

Characterizing Communication Networks Associated with Political Hashtags

Misa Maruyama,
Communication and Information
Sciences
University of Hawaii at Manoa
misattm@hawaii.edu

Daniel D. Suthers
Information and Computer
Sciences, Communication and
Information Sciences
University of Hawaii at Manoa
suthers@hawaii.edu

Scott P. Robertson
Information and Computer
Sciences, Communication and
Information Sciences
University of Hawaii at Manoa
scott.robertson@hawaii.edu

Abstract

Among the diverse forms of communication and information networks found in the Web 2.0 environment, “social” and “informational” communication networks have been characterized in terms of their network metrics. Although Twitter is partly based on relationships between actors, activity has been shown to reflect characteristics of information networks. This study examines activity in Twitter within spaces defined by hashtags on political topics. We gathered our own data on a hashtag associated with the 2012 Hawaii senatorial race and compared our results to those from other political hashtag networks, and to typical social and information networks as well as random graphs. Results show that hashtag-centered reply and retweet networks in this domain do not fall clearly into the social or informational categories. There appears to be a third kind of network associated with political debate. More generally, it may be productive to conceive of communication networks in terms of multidimensional characteristics rather than categories.

1 Introduction

Communication networks have been broadly classified in the literature as either “social” networks or “informational” networks [27]. Social networks are formed primarily through interpersonal ties between actors. Informational networks are formed by the movement of information between actors. However, results reported in this paper show that the social vs. informational dichotomy does not accurately reflect the hybrid nature of online topic networks, such as those found on the micro-blogging platform Twitter.

Research using the methods of social network analysis [37], or more generally network analysis [27], has identified certain network metrics characteristic of social networks: “small-world” average path lengths,

high local clustering, skewed degree distribution [29] and reciprocity [9, 10, 21, 25, 30]. Other types of networks typically display some but not all of these characteristics. Here we differentiate between “social” networks broadly understood as any associations among human and non-human actors [20], and more narrowly defined “social” networks of interpersonal communicators. In this study, we use the latter definition to describe interpersonal associations among humans [28].

Much debate has ensued over the nature of the micro-blogging site Twitter. Does it afford social networks that resemble communities [12], or news-like information networks [19]? Researchers have studied follower-followee networks on Twitter and top trending topics to understand whether Twitter is a social network or a news medium [19]. Much research has been conducted recently on Twitter “topic networks” in the political domain [6, 14, 15, 33, 35], which are connections bounded by use of a political keyword or hashtag. These networks are often ephemeral interaction spaces, in that most topics “trend” or receive interaction for a short while before fading, although a smaller percentage of topics persist on the order of weeks or a few months [19].

We studied retweet, reply and mention networks in political hashtag spaces to explore to what extent they show the characteristics of social or informational networks. Our data included interaction networks for four popular national hashtags and one hashtag used in a U.S. Senate race in Hawai‘i. Political topic networks are also of interest because they have the potential to constitute diverse conversations around a topic [6]. Ongoing analyses on the nature of the dialogue will be reported in a future publication.

We found that the political topic networks are not purely social or purely informational in their structure. They exhibit “small-world” average path lengths and alphas similar to the power-law degree distribution of real-world social networks, but they display low clustering and reciprocity. Our study suggests that the

question “Is it social or not?” does not capture the complexity of networks that serve multiple purposes. We conclude with recommendations for future network analysis of the Twitter hashtag space.

2 What is Twitter, anyway?

2.1 Background on Twitter

Twitter, which allows users to post “tweets” in 140 characters or less, is unlike other social network sites (SNSs) in that relationships need not be reciprocal. A user can follow any other user, even if the user does not follow back. Users can interact using a variety of syntactical features, including retweets, mentions or replies. A retweet rebroadcasts a tweet to the retweeter’s followers. A mention includes the @ symbol and a person’s screen name, creating a deictic marker that shows the tweet is referencing the other person. A reply is a type of mention that shows that a user is speaking to another person, not just speaking about them. Hashtags allow users to tag tweets with metadata to identify a topic or a particular audience [40]. A hashtag consists of the character “#” followed by a keyword. Twitter enables one to search for and track all tweets marked with a given hashtag. Hashtags can be embedded anywhere within a tweet and often are designed to play dual roles of talk and metadata. For example, a recent tweet¹ reads in part, “Every time #Obama tries to reach out to the #GOP, he gets his fingers ripped off.” Thus, hashtags make tweets into “searchable talk” with metadata seamlessly integrated to increase “findability” [40]. They allow users to index into a topical discussion beyond their follower-followee network, and essentially enable users to define virtual interaction spaces on the fly. Researchers have studied hashtags as potential spaces for discourse, information discovery and expression [6, 39, 40].

2.2 Social or information network?

Twitter researchers have struggled with whether to categorize the platform as an informational or social network. A structural analysis by Kwak et al. [19] suggested that the network was more like a “news medium.” Unlike other social networks, the follower-followee network had a non-power law follower distribution, a short effective diameter, and low reciprocity. Using a measure similar to PageRank, Weng et al. [38] analyzed a non-random sample of more than 6,700 Twitter users in Singapore and found high reciprocity (72 percent). Cha, Haddadi,

Benevenuto and Gummadi [5] found contradictory results. Their study of 2 billion follow relationships among 54 million Twitter users found low reciprocity (10 percent).

Other researchers have conducted interviews, surveys and content analyses of tweets to study why people tweet. Taken together, their studies suggest that people use Twitter for social and informational purposes. A content analysis of tweets found four user intentions on Twitter: daily chatter, conversations, sharing information and reporting news [17]. Naaman, Boase and Lai [26] analyzed tweets to develop a taxonomy of reasons for using Twitter, including: opinions/complaints, information sharing, self-promotion, statements and random thoughts, “me now,” question to followers; presence maintenance, anecdote (me), and anecdote (others). The most common of all tweet categories in their data was information sharing, with 22 percent of the tweet share. Based on interviews and a small survey, Ramage Dumais and Liebling [31] identified four reasons for following on Twitter: social (interacting with others), substance (an interesting topic), status (interest in personal updates) and style (i.e., humor or wit).

Some scholars have classified a follower-followee network as a community. Gruzd et al. [12] assessed the extent to which the follower-followee network of the second author of that paper (Wellman) was a “community” using several frameworks, including Anderson’s [1] “imagined community”, where people were unified in identity despite not knowing each other, Jones’ [18] notion of a “virtual settlement” and McMillan’s and Chavis’ [23] sense of community. The authors concluded the network was not only an “imagined community,” but also a real community where people interacted with each other, engaged in long-term membership, maintained a sense of belonging, exchanged support and shared an emotional connection [18, 23].

2.3 Twitter topic networks

An increasing number of researchers have analyzed political “topic networks” on Twitter, defined as a collection of connections pertaining to a particular topic during a given time [6, 14, 15, 33]. The topic network comprises interactions bounded by use of a hashtag or keyword. Data collected in this manner provides a snapshot of the potential (follower-followee) or actual (retweets and mention) communication networks regarding a particular topic.

By collecting tweets that revolve around a particular topic, researchers limit their scope to topic-based interaction. Only social ties that pertain to a particular interest are visible. Like any network

¹ @MWM4444, June 11, 2013, paraphrasing Paul Krugman

boundary, this limits analysis. However, research has shown that social and topical structures are interrelated [32]. Romero, Tan & Kleinberg [32] found that the use of hashtags can predict social relationships and that social structure can predict the popularity of hashtags.

Dialogue within political topical networks on Twitter can be fleeting. Shamma et al. [33] analyzed tweets that contained keywords related to the 2009 inauguration of President Barack Obama. The topical conversation quieted within an hour after the debate. In subsequent studies, Shamma, Kennedy and Churchill [34, 35] found that most topics on Twitter were “peaky,” rising and declining in popularity over several minutes, while “persistent” topics were much less salient and declined within an hour.

This suggests that topic networks on Twitter may provide an ephemeral space where social aggregates coalesce around a shared interest. However, the social aggregate may weaken or dissolve once the topic is no longer relevant. Without long-standing interaction and emotional connections, can the network be described as anything more than an information network?

To pursue this question, we first asked to what extent Twitter topic networks show the characteristics of social or informational networks. We analyzed topographical characteristics of networks of political hashtag interaction, looking for clues that the spaces may be social networks [11, 25, 29]. We also compared the hashtag interaction networks to random graphs and real-world social and information networks.

3 Method

3.1 Data

We collected 2,432 tweets related to a 2012 U.S. Senate election in Hawaii using a Java program designed to monitor selected public streams using via Twitter’s streaming API. All of the tweets included the hashtag #HISen. They were collected during a five-week period around the November 6, 2012, election (October 7 to November 15). We selected the race due to its prominence in local politics, as well as its visibility on the national stage. It was one of several contested Senate races during a year when control of the Senate was within reach for Democrats and Republicans. U.S. Senator Daniel Akaka, D-Hawaii, retired after 22 years in the Senate, and two well-known politicians were vying for his seat: U.S. Congresswoman Mazie Hirono (D), who served two terms as Hawaii’s lieutenant governor, and former two-term Governor Linda Lingle (R). Both candidates and their supporters used the #HISen hashtag leading up to the election. We selected the hashtag because we anticipated it would be an active site for discussion.

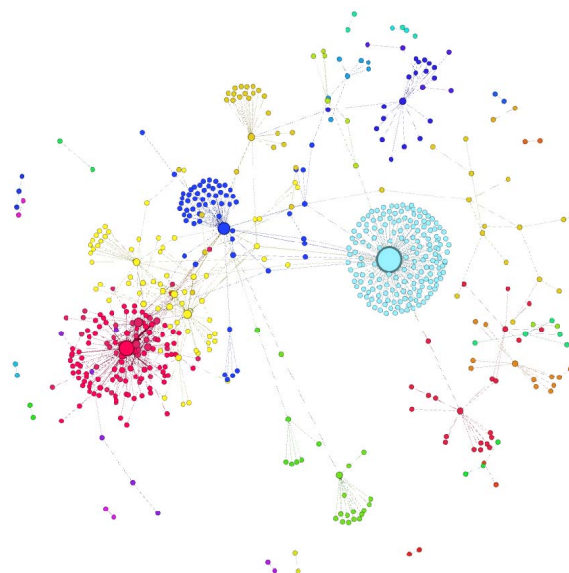


Figure 1: #HISen RT network visualized with a force-directed Yifan Hu layout algorithm

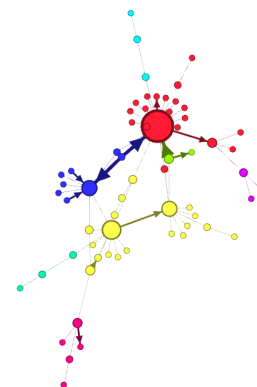


Figure 2: #HISen reply network visualized with a force-directed Yifan Hu layout algorithm

We created retweet and reply graphs from the data. We operationalized retweets as a tweet that included the phrase “RT @” to include those posted using Twitter’s automatic “retweet” button and those that were manually typed. Likewise, replies were all tweets that began with “@username.” We removed misuses of the @reply convention, in which users were speaking about about rather than speaking to someone at the beginning of a tweet. For example, “@lingle2012 said Obamacare would take \$716 billion away from Medicare ...” was removed from the reply corpus because it appeared to be a mention that happened to occur at the beginning of a tweet rather than a direct address to @lingle2012.

Table 1. Sociometrics for Twitter political networks studied

	#HISen				#p2		#tcot		#obama		#gop	
	RT	RT G(n,p)	Reply	Reply G(n,p)	RT	Men	RT	Men	RT	Men	RT	Men
nodes	620	620	90	90	4352	683	4094	837	3275	487	4183	1243
edges	1163	1661	100	173	5042	563	4870	663	3364	379	4801	1065
avg. deg.	2.613	5.358	1.911	3.844	1.062	0.119	1.072	0.146	0.936	0.106	0.975	0.216
reciprocity	0.038	0.002	0.024	0.006	0.001	0.002	0.001	0.002	0.001	0.000	0.001	0.001
network clustering	0.013	0.004	0.012	0.035	0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.000
avg. path length	4.372	6.299	2.954	5.593	2.701	1.459	3.027	1.195	1.281	1.073	3.068	1.264
diameter	10	16	8	14	9	4	10	3	6	3	13	4
α (in)	1.783	1.729	2.869	1.951	1.652	2.437	1.727	0.479	2.708	2.708	1.741	2.367
α (out)	2.974	1.718	2.377	1.906	3.079	2.848	2.785	2.914	3.735	3.076	2.933	2.905

In addition to studying the #HISen data corpus, we also analyzed four of the most popular political hashtags in the United States during the 2012 election. The data were made available by the Truthy project at Indiana University [22]. In May 2013, we downloaded data sets for four political hashtags: #p2 (Progressives 2.0), #tcot (Top Conservatives on Twitter), #obama (President of the United States) and #gop (Grand Old Party). For comparison purposes, we transformed the multi-modal networks into separate retweet and mention graphs, creating a total of eight Truthy networks. We also reversed the direction of the edges in the Truthy retweet graphs to be consistent with our #HISen retweet graph, where A points to B if A retweets B. We chose this convention so that a node's in-degree would represent its influence as an authority in the retweet, reply and mention graphs.

The resulting networks are visualized in Figures 1 and 2, using Gephi [2]. Node size reflects combined in- and out-degree, and edge thickness represents number of retweet or reply events between the given two nodes. The color of the nodes represents their modularity class, or clusters in which the density of weighted edges inside partitions is higher than those expected in a random graph [3]. The largest light blue node in Figure 1 is "emilyslist," a political action committee that supports pro-choice female Democrats; the largest red node is Linda Lingle's 2012 campaign Twitter account, "lingle2012;" the largest dark blue node is Maize Hirono's 2012 campaign Twitter account, "maizeforhawaii." The largest red and dark blue nodes in Figure 2 are again "lingle2012" and "maizeforhawaii;" and the largest yellow node is "colin9," a citizen's account whose username has been changed for anonymity.

3.2 Data Analysis

We used both Gephi and Pajek [8] to calculate network metrics, including reciprocity, network clustering coefficient, average path length and degree distribution. The igraph package [7] in R was used to visualize and calculate the networks' degree distributions and calculate associated power law alphas.

We compared the hashtag networks to a directed "Bernoulli/Poisson" random graph generated in Pajek. (This choice is discussed later.) The Bernoulli or Poisson random graphs, known as $G(n,p)$, are constrained by the number nodes (n) and the probability of being connected to other vertices (p); we generated graphs to match these parameters from our data. The graph tends to have a Poisson degree distribution for large n [27]. The random graphs are a baseline against which we can compare the Twitter networks. For example, Newman et al. [27] write that social networks are marked by larger clustering coefficients than would be expected at random. For our analysis we focus on several dimensions that have been described as characteristic of social networks, including (1) high reciprocity [10, 30], (2) a larger clustering coefficient than expected at random [29], (3) a "small-world" effect [29], and a (4) skewed degree distribution [29].

4 Results

Of the 2,432 tweets in the #HISen data corpus, there were 1,163 retweets (47.8 percent) and 100 replies (4.1 percent). Table 1 shows that the Truthy retweet networks contained between 3,364 and 5,042

retweets, and the mention networks contained between 379 and 1,065 mentions. Retweets comprised between 82 to 90 percent the data for each hashtag. Replies comprised the remaining 10 to 18 percent. We calculated the metric associated with each characteristic of social networks, shown in Table 1.

4.1 Reciprocity

Social relationships, such as friendships, tend to be reciprocal. Reciprocity refers to responding to one action with a similar action, such as repaying a favor or responding to a greeting in kind. It has been commonly dubbed the “Golden Rule.” Reciprocity leads to mutual support [30] and social cohesion [10]. Prior Twitter studies have operationalized social ties as reciprocated and informational ties as unreciprocated [17, 32]. In network terms, reciprocity is the frequency of pairs of dyads that share edges in both directions. Our reciprocity analysis was based on the dyad method, which is the ratio of the number of pairs with a reciprocated tie relative to the number of pairs with any tie.

All of our hashtag networks exhibited low reciprocity, but the national networks displayed lower reciprocity than the local network (#HISen). Table 1 shows that the retweet and mention graphs for the four national hashtags (#p2, #tcot, #obama and #gop) had almost non-existent reciprocity, ranging from 0 to .2 percent. The #HISen retweet and reply networks had reciprocity of 3.8 and 2.4 percent, respectively. Reciprocity in the #HISen retweet graph was about 19 times greater than the corresponding $G(n,p)$ random graph (.2 percent). The #HISen reply graph displayed reciprocity that was about four times greater than its corresponding random graph. This suggests that while reciprocity in the #HISen networks was low, it was more than what would be expected at random. It is surprising to discover that retweet reciprocity was an order of magnitude greater than random because a retweet is generally conceived as a one-to-many rebroadcast rather than one-to-one communication, so reciprocity would not necessarily be expected.

We postulate that the difference in reciprocity between the Truthy and #HISen networks may be due to their nature. Smaller networks tend to be denser, which would lead to higher reciprocity in the #HISen networks. The national vs. local nature of the networks may also explain the difference. A study of topic networks related to four gubernatorial races found that, when controlling for network size, local clusters were denser than national clusters [13]. On the national stage, actors in the Truthy networks may be less likely to know each other on a personal level. For instance, national media outlets are frequently retweeted, but

they rarely retweet average users. On the local level, people are more likely to be acquainted online and offline. This is especially true within Hawaii’s geographically isolated population, where the center of both population and political activity is the city of Honolulu. Thus, one would expect more reciprocated interactions (e.g., returning the favor of a retweet) among these interpersonal relationships.

But why would #HISen retweets have higher reciprocity than replies? One possible explanation is users’ re-appropriation of retweets as a conversation tool. We found that 132 retweets (or 11.34 percent) had been manually modified to include text before “RT @.” This included comments such as “Agree,” “LOL!,” “Fact Check!,” and “So cool that you’re doing that!” In these cases, the retweet was re-appropriated for interpersonal communication, or one-to-one communication that was publicly viewable. When the modified retweets were removed from the graph, reciprocity for the remaining retweet network (only tweets that began with “RT @username”) decreased to .028, which is closer to the reciprocity in the reply graph (.024). This suggests that the modified retweets contributed to reciprocity in the network.

Kwak et al. [19] claimed that a 22.1% reciprocity rate among follower-followee relationships was evidence that Twitter was not a social network. While reciprocity in the #HISen networks was much lower, the follow relationships cannot be compared to retweets and mentions. Data on follower-followee networks are cumulative such that all ties that were formed and maintained are visible. By contrast, data collected on retweets or mentions provide a snapshot of communication within a given period. The cumulative nature of follower-followee networks may be the reason why studies on similar online relationships have yielded higher reciprocity [19, 25].

4.2 Clustering Coefficient

In many real-world networks, the likelihood that two people will be friends is much greater if they share a mutual friend. The social explanation is that introductions often happen through mutual friends. Closed triads are measured by the clustering coefficient. In the same way that reciprocity measures the frequency of loops of length two in a directed network, the clustering coefficient measures the frequency of loops of length three [27]. We calculate the relevant metric using Pajek’s “network clustering coefficient,” which is the network-level proportion of closed triads to total triads. A slightly different metric is the Watts-Strogatz clustering coefficient, which takes each node’s ego-centric ratio of closed to total triads then averages the individual coefficients. We use

the former metric because the latter biases the clustering coefficient toward low degree nodes [27].

The clustering coefficients for the #HISen and Truthy networks are shown in Table 1. The network clustering coefficients for the #HISen retweet and reply graphs were 1.3 percent and 1.2 percent. While the clustering coefficients were low, the retweet network's clustering was greater than what would be expected at random. The four national hashtags displayed very low clustering, with coefficients of 0 to .1 percent. This suggests that the Truthy networks may be more like information networks than social networks.

Newman [27] writes that clustering coefficients in real-world networks are often “of the order of tens of percent.” Table 2 compares the #HISen data to large directed social and information networks [27]. Social networks such as e-mail messages and e-mail address contacts have clustering coefficients of 16 and 13 percent, respectively, while information networks such as a thesaurus word network and the World Wide Web have clustering coefficients of 15 and 29 percent, respectively.

Interestingly, the clustering coefficient for the #HISen retweet graph was greater than would be expected at random, while the opposite was true for the #HISen reply graph. In other words, the #HISen retweet graph comes closer to our expectation for a social network than the reply graph. This is counter-intuitive given that retweets are often used as a one-to-many broadcasting tool and a reply resembles one-to-one communication. Again, the re-appropriation of retweets for conversational purposes may influence clustering just as it may have affected reciprocity. When we removed modified retweets, clustering decreased to .008, which suggests that modified retweets contributed to higher-than-random clustering in the retweet graph. We did not analyze the influence of modified retweets in the Truthy networks because

the text for each tweet was unavailable in our data set.

4.3 “Small-world” effect

Many social networks display a “small-world” effect, which refers to the observation that the geodesic (shortest path distance) between any two vertices in a real-world network tends to be small [29]. The effect is measured by average shortest path length. The mathematical explanation of short average path lengths is that the number of vertices within a particular distance from a given vertex grows exponentially with the distance, so the diameter is logarithmic in n [27]. Milgram [24] tested the mathematical argument in a famous experiment in which he asked recipients to get packages from Omaha, Nebraska, to a target in Boston, Massachusetts. On average, the length of completed routes was 5.9 hops [24], originating the popular catchphrase “6 degrees of separation.”

Average path lengths for the Twitter networks suggest a “small-world effect.” The mean geodesics for the #HISen retweet and reply networks were 4.37 and 2.95, respectively. Both networks displayed average path lengths less than what would be expected at random. The four national hashtags displayed similar average path lengths, ranging from 1.07 to 3.07. The average number of hops was slightly less than what Mislove et al. [25] found in their analysis of the Flickr (5.67), LiveJournal (5.88), Orkut (4.25) and YouTube (5.10). Table 2 shows that the mean geodesics in the Twitter networks were similar to the average path lengths for the social e-mail networks (4.95 and 5.22), but far less than the informational World Wide Web networks (11.27 and 16.18) (from 19).

Table 2. Sociometrics for typical social and information networks (from Newman, 2010, Table 8.1) compared to HISen

	#HISen		Social – directed networks		Information – directed networks			
	RT	Reply	E-mail messages	E-mail address books	WWW (nd.edu)	WWW (AltaVista)	Citation network	Roget's thesaurus
nodes	620	90	59812	16881	269504	203549046	783339	1022
edges	1163	100	86300	57029	1497135	1466000000	6716198	5103
avg. degree	2.613	1.911	1.44	3.38	5.55	7.2	8.57	4.99
Network cc	0.013	0.012	0.16	0.13	0.29			0.15
alpha(in)	1.8	2.9	1.5		2.1	2.1	3	
alpha(out)	3.0	2.4	2		2.4	2.7		
avg. path l	4.372	2.954	4.95	5.22	11.27	16.18		4.87

4.4 Skewed degree distribution

Newman [27] writes that many real-world networks follow a power law degree distribution, from citation networks to the Internet. While not all networks that follow a power law are social, many social networks follow a power-law degree distribution. Right-skewed degree distributions have been found in scientific collaboration networks, movie actor co-appearances and Fortune 1000 company directors who sit on the same boards [27].

One way to tell whether a real-world network follows a power law is to view a histogram of the cumulative degree distribution on a log-log plot to see whether it follows a straight line. Figure 3 shows the cumulative in-degree distribution for the #HISen retweet network, and Figure 4 shows its cumulative out-degree distribution. Both in-degree and out-degree appear to follow a right-skewed degree distribution. Similar results were found for the #HISen reply graph, although we interpret the graphs cautiously with a small n .

Distributions that vary as a power of degree are power laws, and the exponent of the power law is known as alpha, or α . Newman [27] writes that typical values are between 2 and 3.

The degree distributions for the Twitter graphs suggest they are like social networks. Table 1 shows the α for the Twitter networks. They ranged from 1.78 to 2.97 for the #HISen graphs. Most of the exponents for the national hashtags ranged from 1.7 (#p2 retweet in-degree) to 3.7 (#obama retweet out-degree), with the exception of an outlier (#tcot mention in-degree). Because n is fairly small, we hesitate to draw definitive conclusions but intend to conduct further research on power law distribution in the political topic network domain.

5 Discussion

Communication networks associated with political hash tags such as #HISen and those in the Truthy.com data display some features of and fail to display other features of both information and social networks. Like social networks, these political hash tag networks have “small world” average path lengths and power law distributions with alpha values similar to other social networks. But unlike social networks and more like information networks, they have low clustering coefficients and very low reciprocity. Thus, the social vs. information network dichotomy previously proposed in the Twitter literature [19] does not serve well to describe these networks. They appear to be a third kind of communication network, in which local

structure (clustering and reciprocity) is not important, but with global connectivity and differentiation into a few important actors. More generally, the sociometrics of these networks suggest that we think about communication networks as varying in multiple ways rather than falling into discrete categories.

The connectivity of the network has important implications for the way information travels and relationships form. A network that exhibits “small-world” path lengths allows for connectivity through the network. The path lengths are *potential* routes of information diffusion. The shorter the path, the more efficient the diffusion can be. Because the topic networks exhibit short path lengths, news, gossip or ideas can potentially travel quickly through the bounded hashtag space. Similar to short average path lengths, a power-law degree distribution can have special consequences for a network. Within a given exponent range [27], networks that follow a power-law can comprise a giant component, which is the largest group of connected vertices in a network. If Twitter users in the hashtag space are connected in a giant component, this can also aid the transmission of information and group socialization within the hashtag space.

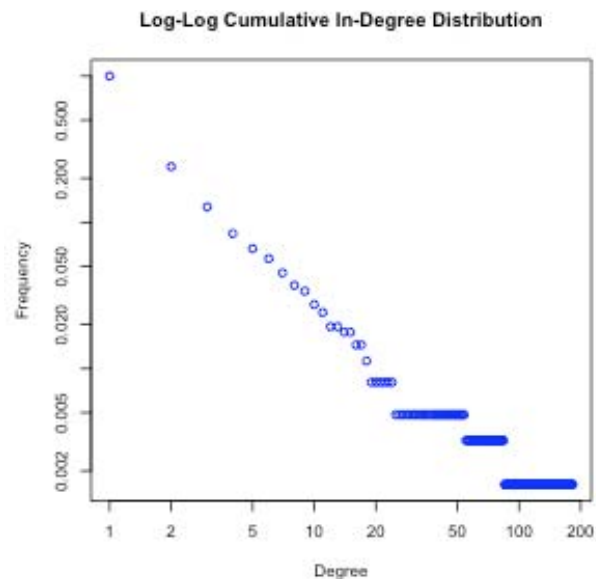


Figure 3. #HISen Retweet Cumulative In-Degree Distribution

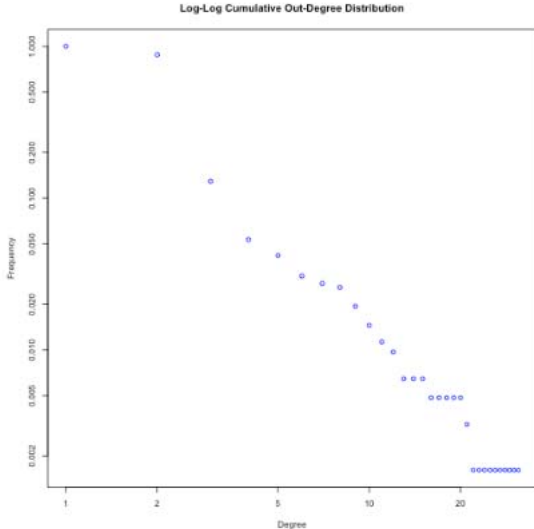


Figure 4. #HISen Retweet Cumulative Out-Degree Distribution

The lack of local structure in the topic network suggests the interactions may neither follow the “Golden rule” (reciprocity) nor manifest the saying, “a friend of my friend is also my friend” (clustering). Research has shown that people on both sides of the political aisle use hashtags that seem to be designed for one party, such as #p2 or #tcot [6]. It is not only a way to tag the topic of a message, but also to broadcast content to political opponents. Our review of the #HISen tweets reveals that supporters of both candidates used the hashtag, a topic that will be elaborated on in a future paper. When political opponents share a topic network, users may not be motivated to reply or may respond with vitriol [39]. Yardi and boyd [39] found that while Twitter affords cross-linking and exposure to diverse political opinions, it can also foster defensiveness and confrontation rather than civil cross-ideological debate. Their longitudinal analysis of reply pairs suggests that few were discussing the same topic a month later. They concluded that the conversations could be the result of news and current event-related interest [39], which may not lead to the reciprocity and transitivity found in long-lasting social relationships.

One limitation of the study is that we did not analyze the content of the tweets, which would provide richer insight whether the networks were informational or social. A preliminary review of #HISen retweets and replies suggests communication that is not only interpersonal, occurring between two or three people, but also relational. Relational communication is a type of interpersonal, socio-emotive communication that defines the relationship between people [4]. Coding for relational communication [36] could shed further light

on the extent to which political topic networks are social.

A technical consideration of the study is our use of the Poisson random graph model as a baseline against which to compare our Twitter networks. We chose to use the $G(n,p)$ random graph rather than other random graph models—such as small-world or scale-free models—because it modeled the Twitter networks’ lack of reciprocity and transitivity. This led to deeper insights about the hybrid nature of our Twitter networks.

In this study, we did not analyze follower-followee relationships, which may have found topographical differences. The follower-followee networks can be denser due to their cumulative nature, which may result in higher reciprocity and clustering. Researchers have expressed limitations of the follower-followee network in accurately reflecting real influence in terms of retweets or mentions [5, 16]. Still, we intend to analyze our topic networks in terms of follower-followee relationships, as well as to control for density when comparing national and local networks, to assess how the type of tie influences network structure.

Additionally, we acknowledge that not all Twitter users choose to use hashtags. By bounding our data in this way, we may bias our sample toward more elite users who are more knowledgeable of Twitter conventions. Future research may focus on topic networks bounded by keywords (“HISen”) rather than hashtags (“#HISen”).

6 Conclusion

Prior research has analyzed the topological characteristics of follower-followee networks to determine whether Twitter is “a social network or a news media” [19]. We analyzed local and national political hashtag spaces, where Twitter users discuss political topics of shared interest. We concluded that the Twitter networks exhibit characteristics of social networks in some ways but not in others. Social networks, as we have defined them, are formed through interpersonal ties among actors. The topic networks display global connectivity through “small-world” average path lengths and skewed degree distributions, but their local structure reveals low reciprocity and clustering.

Topic networks of national interest (i.e., #p2, #tcot, etc.) displayed lower reciprocity than local networks (#HISen), possibly due fewer interpersonal relationships sprawling a larger geographical area. Our results suggest that the networks we studied do not fit into a social vs. informational dichotomy. Even a particular type of network – a political topic network on Twitter – can display different social or

informational characteristics depending on the topic of interest and the actors involved. Instead of categorizing networks, researchers should consider how various features of a socio-technical system afford both information sharing and socialization.

7 Acknowledgements

This research was funded by NSF award Nos. ACI-0943147 and IIS-1064852. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8 References

- [1] Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. London: Verso.
- [2] Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *ICWSM*.
- [3] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, P10008. <http://dx.doi.org/10.1088/1742-5468/2008/10/P10008>.
- [4] Burgoon, J. K., & Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs*, 51(3), 193-214.
- [5] Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K. (2010). Measuring user influence in Twitter: The million follower fallacy. In *ICWSM*.
- [6] Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Political polarization on twitter. In *ICWSM*.
- [7] Csárdi, G., & Nepusz, T. (2012). igraph Reference Manual. Retrieved from <http://igraph.sourceforge.net/doc/igraph-docs.pdf>.
- [8] de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. New York: Cambridge University Press.
- [9] Golder, S.A., and Yardi, S. (2010). Structural predictors of tie formation in Twitter: transitivity and mutuality. In *SocialCom*, 88-95.
- [10] Gouldner, A. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25(2), 161-178.
- [11] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- [12] Gruz, A., Wellman, B., and Takhteyev, Y. (2011). Imagining Twitter as an Imagined Community. *American Behavioral Scientist*, 55(10), 1294-1318.
- [13] Himelboim, I., Hansen, D., & Bowser, A. (2012). Playing in the same Twitter network: political information seeking in the 2010 US gubernatorial elections. *Information, Communication & Society*, 1-24.
- [14] Himelboim, I., McCreery, S., Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, 18, 157-174.
- [15] Himelboim, I., Smith, M., and Shneiderman, B. (2013). Tweeting apart: Applying network analysis to detect selective exposure clusters in Twitter. *Communication methods and measures*, 7, 169-197.
- [16] Huberman, B., Romero, D., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. Available at SSRN 1313405.
- [17] Java, A., Song, X., Finin, T., & Tseng, B. L. (2007). Why we twitter: understanding microblogging usage and communities. In *WebKDD and SNA-KDD 2007 workshop on Web mining and social network analysis*, 56-65.
- [18] Jones, Q. (1997). Virtual communities, virtual settlements and cyber-archaeology. *Journal of Computer Mediated Communication*, 3(3), n.p.
- [19] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *WWW*, 591-600.
- [20] Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. New York: Oxford University Press.
- [21] Lin, Y.R., Sundaram, H., Chi, Y., Tatemura, J., and Tseng, B. (2006). Discovery of Blog Communities based on Mutual Awareness. In *WWE: Aggregation, Analysis and Dynamics*.
- [22] McKelvey, K. R. & Menczer, F. (2013). Truthy: enabling the study of online social networks. In *CSCW Companion*, 23-26.
- [23] McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology*, 14(1), 6-23.
- [24] Milgram, S. (1967). The Small-World Problem. *Psychology Today*, 1, 62-67.

- [25] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *IMC*, 29-42.
- [26] Naaman, M., Boase, J., and Lai, C. (2010). Is it really about me?: message content in social awareness streams. In *CSCW*, 189-192.
- [27] Newman, M.E. (2009). *Networks: an introduction*. Oxford University Press.
- [28] Newman, M.E., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review*, 68(3), 036122.
- [29] Newman, M.E., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. In *Proc. of the National Academy of Sciences of the United States of America*, 99(Suppl 1), 2566-2572.
- [30] Plickert, G., Cote, R.R., Wellman, B. (2007). It's not who you know; it's how you know them: Who exchanges what with whom?, *Social Networks*, 29, 405-429.
- [31] Ramage, D., Dumais, S., Liebling, D. (2010). Characterizing Microblogs with Topic Models. In *ICWSM*.
- [32] Romero, D. M., Tan, C., & Kleinberg, J. (2013). On the interplay between social and topical structure. In *ICWSM*.
- [33] Shamma, D.A., Kennedy, L., Churchhill, E.F. (2009). Tweet the debates: understanding community annotation of uncollected sources, In *SIGMM workshop on social media*, 3-10.
- [34] Shamma, D.A., Kennedy, L., Churchhill, E.F. (2010). Tweetgeist: Can the Twitter timeline reveal the structure of broadcast events? In *CSCW Horizons*.
- [35] Shamma, D.A., Kennedy, L., Churchhill, E.F. (2011). Peaks and Persistence: Modeling the Shape of Microblog Conversations, In *CSCW*, 355-358.
- [36] Walther, J. B., & Burgoon, J. K. (1992). Relational communication in computer-mediated interaction. *Human communication research*, 19(1), 50-88.
- [37] Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.
- [38] Weng, J., Lim, E.-P., Jiang, J., and He, Q. 2010. TwitterRank: Finding Topic-Sensitive Influential Twitterers. In *WSDM*, 261-270.
- [39] Yardi, S., and boyd, d. 2010. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology and Society*, 20, S1-S8.
- [40] Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, 13(5), 788-806. doi:10.1177/1461444810385097.