the effect of this geospatial concentration on communication flows between states it is instructive to visualize the structure of these networks. However, owing to the geographic aggregation process outlined in section Geographic Profile both networks are highly dense, with edges spanning most pairs of states. To address this issue we utilize a technique known as multiscale backbone extraction [36], which is useful for identifying statistically significant edges in weighted networks, regardless of the absolute value associated with the weight of that edge. This technique selects for edges with weights significantly above the expectation given by an analytically defined probability distribution that models a random allocation of each node’s strength among its adjacent edges. Parameterized by a confidence level factor, $$z$$, this technique allows for the selection of statistically
significant edges across all weight scales, a feature that is especially valuable when working with networks with heterogeneous weight distributions such as those associated with communication or human mobility.

Applying this technique to both networks reveals a communication backbone for the Occupy network that exhibits the highly concentrated hub and spoke structure described above. Figure 4 shows that the Occupy Wall Street network is characterized by minimal state-to-state connectivity, with the majority of statistically significant traffic flowing to and from New York, California and Washington D.C. This is in contrast to the communication backbone for the network of domestic political communication, in which we observe extensive interactions among many pairs of states.

Localization

In Figure 5 we present interstate connectivity for each communication network as a matrix in which the weight of an edge is mapped to a grayscale hue ranging from white for weak relationships to black for the strongest relationships. Inspecting these plots, one of the most striking ways in which the topology of the Occupy Wall Street communication network departs from that of the domestic political communication network is the high degree of localization. This is evidenced by the presence of a strong diagonal in the Occupy Wall Street connectivity matrix, as well as the significant off-diagonal mass in the domestic political communication matrix. We find that 40% of Occupy retweets originate and terminate with users in the same state. In contrast, 11% of retweets from the domestic political stream exhibit this type of locality, an increase of more than 350%.

Textual Analysis

To study the relationship between geography, resource mobilization, and collective framing, we focus on the content of tweets flowing within and between states. Restricting our analysis to tokens that account for at least 0.1% of both the intrastate and interstate tweet text, Table 1 presents the ten tokens most overrepresented in both intrastate communication as well as interstate communication.

Discussion

The analysis of interstate connectivity patterns reveals that, relative to stable domestic political communication, the Occupy network has a highly localized geospatial structure, with a disproportionately large amount of traffic being produced and consumed by users in the same state. We propose that this
phenomenon may be related to the issue of resource mobilization, that is, the process whereby any social movement must marshal resources such as money, infrastructure and human capital to further the goals of the movement. In the case of Occupy Wall Street, such resources are often quite tangible, and include not only tents and food, but also the participants required to facilitate large-scale protest action and extended encampments in cities across the country. In this light, it is easy to understand why such a disproportionately large fraction of attention is allocated to communication at the local level.

With respect to the finding that the majority of widely rebroadcast content is produced by users in a small number of

Figure 2. Divergences in geographic distribution of users. This cartogram uses color to represent the extent to which the number of Occupy Wall Street tweets in each state deviates from the domestic political communication baseline, computed as $\frac{\text{Occupy}}{\text{Domestic}}$. Redder colors indicate that proportionally more Occupy content originated from the associated state, while whiter colors indicate the opposite. To minimize the effect of outliers on the visualization and to highlight variation between states, colors for Maine and Oregon have been fixed, indicating that the deviation from baseline is more than three times the expected rate.

doi:10.1371/journal.pone.0055957.g002

Figure 3. Ratio of content production versus content consumption, by stream. Occupy Wall Street users, by state, exhibit a lower content production to consumption ratio relative to users in the domestic political communication stream. The disproportionately high ratio observed for Kentucky can be attributed to the activity of a prolific, highly popular left-leaning user from that state.

doi:10.1371/journal.pone.0055957.g003
high profile locations, we observe that these states represent sites of major encampment and decision making activity. Despite the fact that all users can contribute equally to the Occupy stream, it appears that proximity to events on the ground plays a major role in determining which content receives the most attention. This is in contrast to the stream of domestic political communication, in which content from users across the United States is allocated a significant share of attention. Where the stream of domestic political communication looks more like a conversation taking place at the national level, the structure of the Occupy stream is more akin to a broadcast, with just a few locations playing the role of net content producers.

Finally, we propose that interstate communication plays a significant role in the propagation of narrative imagery associated with collective framing processes, and that intrastate communication is driven more predominantly by the pressures of resource mobilization. Looking to the lists of tokens most overrepresented in each type of traffic (Table 1), we find that those more common in interstate communication include references to core framing language and the news media. This finding suggests that when users engage in communication across state boundaries they allocate proportionately higher levels of attention to speech associated with collective framing processes. In contrast, the tokens more common in intrastate traffic relate to protest action and specific times and places. From this we conclude that the

**Figure 4.** Multiscale backbone ($\alpha = .15$) of the continental interstate communication networks. Stable domestic political communication is shown at left, Occupy Wall Street at right. Edges adjacent to New York, California, and Washington D.C. are shown in red. Note that Occupy Wall Street exhibits a clear hub-and-spoke pattern, with the majority of traffic flowing to and from New York, California, and Washington, D.C. Likewise, observe that the Occupy Wall Street network exhibits diminished levels of interstate connectivity compared to the network of domestic political communication. We note that the structure of this network backbone is robust to different parameterizations of alpha. doi:10.1371/journal.pone.0055957.g004

**Figure 5.** Connectivity matrices describing directed interstate communication volume. The edge weight corresponding to each cell is mapped to a color hue on a logarithmic scale ranging from white for edges with the least weight to black for edges with the most weight. The strong diaonalization and limited off-diagonal mass apparent in the Occupy Wall Street matrix is indicative of highly localized communication activity. doi:10.1371/journal.pone.0055957.g005
content of intrastate tweets deals much more frequently with rallying the movement’s participants, a core function of resource mobilization.

The findings outlined in this paper dovetail nicely with established literature on social movement theory. However, statistical measures are limited in the extent to which they can accurately represent nuanced features of communication, and future work in this domain would benefit from rigorous qualitative content analysis. Moreover, there remains room to improve our understanding of how closely the structure of social media communication mirrors that of other forms of communication. For example, Mislove, et al. found that the geographical distribution of Twitter users tends to over-represent populous counties and metropolitan areas, suggesting that entire rural regions may be significantly under-represented – with similar findings holding true for ethnicity and gender as well [37]. In this respect as well, work of this nature would benefit from deeper involvement from scholars in the social sciences, and we hope that this type of interdisciplinary collaboration will become increasingly common.

Acknowledgments

We would like to thank Alex Radnick, Jacob Ratkiewicz, Mark Meiss, and other current and past members of the Truthy group at Indiana University (cnets.indiana.edu/groups/nan/truthy) for their contributions to the Truthy Project. Additionally, we would like to thank Fabio Rojas, Brian Keegan, and Bruno Gonçalves for their constructive, insightful feedback during the preparation of this manuscript.

Author Contributions

Conceived and designed the experiments: MDC AF FM EF CD KM. Performed the experiments: MDC CD. Analyzed the data: MDC AF FM EF CD KM. Wrote the paper: MDC AF FM EF.

References

1. Chomsky N (2012) Occupy. Zuccotti Park Press.
2. Byrne J (2012) Occupy the media: Journalism for (and by) the 99 percent. In: Byrne J, editor, The Occupy Handbook. Little, Brown, 256–264.
3. Myers D (1994) Communication technology and social movements: Contributions of computer networks to activism. Social Science Computer Review 12: 250–260.
4. Van Laer J, Van Aelst P (2009) Cyber-protest and civil society: the internet and action repertoires in social movements. Handbook on Internet Crime : 230–254.
5. Bray S (1999) On electronic civil disobedience. Peace Review 11: 107–111.
6. Bennett W (2003) Changing citizenship in the digital age. The John D and Catherine T MacArthur Foundation Series on Digital Media and Learning: 1–24.
7. Bennett W (2003) Communicating global activism: Strength and vulnerabilities of networked politics. Information, Communication & Society 6: 143–168.
8. McCarthy J, Zald M (1977) Resource mobilization and social movements: A partial theory. American Journal of Sociology: 1212–1241.
9. Benford R, Snow D (2000) Framing processes and social movements: An overview and assessment. Annual Review of Sociology: 611–639.
10. Garrett R (2006) Protest in an information society: A review of literature on social movements and new ICTs. Information, Communication & Society 9: 202–224.
11. MotherJones website (Accessed 20 Jan 2013); “We Are the 99 Percent!” Creators Revealed. URL http://www.motherjones.com/politics/2011/10/we-are-the-99-percent-creators.
12. Choudhary A, Hendrix W, Lee K, Paletta D, Liao W (2012) Social media evolution of the egyptian revolution. Communications of the ACM 55: 74–80.
13. Lotan G, Graeff E, Ananny M, Gaffney D, Pearce I, et al. (2011) The revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. International Journal of Communication 5.
14. González-Balbastro S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. Scientific reports 1.
15. Java A, Song X, Finin T, Tseng B (2007) Why we Twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 56–63.
16. Krishnamurthy B, Gill P, Arlt M (2008) A few chirps about Twitter. In: Proceedings of the 1st workshop on Online social networks. ACM, 19–24.
17. Huberman B, Romero D, Wu F (2009) Social networks that matter: Twitter under the microscope. First Monday 14: 1–3.
32. De Choudhury M, Counts S, Gamon M (2012) Not all moods are created equal! exploring human emotional states in social media. In: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media.
33. Ciulla F, Mocanu D, Baronchelli A, Goncalves B, Perra N, et al. (2012) Beating the news using social media: the case study of american idol. EPJ Data Science 1: 8.
34. Onnela J, Arbesman S, Gonzalez M, Barabasi A, Christakis N (2011) Geographic constraints on social network groups.
35. Ratcliff J, Metzner D, et al. (1988) Pattern matching: The gestalt approach. Dr Dobb’s Journal 7: 46.
36. Serrano M, Boguna M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proceedings of the National Academy of Sciences 106: 6483–6488.
37. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J (2011) Understanding the demographics of Twitter users. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media.
The Digital Evolution of Occupy Wall Street

Michael D. Conover*, Emilio Ferrara, Filippo Menczer, Alessandro Flammini
Center for Complex Networks and Systems Research
School of Informatics and Computing, Indiana University, Bloomington, IN USA
* Corresponding Author: midconov@indiana.edu

Abstract

We examine the temporal evolution of digital communication activity relating to the American anti-capitalist movement Occupy Wall Street. Using a high-volume sample from the microblogging site Twitter, we investigate changes in Occupy participant engagement, interests, and social connectivity over a fifteen month period starting three months prior to the movement’s first protest action. The results of this analysis indicate that, on Twitter, the Occupy movement tended to elicit participation from a set of highly interconnected users with pre-existing interests in domestic politics and foreign social movements. These users, while highly vocal in the months immediately following the birth of the movement, appear to have lost interest in Occupy related communication over the remainder of the study period.

Introduction

Information communications technologies play a crucial role in the development and persistence of many modern social movements [1–3]. Among these, the American anti-capitalist movement Occupy Wall Street (‘Occupy’) is remarkable for the prominent role social media, and in particular Twitter, played in facilitating communication among its participants [4, 5]. Functioning as a high-visibility forum in which adherents and prospective participants could interact and share information, Twitter represented a valuable resource for supporting the movement’s political and social objectives. In time, however, activity on the platform substantially diminished, mirroring the fading prominence of protest action on the ground. In light of this decline, we seek to understand more about the population from which Occupy drew its support, and specifically whether these individuals exhibited changes in behavior or social connectivity over the course of the movement’s evolution.

The Twitter platform, like other information communication technologies, has the potential to confer a number of benefits to burgeoning social movements [6–8]. Chief among these is the opportunity to connect individuals in service of the dual goals of resource mobilization and collective framing [9]. These factors, well studied in the social sciences literature, are critical to the success of social movements. Resource mobilization refers to the process whereby a social movement works to marshal the physical and technological infrastructure, human resources, and financial capital necessary to sustain its ongoing activity [10,11]. Collective framing refers to the social processes whereby movement participants negotiate the shared language and narrative frames that help define the movement’s identity and goals [12,13].

In related work [9], we report on evidence that Occupy users leveraged Twitter to communicate, at the local level, time-sensitive information about protest and police action. We also find that users relied on these channels to facilitate interstate communication relating to the news media and narrative frames such as “We are the 99%,” suggesting that long-distance communication on Twitter played a role in the collective framing processes that imbue social movements with a shared language, purpose and identity. This evidence indicates that Occupy participants used the Twitter platform to address critical issues facing any burgeoning social movement, and that during peak periods these streams were rich with actionable, relevant information.

To establish the extent of Occupy participant engagement with Twitter over time, here we study the total amount of Occupy-related traffic on the platform from September 2011 through September 2012. With respect to this measure of activity, we find that Occupy traffic has diminished by orders of
magnitude relative to peak activity volumes in late 2011. This effect is evident even in concerted attempts to revive the movement’s flagging levels of engagement, with activity returning to baseline within a week of May 1st, 2012 reoccupation efforts.

Finding little evidence of sustained activity, we turn our attention to Occupy participants themselves, in hopes of understanding how these users were changed as a result of engaging with the movement online. Using a random sample of 25,000 Occupy users, we study changes in behavior at the individual level with respect to attention allocation and social connectivity. From this analysis we are left to conclude that, on Twitter, Occupy evoked interest from a highly-interconnected community of users with pre-existing interest in domestic politics and foreign social movements. Though we find statistically significant changes in political interests and social connectivity over the study period, the magnitude of these changes pales in comparison to the amount of attention these individuals allocated to the Occupy Wall Street cause.

Materials and Methods

Twitter Platform

Twitter is a social networking platform that allows individuals to consume content from and contribute content to streams comprised of 140-character messages known as tweets \[14\]. The Twitter stream has been extensively explored in the recent literature, with focus on user activity modeling \[15–19\], content classification \[20–23\], sentiment analysis \[24–26\] and event detection \[27–29\]. Broadly speaking, there are two types of content streams: those associated with individual accounts and those associated with topic-specific tokens known as hashtags. By following one or more accounts, a user creates a personalized feed that aggregates into a single, private stream the content produced by the followed accounts. Hashtags, short tokens prepended with a pound sign (e.g., #taxes or #obama), allow the content produced by many individuals to be aggregated into a public, topic-specific stream including all the tweets containing a given token.

Although by default each user’s tweets are publicly visible, the audience for an individual’s content is largely limited to his or her network of immediate followers, attaining greater levels of visibility only when it is rebroadcast by large numbers of other users. By including a hashtag in a tweet, however, an individual can contribute content to a high-profile stream, and thereby engage with users who might never otherwise see the content. It is this kind of communication, which represents engagement with a topically cohesive community of users unconstrained by social network structure, that is the primary focus of this study.

In addition to engaging with different content streams, users can interact with one another in two primary ways. A user can retweet content produced by another individual, rebroadcasting it to his or her audience of followers, or mention another user in a tweet, which functions as a publicly-visible message targeting that individual.

Data

We rely on two primary datasets extracted over a 15-month period from an approximately 5-10% sample of the entire public Twitter stream (https://dev.twitter.com/docs/streaming-apis(streaming-apis/streams/public)). In addition to information about the content and users associated with a tweet, the Twitter streaming API provides timestamp metadata that allow for the historical reconstruction of the time series presented in this study.

To identify Occupy-related content, we deem relevant any tweet containing a hashtag matching either #ows or #occupy*, where * represents a wildcard character. This set includes high-profile tags such as #occupy as well as location-specific tokens such as #occupyoakland and #occupyseattle. While this approach does not allow us to study content that does not contain an Occupy-specific hashtag, we argue
that it is appropriate for two reasons. As outlined above, hashtags allow a user to reach an audience beyond his or her immediate followers, and it is this kind of expressly public engagement in which we are primarily interested. Moreover, while topic modeling techniques may allow for the analysis of untagged tweets, their use would introduce noise that could cloud the interpretation of any analytical results [30]. Based on the criteria outlined above, we produce a corpus of all sampled tweets containing at least one of these hashtags from the year-long period between September 1st, 2011 to August 31st, 2012. Referred to hereafter as the Occupy corpus, this dataset contains approximately 1.82 million tweets produced by 447,241 distinct accounts.

In addition to changes in activity explicitly related to the Occupy movement, we are also interested in changes to the behavior of individual users over time. To this end, we identified a random sample of 25,000 random users who produced at least one tweet in the Occupy corpus. We then produced a second corpus containing any tweet, regardless of content, produced by each account in this sample during the 15-month period spanning June 1st, 2011 through August 31st, 2012. Including tweets from the three-month period preceding the start of the Occupy Wall Street movement allows us to study the behavior of these users before, during, and after the movement’s primary period of activity. Referred to hereafter as the random sample, this dataset contains approximately 7.74 million tweets produced by 25,000 unique users.

To facilitate analysis relating to the attention allocation habits of these individuals, we rely on three non-overlapping sets of hashtags: those related to Occupy Wall Street (defined above), a second set relating to foreign social movements, and a third relating to domestic political communication. As we are interested exclusively in the attention allocation habits of Occupy users, we identified the set of hashtags relating to domestic political communication and foreign social movements by manually inspecting the 300 hashtags most frequently used by individuals in the random sample. Table 1 lists the hashtags associated with each topic. While not exhaustive due to a long-tail use distribution, the 300 most popular hashtags account for 70.8% of all tagging activity, with the 300th most popular tag constituting just 0.027% of all tags. We therefore believe that the inclusion of additional tags in our topic lists is not likely to affect the results of this study.

Methods

All of the analyses in this article rely on time series describing changes to measured quantities over the course of the study period. Each time series is produced by computing a single statistic on disjoint sets of tweets partitioned into adjacent, temporally non-overlapping bins of $k$ hours. For all of these analyses we use one of three temporal resolutions to reveal different characteristics of the signal under study: 12 hours, 24 hours, or one week.

At various times over the course of the study period, our system experienced service outages that affected our ability to collect data from the Twitter API. Amounting to 15 days in total, these periods are: September 29 to October 4, 2011; October 11–12, 2011; December 28–30, 2011; February 11–13, 2012; February 16–17, 2012; and May 28–31, 2012. Owing to the fact that the measures we employ reflect relative composition of the stream rather than its absolute volume, these outages do not unduly influence the statistical character of our results.

Results

Let us first focus to the total number of tweets in the Occupy corpus over the course of the year. Figure 1 shows that, in general, Occupy traffic closely mirrors activity on the ground, and is characterized by peak levels during the month-long period following the movement’s initial protests, with significantly diminished activity levels over the following eleven months. In terms of relative change, average levels of
Occupy traffic in the second half of the period from September 17th, 2011 to August 31st, 2012 decreased 80.8% relative to the first half of the same period.

In light of this finding, we wish to gain insights into the character of the individuals from which Occupy drew its support. We begin by studying how Occupy user interests changed in time, examining the frequency with which 25,000 random individuals produced content relating to one of three topics: Occupy Wall Street, foreign social movements, and domestic politics. Based on the random sample described in §Data, the results of this analysis describe activity from June 1, 2011 to August 31, 2012, a period including the three months prior to the initial protest action.

As we are interested in the behavior of individuals who were active on Twitter at a given time, we identify the set of users \(U_i\) from whom we observe at least one tweet at time step \(i\), regardless of its content. Within this set we isolate, at each timestep, the set of users \(U_{it}\) from whom we observe, in any of their tweets, at least one hashtag relating to topic \(t\). The *engaged user ratio* \(|U_{it}|/|U_i|\) describes the extent to which individuals chose to engage in communication relating to each of the three topic areas.

Among the set of users engaged with a topic, we next examine the extent to which that topic tends to dominate their content production activity. To accomplish this, let us consider, for each user \(u \in U_{it}\), the collection \(H_{it}\) of hashtags contained in his or her tweets at time step \(i\). From this we compute the proportion of each user’s tagging activity that is associated with a given topic, \(|H_{it}|/|H_{iu}|\), where \(H_{iu}\) is the set of tags from topic \(t\) produced by \(u\) at time step \(i\). Averaging this value across all engaged users provides a lens on the behavior of these individuals as a whole, and is reported as the *engaged user attention ratio*. Figure 2 presents this value alongside the engaged user ratio to show how the amount of attention allocated to the three topics changed over time.

As expected, a large fraction of users produced Occupy related content during the period of peak activity, with more than 40% of sampled users allocating on average 64% of their attention to the topic during the third week following the initial protests. However, this intense focus on the subject is not sustained over the course of the following year, with the engaged user ratio decaying to less than 5% in the last three months of the study period. Moreover, comparing the engaged user attention ratios from the first half of the period following the initial Occupy protests (\(\mu = .439\)) to those from the second half (\(\mu = .318\)), we find that individuals who continue to produce Occupy content do so with significantly lower frequency. Computed using a two location t-test for a difference in sample means, we reject the null hypothesis (\(p < 10^{-3}\)) that the mean of the engaged user attention ratios in the first half of the study period is greater than or equal to that of the observations in the second half of the study period, a finding suggestive of diminished enthusiasm even among the most persistent individuals.

With respect to foreign social movements and domestic political communication, we observe that users who would go on to engage with the Occupy movement online tended to exhibit interest in these topics before the initial protest activity in September, 2011. Comparing the engaged user ratios in the first 12 weeks of the study period with those observed during the last 12 weeks of the study period, we find a significant but small increase in domestic political communication activity. This conclusion is based on a two location t-test for a difference in sample means, in which we reject the null hypothesis (\(p < 10^{-3}\)) that the mean of the engaged user ratios from the first twelve weeks (\(\mu = 0.066\)) is greater than or equal to that of the latter half of the study period (\(\mu = 0.077\)). With respect to interest in foreign social movements, we observe a significant (\(p < 0.05\)) but small decrease in engagement for the same periods (from \(\mu = 0.074\) to \(\mu = 0.057\)). These differences suggest that the changes in individual behaviors in response to the Occupy Wall Street movement were limited.

Finally, let us examine the extent to which Occupy users tended to interact with one another over the course of the study period. To this end we focus on the proportion of retweets and mentions produced by active users in the random sample that involved another user who produced at least one Occupy-related tweet during the year following the movement’s inception. This proportion is computed with respect to all of a user’s retweets and mentions, regardless of content, rather than just those related to Occupy Wall Street. Inspecting the 95% confidence interval bands in Figure 3 we observe a statistically significant
increase in in-group retweet and mention activity during the peak period of Occupy activity, followed by a gradual decay to values approaching pre-Occupy levels. Comparing the fifteen week period before the inception of the movement to the fifteen week period at the study’s close, we use a two location t-test to identify a small but significant increase in both in-group retweets ($p < 10^{-6}$) and mentions ($p < 10^{-3}$), with the mean connectivity increasing 5.1% for retweets and 3.2% for mentions. Although these changes are statistically significant, they can hardly be interpreted as evidence that this community’s long-term social connectivity has been dramatically altered in response to participation in the Occupy Wall Street movement. Moreover, it’s notable that even in the period preceding the Occupy events, nearly 30% of these individuals’ targeted retweeting activity and almost a quarter of their mentioning were originated from or were directed to other Occupy users, suggesting that the movement elicited engagement from an already tightly interconnected community of users, rather than uniting disparate social groups behind a common cause.

**Discussion**

While interest and activity relating to the Occupy movement has substantially diminished, one could envision that increased levels of engagement with the political process online might constitute a positive outcome for the movement’s participants. Along these lines, however, Occupy users remain barely changed, exhibiting a slight increase in attention paid to domestic politics and a slight decrease in attention paid to foreign social movements. Relative to the dramatic behavioral changes these users exhibited in the early stages of the movement, and the magnitude of Occupy-related communication in general, these changes constitute a somewhat underwhelming long-term effect.

Similarly, a supporter of the movement might take as a promising outcome increased levels of interaction among Occupy users. Such a scenario could indicate that these individuals formed a more tight-knit community over the course of the year, creating social and communication bonds that may help to facilitate the efficient spread of information, potentially even reinforcing individual propensity for offline activity [31,32]. The data, however, provide little evidence to indicate that Occupy precipitated a dramatic rewiring of these users’ information sharing networks. While we observe significant increases in the proportion of in-group retweet and mention activity during the movement’s peak, the trend suggests that these values are slowly returning to those observed before the movement’s birth. What’s more, in the months preceding the initial protests we find evidence indicating that these users were already highly interconnected, with more than a quarter of their directed communication (either retweeting or mentioning) involving another individual who would go on to create Occupy related content.

Taken together, these data suggest that, on Twitter, the Occupy movement tended to elicit participation from a set of highly interconnected users with pre-existing interests in domestic politics and foreign social movements. These same users, while highly vocal in the months immediately following the movement’s birth, appear to have lost interest in Occupy-related communication over the remainder of the study period, and have exhibited only marginal changes in their attention allocation habits and social connectivity as a result of their participation.

These findings should not be taken to suggest that the Occupy movement itself has failed, as an argument can be made that the movement played a role in increasing the prominence of social and economic inequality in the public discourse. Though it would be unreasonable to argue that users could have maintained the frenetic pace of Occupy’s earliest days, it is doubtless that supporters may have hoped for a more sustained discourse than is evident from the near-complete abandonment of these once high-profile communication channels.
Acknowledgments

Thanks to Twitter for making data available through their streaming API; and to Karissa McKelvey, Clayton Davis, Bruno Gonçalves, Jacob Ratkiewicz, and other current and past members of the Truthy Project at Indiana University (cnets.indiana.edu/groups/nan/truthy) for facilitating access to this data.

References

1. Garrett R (2006) Protest in an information society: A review of literature on social movements and new ICTs. Information, Communication & Society 9: 202–224.

2. Bennett W (2003) Communicating global activism: Strength and vulnerabilities of networked politics. Information, Communication & Society 6: 143–168.

3. Myers D (1994) Communication technology and social movements: Contributions of computer networks to activism. Social Science Computer Review 12: 250–260.

4. Chomsky N (2012) Occupy. Zuccotti Park Press.

5. Byrne J (2012) Occupy the media: Journalism for (and by) the 99 percent. In: Bryne J, editor, The Occupy Handbook. Little, Brown, pp. 256-264.

6. Lotan G, Graeff E, Ananny M, Gaffney D, Pearce I, et al. (2011) The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. International Journal of Communication 5: 1375–1405.

7. Howard P, Duffy A, Freelon D, Hussain M, Mari W, et al. (2011) Opening closed regimes: What was the role of social media during the arab spring. Technical Report 2011.1, Project on Information Technology and Political Islam.

8. Aday S, Farrel H, Lynch M, Sides J, Kelly J, et al. (2010) Blogs and bullets: New media in contentious politics. Technical report, U.S. Institute of Peace.

9. Conover MD, Davis C, Ferrara E, McKelvey K, Menczer F, et al. (2013) The geospatial characteristics of a social movement communication network. PLoS ONE 8: e55957.

10. McCarthy J, Zald M (1977) Resource mobilization and social movements: A partial theory. American Journal of Sociology 82: 1212–1241.

11. Jenkins J (1983) Resource mobilization theory and the study of social movements. Annual Review of Sociology 9: 527–553.

12. Benford R, Snow D (2000) Framing processes and social movements: An overview and assessment. Annual Review of Sociology : 611–639.

13. Benford R (1997) An insider’s critique of the social movement framing perspective. Sociological Inquiry 67: 409–430.

14. Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World wide web. ACM, pp. 591–600.

15. Conover M, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F (2011) Predicting the political alignment of Twitter users. In: Proceedings of the 3rd IEEE International Conference on Social Computing. IEEE, pp. 192–199.
16. Gonçalves B, Perra N, Vespignani A (2011) Modeling users’ activity on Twitter networks: validation of dunbar’s number. PLoS ONE 6: e22656.

17. Abel F, Gao Q, Houwen G, Tao K (2011) Analyzing user modeling on Twitter for personalized news recommendations. User Modeling, Adaption and Personalization (LNCS) 6787: 1–12.

18. Lehmann J, Gonçalves B, Ramasco J, Cattuto C (2012) Dynamical classes of collective attention in Twitter. In: Proceedings of the 21st International Conference on World Wide Web. ACM, pp. 251–260.

19. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. Scientific Reports 2.

20. Agichtein E, Castillo C, Donato D, Gionis A, Mishne G (2008) Finding high-quality content in social media. In: Proceedings of the International Conference on Web Search and Web Data Mining. ACM, pp. 183–194.

21. Boyd D, Golder S, Lotan G (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. IEEE, pp. 1–10.

22. Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE 5: e14118.

23. Wu S, Hofman J, Mason W, Watts D (2011) Who says what to whom on Twitter. In: Proceedings of the 20th International Conference on World wide web. ACM, pp. 705–714.

24. Diakopoulos N, Shamma D (2010) Characterizing debate performance via aggregated Twitter sentiment. In: Proceedings of the 28th International Conference on Human factors in computing systems. ACM, pp. 1195–1198.

25. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. Journal of Computational Science 2: 1–8.

26. Bollen J, Pepe A, Mao H (2011) Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. pp. 450–453.

27. Hughes A, Palen L (2009) Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management 6: 248–260.

28. Vieweg S, Hughes A, Starbird K, Palen L (2010) Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the 28th International Conference on Human factors in computing systems. ACM, pp. 1079–1088.

29. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World wide web. ACM, pp. 851–860.

30. Hong L, Davison B (2010) Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics. ACM, pp. 80–88.

31. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. Scientific Reports 1.

32. Karpf D (2012) Online political mobilization from the advocacy group’s perspective: looking beyond clicktivism. Policy & Internet 2: 7–41.
Figure Legends

Figure 1. Total number of tweets related to Occupy Wall Street between September 2011 and September 2012. Each timestep represents a 12-hour period, with vertical blue bars overlaid on periods during which access to the Twitter streaming API was interrupted. Large bursts in activity tend to correspond to protest or police action on the ground, demarcated with circles. From left to right, the events are: initial Occupy Wall Street protest in Zuccotti Park; initial NYPD arrests of protesters; march from Foley Square to Zuccotti Park; protest at U.S. Armed Forces recruiting station in Times Square; protest in support of Iraq veteran injured by police-fired projectile; NYPD action to clear Zuccotti Park; protest against eviction from Zuccotti Park; first round of Egyptian elections; ‘May Day’ general strike and planned reoccupation of former encampments.

Figure 2. Attention allocation of 25,000 randomly selected Occupy users to each of three topics: Occupy Wall Street, domestic politics, and revolutionary social movements. Engaged User Ratio describes the proportion of active users in each timestep who produced at least one topically-relevant tweet. Engaged User Attention Ratio describes, among these users, the share of average attention allocated to each topic. The Engaged User Attention Ratio did not exhibit meaningful trends for either domestic politics or foreign social movements, and so it is omitted from the figure for sake of visual clarity. Refer to \textsection Results for the full derivation of these measures. The dashed vertical line corresponds to the date of the first Occupy protest.
Figure 3. Proportion of all retweet and mention traffic, regardless of content, from 25,000 randomly selected Occupy users involving another individual who produced at least one Occupy-related tweet. Shown are means and 95% confidence intervals for each time step. The dashed vertical line corresponds to the date of the first Occupy protest.

Tables
Table 1. Lists of topic-specific hashtags. Hashtags were manually selected from among the 300 most frequently used by individuals in the 25,000-person random sample of Occupy users.

| Domestic Politics   | Social Movements |
|---------------------|------------------|
| #tcot               | #syria           |
| #p2                 | #bahrain         |
| #teaparty           | #egypt           |
| #gop                | #yemen           |
| #anonymous          | #libya           |
| #obama              | #tahrir          |
| #tlot               | #wiunion         |
| #jobs               | #iranelection    |
| #ronpaul            | #assange         |
| #romney             | #wikileaks       |
| #sopa               | #jan25           |
| #ndaa               | #14feb           |
| #obama2012          | #assad           |
| #ocra               | #greece          |
| #twisters           | #damascus        |
| #sgp                | #gaddafi         |
| #politics           | #feb14           |
| #solidarity         | #scaf            |
| #gop2012            | #antisec         |
| #p21                | #arabspring      |
| #topprog            | #tunisia         |
| #obamacare          | #noscaf          |
| #mapoli             | #syrian          |
| #acta               |                  |
| #sotu               |                  |
| #newt               |                  |
| #santorum           |                  |
| #mittromney         |                  |
| #gopdebate          |                  |
| #dem                |                  |
What role did social media play in recent manifestation of social unrest?

What can be “discovered” about social movements by analyzing their “reflection” on social media?
The rise and decline of the Occupy Wall Street movement from a digital perspective

Alessandro Flammini
Indiana University

Thanks to:
THE OCCUPY PROTEST

- Social and wealth inequalities, taxation disparities
- Non sustainable capitalistic market model
- Political corruption, corporate influence of government

- Pacific movement
- Occupation of public places
- Resistance to authorities
# TIMELINE

**Few key dates**

| Date       | Event                                                                 |
|------------|----------------------------------------------------------------------|
| 7 Jul. 2011 | Call for action by magazine Adbuster                                 |
| 17 Sept. 2011 | Occupation of Zuccotti Park in NY                                    |
| 15 Oct. 2011 | Public places are occupied world wide                               |
| 15 Nov. 2011 | Police starts clearing Zuccotti Park                                |
| Dec. 2011   | Intensity of movement starts to slowly decline, punctuated by many attempts to revive it |
TWITTER TRAFFIC REFLECTS INTENSITY of MOVEMENT

Occupy Wall Street Tweet Volume (12 Hour Bins)

Occupy Hashtag Use Over Time

Event
- Initial Protests
- Initial Arrests
- Foley Square March
- USAF Protests
- Zuccotti Cleared
- Counter Protest
- Egyptian Elections
- May Day Strike

Hashtag
- #occupy
- #occupywallstreet
- #ows
DATA

Occupy corpus

From Jul 3rd, 2011 to Aug. 31st, 2012
Any tweet containing #ows or #occupy+"anything"
Total of 1.5M tweets [676K RT]
Total of 257K distinct users

Domestic politics corpus (baseline)

Any tweet containing #tcot (top conservatives on Twitter) or #p2 (Progressives 2.0)
Total of 825K tweets [259K RT]
Total of 68K distinct users
NY, CA and DC account for more than 60% of traffic less than 30% in political conversations.
**Collective framing**: the social processes whereby movement participants negotiate the shared language and narrative frames that help define the movement's identity and goals.

**Resource mobilization**: the work to marshal the physical and technological infrastructure, human resources, and financial capital necessary to sustain ongoing activity.
LOCAL AND GLOBAL HASHTAGS

Regional Hashtag Use Over Time

Hashtag
- #occupyboston
- #occupycabinet
- #occupydc
- #occupydanver
- #occupyla
- #occupylsx
- #occupynigeria
- #occupyoakland
- #occupysf

Proportion of Traffic

09/11 10/11 11/11 12/11 01/12 02/12 03/12 04/12 05/12 06/12 07/12 08/12 09/12
Occupy discourse on Twitter has highly localized geospatial structure: a large amount of traffic is produced and consumed locally per state. This might be explained by Resource Mobilization.

Interstate communication is driven by high-profile locations acting as information broadcasters. This might represent the Collective Framing process.

Proximity to on-the-ground events plays a big role: users from NY, DC and California are the main actors of the discourse. They produce much more Occupy-related information than that they consume, unlike other states.
| Domestic Politics       | Social Movements   |
|------------------------|--------------------|
| #tcot                  | #syria             |
| #p2                    | #bahrain           |
| #teaparty              | #egypt             |
| #gop                   | #yemen             |
| #anonymous             | #libya             |
| #obama                 | #tahrir            |
| #tlot                  | #wiunion           |
| #jobs                  | #iranelection      |
| #ronpaul               | #assange           |
| #romney                | #wikileaks         |
| #sopa                  | #jan25             |
| #ndaa                  | #14feb             |
| #obama2012             | #assad             |
| #ocra                  | #greece            |
| #twisters              | #damascus          |
| #sgp                   | #gaddafi           |
| #politics              | #feb14             |
| #solidarity            | #scaf              |
| #gop2012               | #antisec           |
| #p21                   | #arabspring        |
| #topprog               | #tunisia           |
| #obamacare             | #noscaf            |
| #mapoli                | #syrian            |
Figure 2. Attention allocation of 25,000 randomly selected Occupy users to each of three topics: Occupy Wall Street, domestic politics, and revolutionary social movements. Engaged User Ratio describes the proportion of active users in each timestep who produced at least one topically-relevant tweet. Engaged User Attention Ratio describes, among these users, the share of average attention allocated to each topic. The Engaged User Attention Ratio did not exhibit meaningful trends for either domestic politics or foreign social movements, and so it is omitted from the figure for sake of visual clarity. Refer to §Results for the full derivation of these measures. The dashed vertical line corresponds to the date of the first Occupy protest.
Figure 3. Proportion of all retweet and mention traffic, regardless of content, from 25,000 randomly selected Occupy users involving another individual who produced at least one Occupy-related tweet. Shown are means and 95% confidence intervals for each time step. The dashed vertical line corresponds to the date of the first Occupy protest.
NEW PARTICIPANTS and WHEN THEY ENTER
Occupy captured the attention of users with pre-existing interests in politics and social protests.

Occupy users' interests and amount of attention dedicated to politics and social discourse did not exhibit any remarkable variation over time.

Occupy users were already highly interconnected before the movement’s start.

The extent to which Occupy users interacted before the movement’s start and at the end of the observation period is substantially unchanged.
A Pointer to Our Framework Project

Truthy

Information diffusion research at Indiana University

Truthy is a research project that helps you understand how communication spreads on Twitter. We currently focus on tweets about politics, social movements and news from the past 90 days.

For Researchers

- **DATA, STATISTICS, AND VISUALIZATIONS**
  - Study online communication networks with interactive interfaces that visualize data and allow you to identify interesting users.
  - Download recent tweets, tweet volume over time, network layout, and statistics about users such as predicted political partisanship, sentiment scores, language, and activity.

For Citizens

- **GALLERY**
  - Read descriptions of interesting memes and take a look at static images to learn what is possible with this research.
- **MOVIES**
  - Create movies of communication over time to see how a hashtag on Twitter evolved during a given time period.

For Journalists

- **POLITICAL COVERAGE**
  - Identify prominent events and pay attention to key players with our interactive visualizations of U.S. political conversation.

For Developers

- API integration.
  - Write scripts to work with our data, statistics, and images using the API.

Recent Updates

- **truthy.indiana**
  - `truthy.indiana` is tracking at least 546733 different memes, which ones are the most interesting? `truthy.indiana` 27 days ago — @truthy.indiana
  - `truthy.indiana` is tracking at least 50000 different memes, which ones are the most interesting? `truthy.indiana` 19 days ago — @truthy.indiana
  - `truthy.indiana` recent Truthy movies about AI, AI, A.I., AI, AI.

System Info

- **New Tweets**
- **All Tweets**

From the Gallery

- A smear campaign against a U.S. Senate candidate by many Twitter bots controlled by one person.

truthy.indiana.edu/
