

Partisan asymmetries in online political activity

Michael D Conover¹, Bruno Gonçalves^{2*}, Alessandro Flammini¹ and Filippo Menczer¹

*Correspondence:

b.goncalves@neu.edu

²College of Computer and Information Sciences, Northeastern University, Boston, MA 02115, USA
Full list of author information is available at the end of the article

Abstract

We examine partisan differences in the behavior, communication patterns and social interactions of more than 18,000 politically-active Twitter users to produce evidence that points to changing levels of partisan engagement with the American online political landscape. Analysis of a network defined by the communication activity of these users in proximity to the 2010 midterm congressional elections reveals a highly segregated, well clustered, partisan community structure. Using cluster membership as a high-fidelity (87% accuracy) proxy for political affiliation, we characterize a wide range of differences in the behavior, communication and social connectivity of left- and right-leaning Twitter users. We find that in contrast to the online political dynamics of the 2008 campaign, right-leaning Twitter users exhibit greater levels of political activity, a more tightly interconnected social structure, and a communication network topology that facilitates the rapid and broad dissemination of political information.

1 Introduction

Digitally-mediated communication has become an integral part of the American political landscape, providing citizens access to an unprecedented wealth of information and organizational resources for political activity. So pervasive is the influence of digital communication on the political process that almost one quarter (24%) of American adults got the majority of their news about the 2010 midterm congressional elections from online sources, a figure that has increased three-fold since the Pew Research Center began monitoring the statistic during the 2002 campaign [1]. Relax the constraint that a majority of a person's political news and information must come from online sources and the figure jumps to include the 54% of adult Americans who went online in 2010 to get political information. Critically, this activity precipitates tangible changes in the beliefs and behaviors of voters, with 35% of Internet users who voted in 2010 reporting that political information they saw or read online made them decide to vote for or against a particular candidate [1].

Within this ecosystem of digital information resources, social media platforms play an especially important role in facilitating the spread of information by connecting and giving voice to the voting public [2–4]. Networked and unmoderated, social media are characterized by the large-scale creation and exchange of user-generated content [5], a production and consumption model that stands in stark contrast to the centralized editorial and distribution processes typical of traditional media outlets [6, 7].

In terms of political organization and engagement, the benefits of social media use are many. For voters, social media make it easier to share political information, draw attention

to ideological issues, and facilitate the formation of advocacy groups with low barriers to entry and participation [8, 9]. The ease with which individual voters can connect with one another directly also makes it easier to aggregate small-scale acts, as in the case of online petitions, fundraising, or web-based phone banking [10]. Together, these features contribute to the widespread use of social media for political purposes among the voting public, with as many as 21% of online adults using social networking sites to engage with the 2010 congressional midterm elections [11]. Moreover, a survey by the Pew Internet and American Life Project finds that online political activity is correlated with more traditional forms of political participation, with individuals who use blogs or social networking sites as a vehicle for civic engagement being more likely to join a political or civic group, compared to other Internet users [12].

Likewise, candidates and traditional political organizations benefit from a constituency that is actively engaged with social media, finding it easier to raise money, organize volunteers and communicate directly with voters who use social media platforms [13]. Social media also facilitate the rapid dissemination of political frames, making it easy for key talking points to be communicated directly to a large number of constituents, rather than having to subject messages to the traditional media filter.

Considered in this light, it becomes clear why social media were argued to have played such an important role in the political success of the Democratic party in the 2008 presidential and congressional elections [14–16]. Survey data from the Pew Research Center showed that, along the seven dimensions used to measure online political activity, Obama voters were substantially more likely to use the Internet as an outlet for political activity [17]. In particular, Obama voters were more likely than McCain voters to create and share political content, and to engage politically on an online social network [17]. Moreover, a 2009 Edelman report found that in addition to a thirteen million member e-mail list, the Obama campaign enjoyed twice as much web traffic, had four times as many YouTube viewers and five times more Facebook friends compared to the McCain campaign [13]. While the direct effect of any one media strategy on the success of a campaign is difficult to assess and quantify, the data show that Obama campaign had a clear advantage in terms of online voter engagement.

Motivated by the connection between the widely reported advantage in on-line mobilization and the result of the 2008 presidential election, we seek to understand structural shifts in the American political landscape with respect to partisan asymmetries in online political engagement. We work toward this goal by examining partisan differences in the behavior, communication patterns and social interactions of more than 18,000 politically-active users of Twitter, a social networking platform that allows individuals to create and share brief 140-character messages. Among all social media services, Twitter makes an appealing analytical target for a number of reasons: the public nature of its content, the accessibility of the data through APIs, a strong focus on news and information sharing, and its prominence as a platform for political discourse in America and abroad [18, 19]. These features make a compelling case for using this platform to study partisan political activity.

For this analysis we build on the findings of a previous study which established the macroscopic structure of US domestic political communication on Twitter. In that work we employed clustering techniques and qualitative content analysis to demonstrate that the network of political retweets exhibits a highly segregated, partisan structure [20]. De-

spite this segregation, we found that politically left- and right-leaning individuals engage in interaction across the partisan divide using mentions, a behavior strongly correlated with a type of cross-ideological provocation we term ‘content injection.’

Having established the large-scale structure of these communication networks, in this study we employ a variety of methods to provide a more detailed picture of domestic political communication on Twitter. We characterize a wide range of differences in the behavior, communication, geography and social connectivity of thousands of politically left- and right-leaning users. Specifically, we demonstrate that right-leaning Twitter users exhibit greater levels of political activity, tighter social bonds, and a communication network topology that facilitates the rapid and broad dissemination of political information, a finding that stands in stark contrast to the online political dynamics of the 2008 campaign.

With respect to individual-level behaviors, we find that right-leaning Twitter users produce more than 50% more total political content and devote a greater proportion of their time to political discourse. Right-leaning users are also more likely to use hyperlinks to share and refer to external content, and are almost twice as likely than left-leaning users to self-identify their political alignment in their profile biographies. At the individual level, these behavioral factors paint a picture of a right-leaning constituency comprised of highly-active, politically-engaged social media users, a trend we see reflected in the communication and social networks in which these individuals participate.

Regarding connectivity patterns among users in these two communities we report findings related to three different networks, described by the set of explicitly declared follower/followee relationships, mentions, and retweets. Casting the declared follower network as the social substrate over which political information is most likely to spread, we find that right-leaning users exhibit a greater propensity for mutually-affirmed social ties, and that right-leaning users tend to form connections with a greater number of individuals in total compared to those on the left. With respect to the way in which information actually propagates over this substrate in the form of retweets, right-leaning users enjoy a network structure that is more likely to facilitate the rapid and broad dissemination of political information. Additionally, right-leaning users exhibit a higher probability to rebroadcast content from and to be rebroadcast by a large number of users, and are more likely to be members of high-order retweet network k -cores and k -cliques, structural features that are associated with the efficient spreading of information and adoption of political behavior and opinions. Pointing definitively to a vocal, socially engaged, densely interconnected constituency of right-leaning users, these topological and behavioral features provide a significantly more nuanced perspective on political communication on this important social media platform. Moreover, through its use of digital trace data to illuminate a complex sociological phenomenon, this article illustrates the explanatory power of data science techniques and underscores the potential of this burgeoning scientific epistemology.

2 Platform and data

2.1 The Twitter platform

Twitter is a popular social networking and microblogging site where users can post 140-character messages containing text and hyperlinks, called *tweets*, and interact with one another in a variety of ways. In the present section we describe four of the platform’s key features: follow relationships, retweets, mentions, and hashtags.

Twitter allows each user to broadcast tweets to an audience of users who have elected to subscribe to the stream of content he or she produces. The act of subscribing to a user's tweets is known as *following*, and represents a directed, non-reciprocal social link between two users. From a content consumption perspective, each user can sample tweets from a variety of content streams, including the stream of tweets produced by the users he or she follows, as well as the set of tweets containing specific keywords known as hashtags.

A hashtag is a tokens prepended with a pound sign (e.g., #token) which, when displayed, functions as a hyperlink to the stream of recent tweets containing the specified tag [21]. While they can be used to specify the topic of a tweet (e.g., #oil or #taxes), when used in political communication hashtags are commonly employed to identify one or more intended audiences, as in the case of the most popular political hashtags, #tcot and #p2, acronyms for 'Top Conservative on Twitter' and 'Progressives 2.0', respectively. In this way, hashtags function to broaden the audience of a tweet, extending its visibility beyond a person's immediate followers to include all users who seek out content associated with the tag's topic or audience. For this reason, as outlined in Section 3.1, we restrict our analysis to the set of tweets containing political hashtags, ensuring that the content under study is broadly public and expressly political in nature.

In addition to broadcasting tweets to the public at large, Twitter users can interact directly with one another in two primary ways: retweets and mentions. Retweets often act as a form of endorsement, allowing individuals to rebroadcast content generated by other users, thus raising the content's visibility [22]. Mentions allow someone to address a specific user directly through the public feed, or, to a lesser extent, refer to an individual in the third person. In this study, we differentiate between mentions that occur in the body of the tweet and those that occur at the beginning of a tweet, as they correspond to distinct modes of interaction. Mentions located at the beginning of a tweet are known as 'replies,' and typically represent actual engagement, while mentions in the body of a tweet typically constitute a third-person reference [23]. Together, retweets and mentions act as the primary mechanisms for explicit, public user-user interaction on Twitter.

2.2 Data

The analysis described in this article relies on data collected from the Twitter 'gardenhose' streaming API^a between September 1st and January 7th, 2011 - the eighteen week period surrounding the November 4th United States congressional midterm elections. The gardenhose provides a sample of approximately 10% of the entire Twitter corpus in a machine-readable format. Each tweet entry is composed of several fields, including a unique identifier, the content of the tweet (including hashtags and hyperlinks), the time it was produced, the username of the account that produced the tweet, and in the case of retweets or mentions, the account names of the other users associated with the tweet.

From this eighteen week period we collected data on 6,747 right-leaning users and 10,741 left-leaning users, responsible for producing a total of 1,390,528 and 2,420,370 tweets, respectively. It's useful to note that we evaluate all gardenhose tweets associated with each user, rather than just those containing political hashtags, in order to facilitate comparisons between the two groups in terms of relative proportions of attention allocated to political communication.

3 Methodology

In order to examine differences in the behavior and connectivity of left- and right-leaning Twitter users we rely on the political hashtags and partisan cluster membership labels established in a previous study on political polarization. In addition to reviewing the approach used to establish these features, we show that the networks and communities under study are representative of domestic political communication on Twitter in general.

3.1 Identifying political content

As outlined in Section 2.1, hashtags are used to specify the topic or intended audience of a tweet, and allow a user to engage a much larger potential audience than just his or her immediate followers. We define the set of pertinent political communication as any tweet containing at least one political hashtag. While an individual can engage in political communication without including a hashtag, the potential audience for such content is limited primarily to his or her immediate followers. Moreover, restricting our analysis to tweets which have been expressly identified as political in nature allows us to define a high-fidelity corpus, avoiding the risk of introducing undue noise through the use of topic detection strategies [24, 25].

To isolate a representative set of political hashtags and to avoid introducing bias into the dataset we performed a simple algorithmic hashtag discovery procedure. We began by seeding our sample with the two most popular political hashtags, #p2 ('Progressives 2.0') and #tcoot ('Top Conservatives on Twitter'). For each seed we identified the set of hashtags with which it co-occurred in at least one tweet, and ranked the results using the Jaccard coefficient. For a set of tweets S containing a seed hashtag, and a set of tweets T containing a second hashtag, the Jaccard coefficient between S and T is

$$\sigma(S, T) = \frac{|S \cap T|}{|S \cup T|}. \quad (1)$$

Thus, when the tweets in which both seed and the second hashtag occur make up a large portion of the tweets in which either occurs, the two are deemed to be related. Using a similarity threshold of 0.005 we identified sixty-six unique hashtags (Table 1), eleven of which were excluded due to overly-broad or ambiguous meanings (Table 2). While it is a common practice among spammers to contribute content to popular hashtag streams, we do not believe this phenomenon plays a substantial role in a shaping the structure of the sample data. During a previous study we found that of 1,000 manually-inspected accounts identified by this methodology fewer than 3% corresponded to foreign language or spam activity [20].

3.2 Representativeness

Using the technique outlined above we identified many high-profile political hashtags, and with them the majority of tweets and users associated with domestic political communication on Twitter. Supporting this claim, Figure 1 shows a roughly exponential decay in hashtag popularity as measured in terms of number of users or tweets associated with the hashtag. This sharp decay in the tag popularity indicates that the inclusion of additional political hashtags is not likely to substantially increase the size or alter the structure of the corpus.

Table 1 Political hashtags related to #p2 and #tcot (acronyms for 'Progressives 2.0' and 'Top Conservatives on Twitter')

Just #p2	#casen #dadt #dc10210 #democrats #du1 #fem2 #gotv #kysen #lgr #ofa #onation #p2b #pledge #rebelleft #truthout #vote #vote2010 #whyimvotingdemocrat #youcut
Both	#cspj #dem #dems #desen #gop #hcr #nvsen #obama #ocra #p2 #p21 #phnm #politics #sgp #tcot #teaparty #tlot #topprog #tpp #twisters #votedem
Just #tcot	#912 #ampat #ftrs #glennbeck #hhrs #iamthemob #ma04 #mapoli #palin #palin12 #spwbt #tsot #tweetcongress #ucot #wethepeople

Tweets containing any of these were included in our sample.

Table 2 Hashtags excluded from the analysis due to ambiguous or overly broad meaning

Excluded from #p2	#economy #gay #glbt #us #wc #lgbt
Excluded from both	#israel #rs
Excluded from #tcot	#news #qsn #politicalhumor

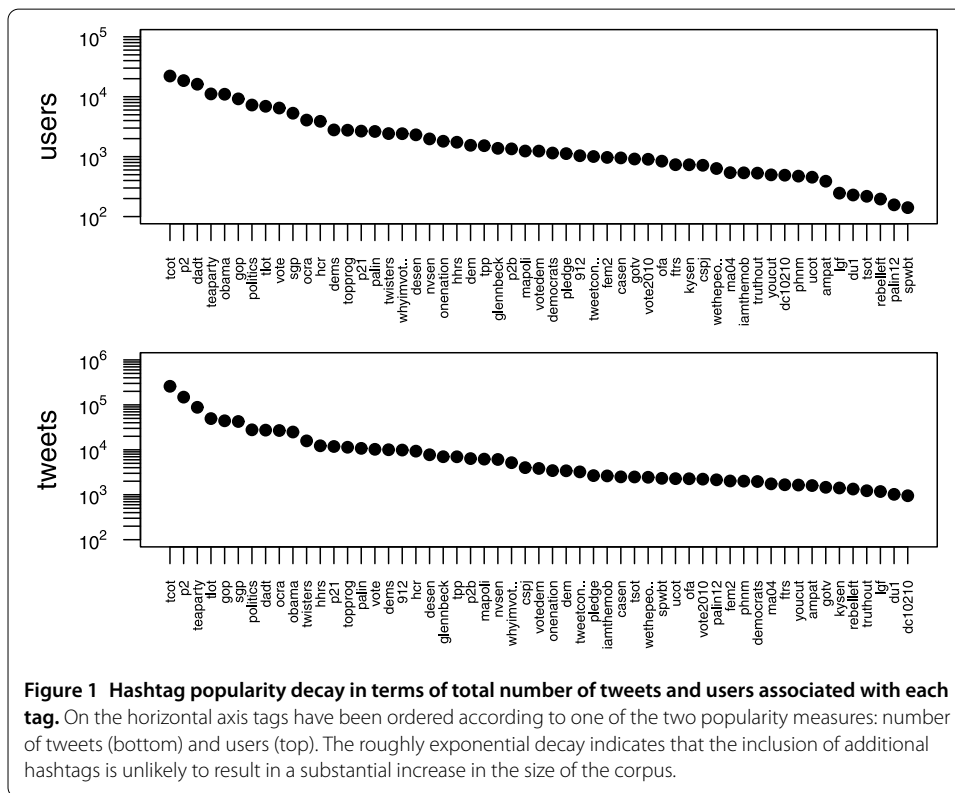
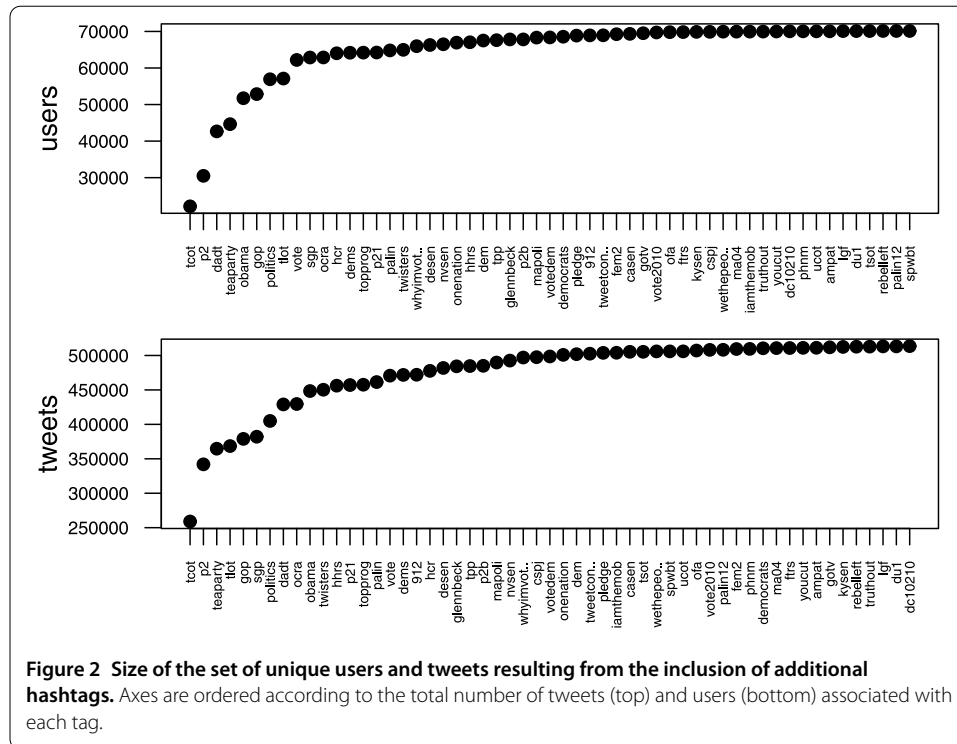


Figure 1 Hashtag popularity decay in terms of total number of tweets and users associated with each tag. On the horizontal axis tags have been ordered according to one of the two popularity measures: number of tweets (bottom) and users (top). The roughly exponential decay indicates that the inclusion of additional hashtags is unlikely to result in a substantial increase in the size of the corpus.

This claim is also supported by Figure 2, which shows that there is a strong effect of diminishing returns with respect to the observed number of unique users and tweets as the number of hashtags included in our analysis increases. This effect is due to the fact that many tweets are annotated with multiple hashtags, and many users utilize several different hashtags over the course of the study period. As a result, the inclusion of a single hashtag may result in the inclusion of many tweets and users also redundantly associated with other hashtags.

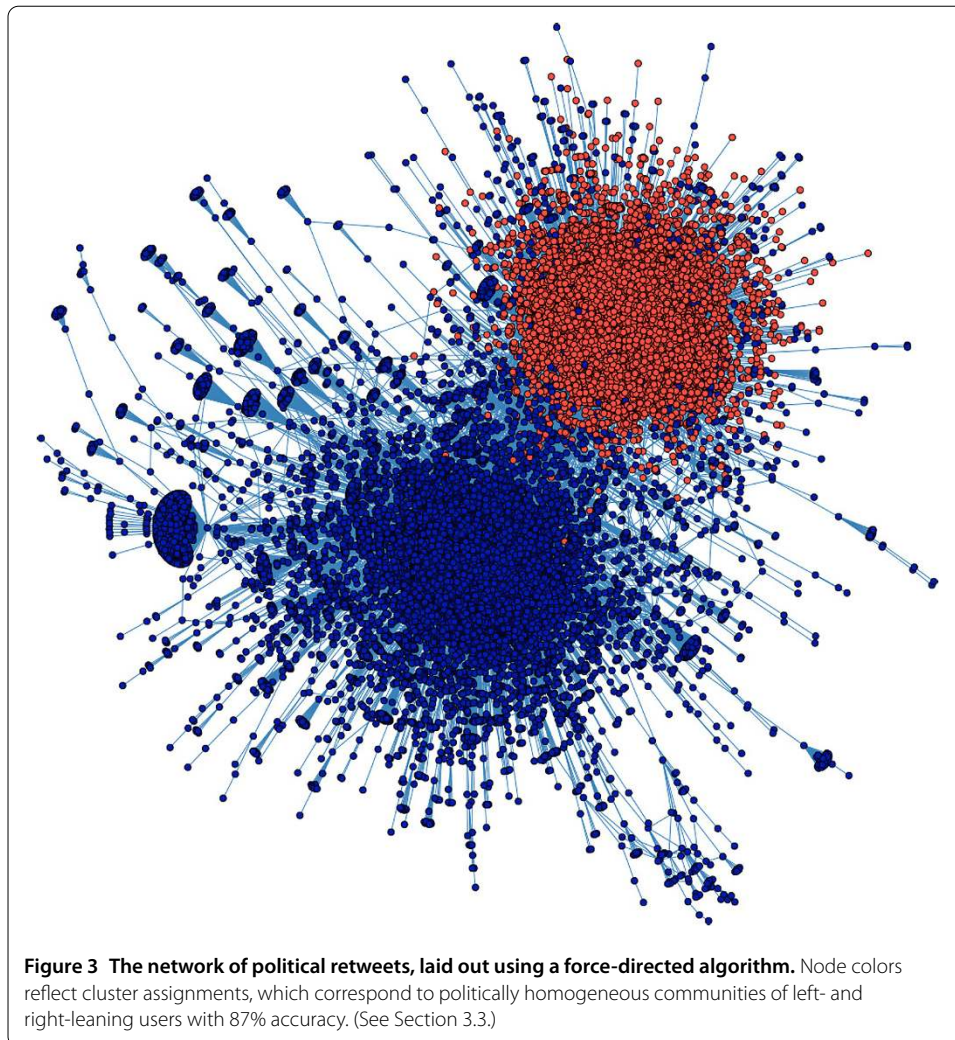


To further support the claim that sampling based on this set of hashtags produces a representative set of political tweets, we selected all the tweets in the gardenhose from the study period that included any one of 2,500 hand-selected political keywords related to the 2010 elections [26]. We considered only the 312,560 tweets in this set containing a hashtag because we use this characteristic to define public political communication on Twitter. We found that 26.4% of these tweets are covered by our target set of hashtags. Furthermore, among the ten most popular hashtags not included in our target set (#2010memories, #2010disappointments, #ff, #p2000, #2010, #business, #uk, #newsjp, #asia, #sports), only one is explicitly political and its volume accounts for less than 2% of public political communication. This coverage confirms that we have isolated a substantial and representative sample of political communication on Twitter.

3.3 Inferring political identities from communication networks

In a previous study we used the set of political tweets from the six weeks preceding the 2010 midterm election to build a network representing political retweet interactions among Twitter users. In this network an edge runs from a node representing user *A* to a node representing user *B* if *B* retweets content originally broadcast by *A*, indicating that information has propagated from *A* to *B*. This network consists of 23,766 non-isolate nodes among a total of 45,365, with 18,470 nodes in its largest connected component and 102 nodes in the next-largest component. We describe the construction of an analogous network of political mentions in Section 5.3.

Using a combination of network clustering algorithms and manually-annotated data we determined that the network of political retweets neatly divides the population of users in the largest connected component into two distinct communities (Figure 3) [20]. In brief, we used Rhagavan's label propagation method seeded with node labels determined by



Newman's leading eigenvector modularity maximization method to assign cluster membership to each node [27, 28]. The final community assignments are consistent and robust to fluctuations in starting conditions [20]. To determine whether these communities were composed of users from the political left and right, respectively, we used qualitative content analysis to evaluate the tweets produced by 1,000 random users appearing in the intersection of the mention and retweet networks [29, 30].

To establish the reproducibility of these results we had two authors, working independently, determine whether the content of a user's tweets express a 'left', 'right' or 'undecidable' political identity according to the coding rubric developed in a previous study [20]. These annotations were compared against the work of an independent non-author judge, and using a well-established measure of inter-annotator agreement we report 'nearly perfect' inter-annotator agreement between author and non-author annotations for the 'left' and 'right' classes (Cohen's Kappa values of 0.80 and 0.82, respectively) and 'fair to moderate' agreement for the 'undecidable' category (Cohen's Kappa value of 0.42) [29, 30]. From these high levels of inter-annotator agreement we conclude that an objective outside party would be able to reproduce our class assignments for most users.

Table 3 Partisan composition of retweet cluster communities as determined through manual annotation of 1,000 random users. (See Section 3.3)

Cluster	Left	Right	Undecidable	# Nodes
A (Top)	1.19%	93.4%	5.36%	7,115
B (Bottom)	80.1%	8.71%	11.1%	11,355

Based on this content analysis, we determined that the retweet network communities are highly politically homogeneous, consisting of 80.1% left- and 93.4% right-leaning users, respectively (Table 3) [20]. In this study we use network community membership as a proxy for the political identities of all 18,470 users in the largest connected component of the retweet network, and hereafter focus on the behavior of these users. Based on the relative proportions of right- and left-leaning users identified during the qualitative content analysis stage, this mechanism results in correct predictions for 87.3% of users in the largest connected component of the retweet network [31].

In the following sections we leverage these data to explore, in detail, how users from the political left and right utilize this important social media platform for political activity in different ways.

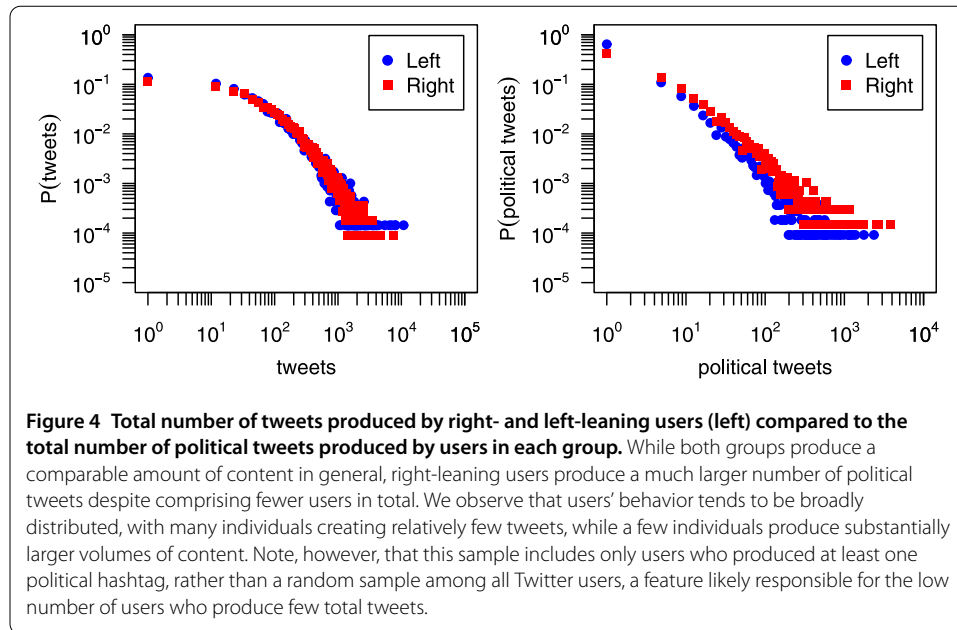
4 Behavior: individual-level political activity

Before examining structural differences in the social and communication networks of left- and right-leaning Twitter users, we first focus on political activity at the individual level. In this section we compare users in the left- and right-leaning communities in terms of their relative rates of content production, the amount of attention they allot to political communication, their respective rates of political self-identification, and their propensity for sharing information resources in the form of hyperlinks.

Right-leaning users are substantially more active and politically engaged with this social media platform. Specifically, our analysis shows that left-leaning users produce less total political content, allocate proportionally less time to creating political content, are less likely to reveal their political ideology in their profile biography, and are less likely to share resources in the form of hyperlinks. All of these findings stand in stark contrast to survey data and media reportage of the 2008 online political dynamics, and provide evidence in support of the notion that right-leaning voters are becoming more politically engaged online.

4.1 Political communication

From the perspective of leveraging social media for political organization, the baseline level of activity among a constituency is one of the most important characteristics of a population. Figure 4 shows that while left- and right-leaning users produce approximately the same number of tweets per user, right-leaning individuals actually produce 54% more total political content despite comprising fewer users altogether. This trend is the result of divergent priorities among left- and right-leaning users, as right-leaning users devote a substantially larger portion of their activity on Twitter to political communication. In fact, right-leaning users were almost twice as likely to create political content, with 22% of all tweets produced by right-leaning users containing one or more of the political hashtags under study, compared to only 12% for left-leaning users.



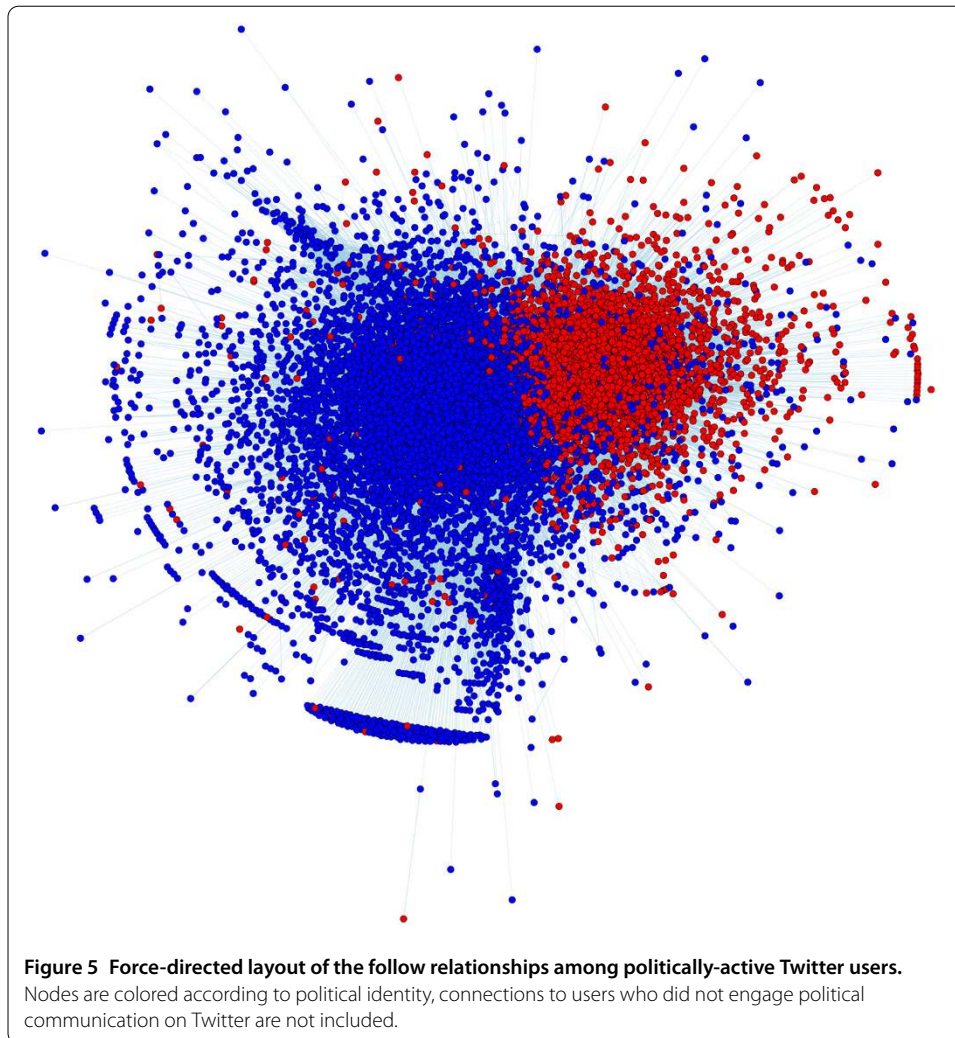
4.2 Partisan self-identification

In addition to devoting a larger proportion of tweets to political content, right-leaning users are much more likely to use their 140-character profile 'biography' to explicitly self-identify their political alignment. A survey of the biographies of 400 random users from the set of individuals selected for qualitative content analysis (Section 3.3) reveals that 38.7% of right-leaning users included reference to their political alignment in this valuable space, as compared with only 24.6% of users in the left-leaning community. Taken together, this analysis demonstrates that right-leaning users are much more likely to use Twitter as an outlet for political communication, and are substantially more inclined to view the Twitter platform as an explicitly political space.

4.3 Resource sharing

One of the key functions of the Twitter platform is to serve as a medium for sharing information in the form of hyperlinks to external content [22]. Given the constraints of the 140-character format, hyperlinking activity is especially important to the dissemination of detailed political information among members of a constituency.

With respect to this aspect of online political engagement, too, we see that right-leaning users are more active than those individuals in the left-leaning community. Among all tweets produced by users in the right-leaning community, 43.4% contained a hyperlink, compared with 36.5% of all tweets from left-leaning users. This trend is even more pronounced if we consider only resource sharing within the set of political tweets, with left-leaning users including a hyperlink in 50.8% of political tweets, as compared to right-leaning users, who include hyperlinks 62.5% of the time. From these observations we conclude that right-leaning users are more inclined to treat Twitter as a platform for aggregating and sharing links to web-based resources, an activity crucial to the efficient spread of political information on the Twitter platform.



5 Connectivity: global-level political activity

Next, we turn our attention to structural differences in social interaction and communication networks of left- and right-leaning users.

5.1 Follower network

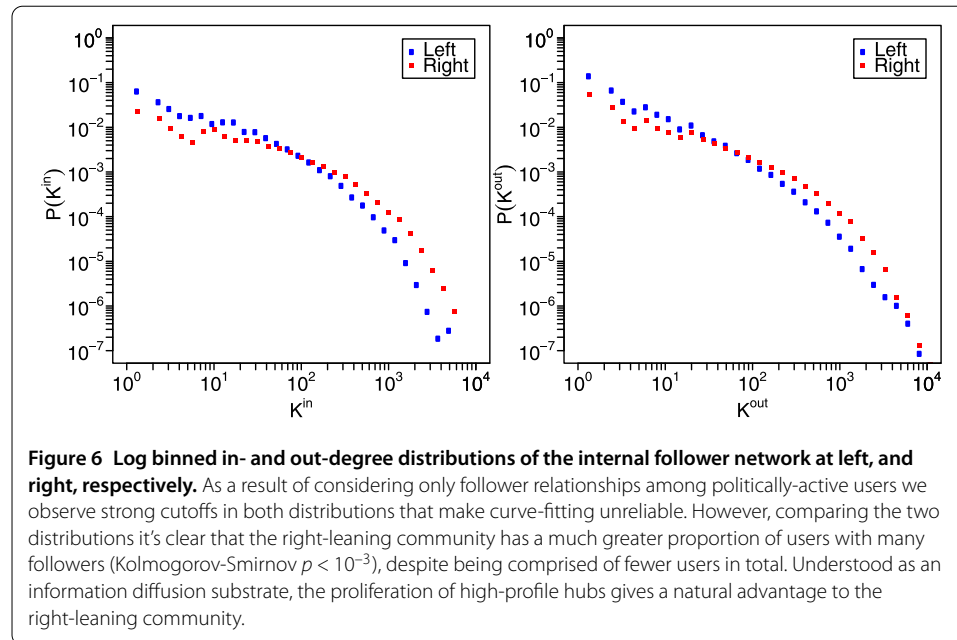
We begin with an analysis of the network defined by the follower/followee relationships shared among members of these two groups (Figure 5). Encoding the fact that a user subscribes to the content produced by another, the follower network is best understood as describing the social substrate over which information is likely to flow between political actors on Twitter. Specifically, though not all connections in the follower network encode equally meaningful social relationships [32], content is broadcast equally along all edges in this network.

We examine the differences in the follower subgraphs induced by considering only connections between users of the same political affiliation. For the purposes of this analysis, a directed edge is drawn from user A to user B if A is a follower of B . Basic statistics about these two subgraphs, including average degree, undirected clustering coefficient, and proportion of reciprocal links are presented in Table 4. We see that along all dimensions, users

Table 4 Follower network statistics for the subgraphs induced by the set of edges among users of the same political affiliation

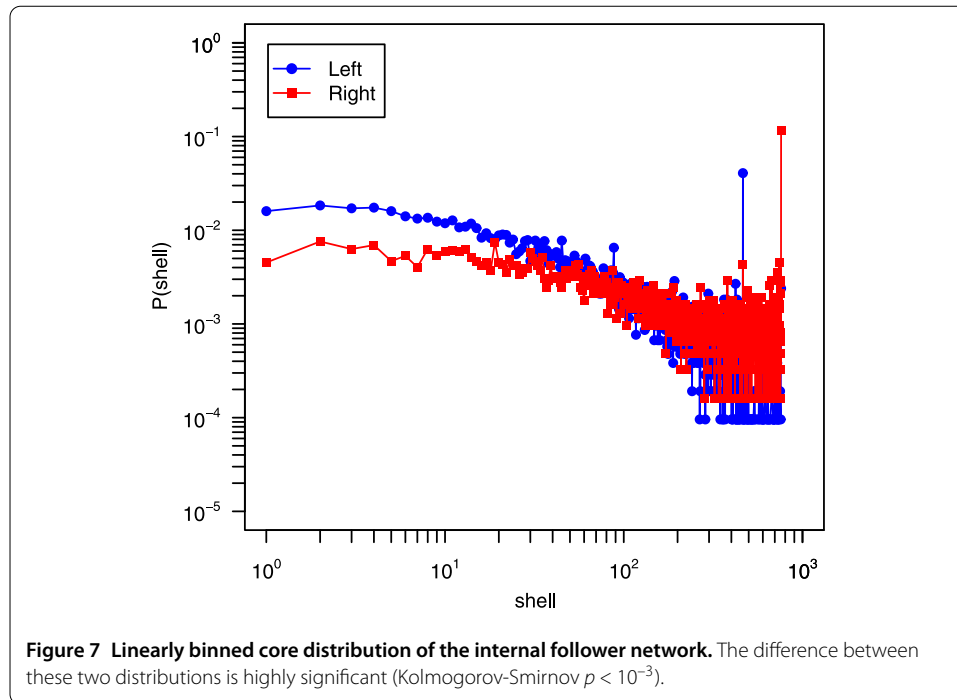
Community	Nodes	Edges	Average degree	Clustering coefficient	Reciprocity
Left	9,941	803,329	80.80	0.134	42.8%
Right	6,426	1,503,417	233.95	0.221	64.8%

Reciprocity is defined as $\frac{D_R}{D}$, where D_R is the number of dyads with an edge in each direction and D is the total number of dyads with at least one edge. Follower data was only available for a subset of the study population, owing to private or deleted accounts.



in the right-leaning community are much more tightly interconnected, with a substantially higher average clustering coefficient and greater average degree. Additionally, we observe a higher proportion of reciprocal links between right-leaning users, indicating the presence of stronger, mutually-affirmed interest among individuals in this community. All of these factors indicate that right-leaning users are more tightly interconnected, resulting in a basic structural advantage with respect to the challenge of efficiently spreading political information on the Twitter platform.

Using the Kolmogorov-Smirnov two-sample test to measure the degree of similarity between the in- and out-degree distribution for left- and right-leaning users we find a significant difference between the in-degree distributions of left- and right-leaning users, but only a marginal difference between the corresponding out-degree distributions (Figure 6). We interpret this to mean that a right-leaning user is more likely to have a large audience of followers who may potentially rebroadcast his or her call to action or piece of political information. For example, left-leaning users are roughly twice as likely as right-leaning users to have in-degree one, while users that are associated with the right are almost four times more likely to have in-degree 1,000 than users associated with the left. Additionally, users in the left-leaning community are more likely to be only peripherally connected into the network, as evidenced by the distribution of the k -core shell indices of users in each community (Figure 7). For a given network, the k -core is the maximal subgraph whose nodes (as members of the subgraph) have at least degree k , or, in other words, have at



least k neighbors in the k -core itself. The shell index, c , of a node refers to the coreness (k) of the highest-order k -core of which the node is a member [33].

These observations lead us to conclude that there are substantial structural differences in the fundamental patterns of social connectivity among politically left- and right-leaning Twitter users, a finding supported by the seminal work of Adamic and Glance [34] on the connectivity patterns of high-profile partisan bloggers. Specifically, the right-leaning community is much more densely interconnected, with more users tightly integrated into the right-leaning social network. In contrast, the network of follower/followee relations among left-leaning users exhibits a much more decentralized, loosely-interconnected structure, with far fewer mutually-affirmed social connections.

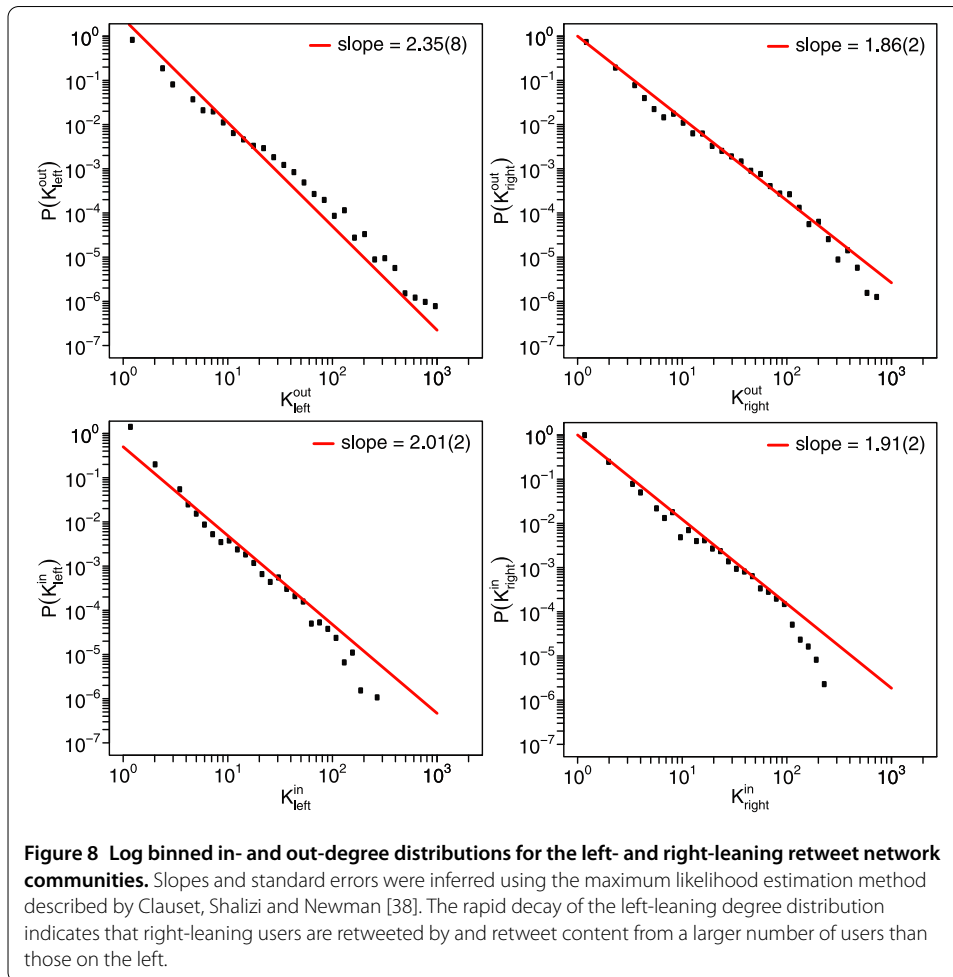
5.2 Retweet network

Next we consider the structure of the network of political retweets in order to understand how information actually spreads on the social substrate characterized in Section 5.1. While each link in the follower network represents a potential pathway along which information may flow, edges in the retweet network correspond to real information propagation events. Specifically, when user A rebroadcasts a tweet produced by user B , she explicitly signifies receipt of the content in question, and thus we draw an edge from user B to user A indicating the direction of information flow. Consequently, the structure of the retweet network reveals much about how information actually spreads within these two communities. Visualized previously in Figure 3, basic statistics describing the networks induced by retweets containing at least one political hashtag between users of the same partisan affiliation are shown in Table 5.

In practice, the tightly-interconnected structure of the retweet network confers communication advantages to the right-leaning community of users. Examining the in- and out-degree distributions for these two communities we find that though the power-law

Table 5 Retweet network statistics for the subgraphs induced by the set of edges among users of the same political affiliation

Community	Nodes	Edges	Average degree	Clustering coefficient	Reciprocity
Left	11,353	32,772	2.88	0.032	13.5%
Right	7,115	39,713	5.58	0.045	12.1%



exponents are similar, the difference between them is statistically significant at the 95% level (Figure 8). The faster decay in the degree distribution of the left-leaning community implies that right-leaning users are rebroadcast by and rebroadcast content from a larger number of individuals than users on the left. That right-leaning users pay attention to more information sources compared to left-leaning individuals is indicative of a higher degree of engagement with the Twitter platform itself. Similarly, an individual wishing to rapidly reach a wide audience has a natural advantage given the structure of the right-leaning retweet network.

With respect to the number of users in high-order k -cores, too, we see that the right-leaning community enjoys structural advantages, with a greater proportion of highly active users connected to other highly active users (Figure 9). This difference could lead to consequences in the spread of information through these networks. Work by Kitsak et al. indicates that it is individuals with high shell index, rather than those who are most cen-

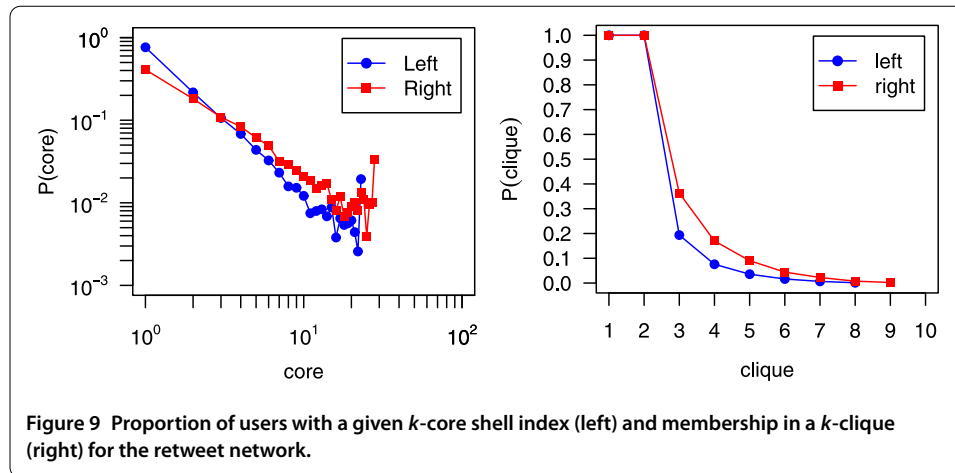


Figure 9 Proportion of users with a given k -core shell index (left) and membership in a k -clique (right) for the retweet network.

Table 6 Mention network statistics for the subgraphs induced by the set of edges among users of the same political affiliation

Community	Nodes	Edges	Average degree	Clustering coefficient	Reciprocity
Left	11,353	50,273	4.42	0.053	20.8%
Right	7,115	64,993	9.13	0.078	24.5%

tral or well connected, who are the most effective spreaders of information under a simple SIR-based information diffusion model [35]. Users on the right therefore, are more likely than those on the left to be wired into the political communication network in such a way that they are able to facilitate the broad and rapid dissemination of political information.

We also find that a substantially higher proportion of right-leaning user participate in fully-connected subgraphs of size k , known as k -cliques. This result is especially important in the context of the complex contagion hypothesis, which posits that repeated exposures to controversial behaviors are essential to the adoption of these behaviors. Work by Romero, Meeder and Kleinberg focused specifically on online social networks indicates that this effect is particularly pronounced for political discourse on Twitter [36]. With fewer users in high-order k -cores, individuals in the left-leaning community will be less likely to encounter multiple users discussing the same partisan talking points or calls to action, exactly the kind of contentious content whose propagation is most likely to benefit from repeated exposure.

5.3 Mention network

Mentions are most strongly associated with direct, conversational engagement when the target username appears at the beginning of a tweet, as opposed to appearing in the body text. Among the mentions in our sample, the overwhelming majority (94.5%) take this form, providing strong evidence that connectivity among and between users in these two groups represents actual political discourse rather than simply third-person references. In Table 6 we report descriptive statistics on the topology of the left- and right-leaning mention networks, where an edge from A to B is drawn between two users of the same political affiliation if A mentions B in a tweet containing at least one political hashtag. Though the two networks exhibit very similar degree distributions, one important distinction is the fact that a greater proportion of mention relationships in the right-leaning commu-

nity are reciprocal. Compared to the number of reciprocal mentions observed in degree-preserving reshufflings of the left- and right-leaning mention networks, the right-leaning community exhibits 7.5 times as many reciprocal mention interactions than is expected by chance alone, compared to a 5.6 times as many reciprocal links in the left-leaning community. Reciprocal interactions suggest the presence of more meaningful social connections, manifest in conversational dialogue, rather than, for example, unidirectional commentary on the content of another user's tweets [32]. Here too, we find that users on the political right are more engaged with one another on Twitter, indicating that they are likely to benefit from a richer dialogue and hence more opportunities for frame-making and consensus building with respect to political topics.

6 Political geography

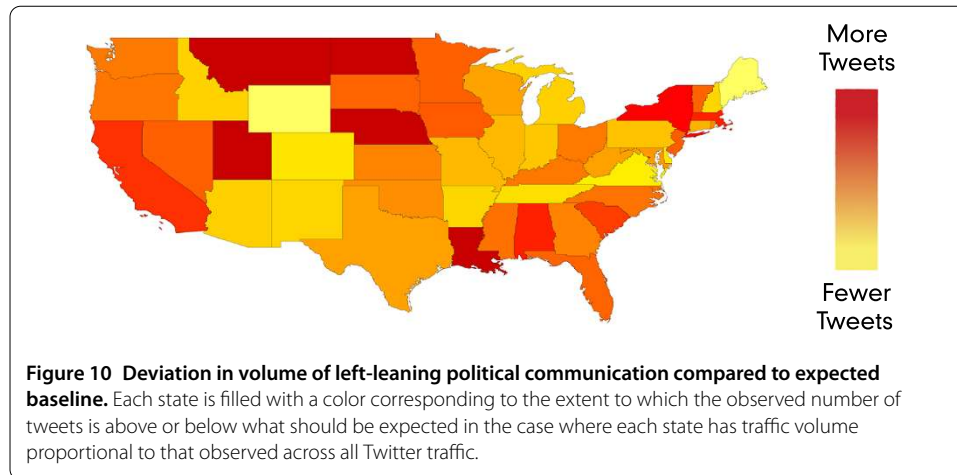
In addition to characterizing differences in behavior and connectivity, we can also examine the geographic distribution of individuals in these two communities. Here we present a cartogram in which the color of each state has been scaled to correspond to the degree to which, in that state, the observed number of tweets originating from the left-leaning community exceeds what we should expect by chance alone.

Because fewer than one percent of Twitter users provide precise geolocation data, we instead rely on the self-declared 'location' field of each user's profile to enable geographic analysis of data at the scale of this study. As a free-text field, users are able to enter in arbitrary data, and non-location responses such as 'the moon' do appear in the results. Complicating this analysis further, some users do not report any location data, though we do not report a partisan bias in terms of non-entries. Despite these caveats, a large number of users do report actual locations, and using the Yahoo Maps Web Service API,^b we are able to make a best-guess estimate about the state with which a user most strongly identifies.

Thus, for each state in which we observe N total tweets, and the relative proportion of tweets originating from left-leaning users (P_l), we can treat the arrival of partisan tweets as a Bernoulli process, and compute the number of tweets we should expect to see from left-leaning users as NP_l . Likewise, we can compute the extent to which the observed number of tweets associated with left-leaning users (T_l) is above or below the expected number, measured in terms of standard deviations, as $\frac{T_l - NP_l}{\sqrt{NP_l(1-P_l)}}$. Figure 10 uses color to encode these deviations for each state, with states in which the volume of activity far exceeds what should be expected by chance shown in deep red, and those in which the observed volume is far below what should be expected by chance shown in light yellow.

Initial inspection of this figure reveals that the geographic distribution of individuals from the left-leaning network community corresponds strongly to the traditional political geography of the United States. We see that left-leaning individuals feature prominently on the coasts and North East, and tend to be underrepresented in the midwest and plains states.

Looking more closely, however, we find that there are some places in which the partisan makeup of tweets is quite different from what might be hypothesized intuitively. For example, Utah, a traditionally conservative state which at the time of this writing had two Republican senators, exhibits a dramatically higher volume of left-leaning content than should be expected by chance alone. One possible explanation for this observation could be that individuals in some states with an ideologically homogeneous population turn to



social media as an outlet for political expression. While this is but one possible explanation among many, and a more rigorous analysis is required to support any definitive claim, this example illustrates the ways in which novel hypotheses can derive from data-driven analyses of political and sociological phenomena.

7 Conclusion

In this study we have described a series of techniques and analyses that indicate a shifting landscape with respect to partisan asymmetries in online political engagement. We find that, in contrast to what might be expected given the online political dynamics of the 2008 campaign, right-leaning Twitter users exhibit greater levels of political activity, tighter social bonds, and a communication network topology that facilitates the rapid and broad dissemination of political information.

In terms of individual behavior, politically right-leaning Twitter users not only produce more political content and devote a greater proportion of their time to political discourse, but are also more likely to view the Twitter platform as an explicitly political space and identify their political leanings in their profiles. With respect to social interactions, the right-leaning community exhibits a higher proportion of reciprocal social and mention relationships, are more likely to rebroadcast content from a large number of sources, and are more likely to be members of high-order retweet network k -cores and k -cliques. Such structural features are directly associated with the efficient spreading of information and adoption of political behavior. Taken together, these features are indicative of a highly-active, densely-interconnected constituency of right-leaning users using this important social media platform to further their political views.

This study is characteristic of an emerging mode of inquiry in the political and social sciences, whereby large-scale behavioral data are aggregated and analyzed to shed quantitative light on questions whose scale was previously considered outside the realm of tractable analysis [37]. Using structural features of a digital communication network one can make high-fidelity inferences about the political identities of thousands of individuals. Such data provide a deeper understanding of the changing landscape of American online political activity. Looking forward, techniques such as these are likely to become increasingly important as the political and social sciences rely in greater measure on large-scale digital trace data describing human opinion and behavior.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

MDC and BG collected the data and performed the analysis. MDC, BG, AF and FM conceived the experiments and wrote the manuscript.

Author details

¹Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA. ²College of Computer and Information Sciences, Northeastern University, Boston, MA 02115, USA.

Endnotes

^a http://dev.twitter.com/pages/streaming_api.

^b <http://developer.yahoo.com/maps/rest/V1/geocode.html>.

Received: 20 January 2012 Accepted: 4 May 2012 Published: 18 June 2012

References

1. Pew Internet and American Life Project (2010) The Internet and Campaign 2010. Technical report, Pew Research Center
2. Bennett L (2003) New media power: The Internet and global activism. In: Couldry N, Curran J (eds) Contesting media power: Alternative media in a networked world. Rowman and Littlefield, Totowa, pp 17-37
3. Aday S, Farrel H, Lynch M, Sides J, Kelly J, Zuckerman E (2010) Blogs and bullets: new media in contentious politics. Technical report, U.S. Institute of Peace
4. Farrell H, Drezner D (2008) The power and politics of blogs. *Public Choice* 134:15-30
5. Kaplan A, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of Social Media. *Bus Horiz* 53:59-68
6. Benkler Y (2006) The wealth of networks: how social production transforms markets and freedom. Yale University Press, New Haven
7. Sunstein CR (2007) Republic.com 2.0. Princeton University Press, Princeton
8. Tolbert C, McNeal R (2003) Unraveling the effects of the Internet on political participation? *Polit Res Q* 56(2):175
9. Garrett R (2006) Protest in an information society: A review of literature on social movements and new ICTs. *Information, Communication & Society* 9(2):202-224
10. Land M (2009) Networked activism. *Harv Hum Rights J* 22:205
11. Pew Internet and American Life Project (2010) Social media and politics in 2010 campaign. Technical report, Pew Research Center
12. Pew Internet and American Life Project (2010) The Internet and civic engagement. Technical report, Pew Research Center
13. Lutz M (2009) The social pulpit: Barack Obama's social media toolkit. Technical report, Edelman
14. Carr D (2008) How Obama tapped into social networks' power. *The New York Times*
15. Creamer M (2008) Obama wins! ... Ad age's marketer of the year. *Advertising Age*
16. Holahan C (2008) John McCain is way behind online. *Bloomberg Businessweek*
17. Pew Internet and American Life Project (2008) Social networking and online videos take off: Internet's broader role in campaign 2008. Technical report, Pew Research Center. <http://people-press.org/reports/display.php3?ReportID=384>
18. 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, New York, pp 591-600
19. Howard P, Duffy A, Freelon D, Hussain M, Marai W, Mazaid M (2011) Opening closed regimes, what was the role of social media during the Arab Spring? Project on Information Technology Political Islam 1-30
20. Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Flammini A, Menczer F (2011) Political polarization on Twitter. In: Fifth international AAAI conference on weblogs and social media, p 89
21. Java A, Song X, Finin T, Tseng B (2007) Why we Twitter: understanding microblogging usage and communities. In: Proc. of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis
22. Boyd D, Golder S, Lotan G (2008) Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: Proc. Hawaii intl. conf. on systems sciences, pp 1-10
23. Honeycutt C, Herring SC (2008) Beyond microblogging: conversation and collaboration via Twitter. In: Proc. 42nd Hawaii intl conf. on system sciences
24. Landauer T, Foltz P, Laham D (1998) An introduction to latent semantic analysis. *Discourse Process* 25(2):259-284
25. Blei D, Ng A, Jordan M (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993-1022
26. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F (2011) Truthy: mapping the spread of astroturf in microblog streams. In: Proc. 20th intl. World Wide Web conf. (WWW)
27. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 74(3):036104
28. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 76(3):036106
29. Krippendorff K (ed) (2004) Content analysis: an introduction to its methodology. Sage, Thousand Oaks
30. Kolbe RH (1991) Content analysis research: an examination of applications with directives for improving research reliability and objectivity. *J Consum Res* 18(2):243-250
31. Conover M, Ratkiewicz J, Gonçalves B, Haff J, Flammini A, Menczer F (2011) Predicting the political alignment of Twitter users. In: Proceedings of 3rd IEEE conference on social computing (SocialCom)
32. Huberman B, Romero D, Wu F (2009) Social networks that matter: Twitter under the microscope. *First Monday* 14:8

33. Barrat A, Barthlemy M, Vespignani A (2008) *Dynamical processes on complex networks*. Cambridge University Press, Cambridge
34. Adamic L, Glance N (2005) The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In: Proc. 3rd intl. workshop on link discovery (LinkKDD), pp 36-43
35. Kitsak M, Gallos L, Havlin S, Liljeros F, Muchnik L, Stanley H, Makse H (2010) Identifying influential spreaders in complex networks. *Nat Phys* 6:888-893
36. Romero D, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th international conference on World Wide Web. ACM, New York, pp 695-704
37. Lazer D, Pentland A, Adamic L, Aral S, Barabasi A, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M et al (2009) Life in the network: the coming age of computational social science. *Science* 323(5915):721
38. Clauset A, Shalizi C, Newman M (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661

doi:10.1140/epjds6

Cite this article as: Conover et al.: Partisan asymmetries in online political activity. *EPJ Data Science* 2012 1:6.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
