Southern Illinois University Carbondale
OpenSIUC

Working Papers | Political Networks Paper Archive

2011

Mapping the Political Twitterverse: Finding Connections Between Political Elites

Leticia Bode
University of Wisconsin - Madison, lbode@wisc.edu

Alexander Hanna
University of Wisconsin - Madison, ahanna@ssc.wisc.edu

Ben Sayre
University of Wisconsin - Madison, bgsayre@wisc.edu

JungHwan Yang
University of Wisconsin - Madison, jyang66@wisc.edu

Dhavan V. Shah
University of Wisconsin - Madison, dshah@wisc.edu

Mapping the Political Twitterverse: Finding Connections Between Political Elites*

Leticia Bode
Alexander Hanna, Ben Sayre, JungHwan Yang, Dhavan Shah
University of Wisconsin-Madison
lbode@wisc.edu

June 17, 2011

*This paper was prepared for the 2011 Political Networks Conference, Ann Arbor, MI.

Abstract

Twitter provides a new and important tool for political actors, and is increasingly being used as such. In the 2010 midterm elections, the vast majority of candidates for the U.S. House of Representatives and virtually all candidates for U.S. Senate and governorships used Twitter to reach out to potential supporters, direct them to particular pieces of information, request campaign contributions, and mobilize their political action. Despite the level of activity, we have little understanding of what the political Twitterverse looks like in terms of communication and discourse. This project seeks to remedy that lack of understanding by mapping candidates for federal office in 2010 and their followers, according to their use of the 4016 most used hashtags (keywords). Our data set is uniquely constructed from tweets of most of the candidates running for the U.S. House of Representatives in 2010, all the candidates for the Senate and governorships, and a random sample of their followers. From this we utilize multidimensional scaling to construct a visual map based on hashtag usage. We find that our data have both local and global interpretations that reflect not only political leaning but also strategies of communication. This study provides insight into innovation in new media usage in political behavior, as well as a snapshot of the political twitterverse in 2010.

1 Introduction

Twitter as a medium of communication has rapidly come of age since its creation in 2006. As usership of Twitter has grown, so too has its adoption in new arenas, including that occupied by politicians in the United States. As political officials adopt and use Twitter in larger numbers, it becomes important for us to understand how strategic politicians and political elites are making use of this new media platform.

This study takes a focused and unique sample from the political Twitterverse – candidates in the U.S. Congress in 2010 and their followers – to create a map of the space occupied by political elites on Twitter. Based on elements of Twitter speech (namely hashtags) we are able to gain a more nuanced understanding of what these universes of discourse look like, and how political users are connecting with one another in this new medium.

With the dramatic growth in popularity of Twitter, political actors have increasingly begun to use tweets as one of many campaign tools. The vast majority of candidates for the U.S. Congress in 2010, for instance, employed Twitter at least marginally in their campaign strategy. Political use of Twitter by candidates included efforts to reach out to potential supporters, direct them to particular pieces of information, request campaign contributions, and mobilize political action. As a result, a large amount of valuable information regarding political behavior is embedded in the political Twitterverse. While this is only one of many campaign tools and strategies, it is useful in that it gives us an easily measurable proxy for candidate outreach efforts, as well as offering an understanding of organic connections amongst political elites on Twitter.

1.1 Political Use of Twitter

The vast majority of research to date on the political use of Twitter has focused on members of Congress. Scholars have considered both what encourages members of Congress to adopt use of Twitter, and what helps them to be "successful" in such use. Lassen & Brown (2010) found members are more likely to adopt Twitter if their party leaders urge them to, if they are young, or if they serve in the Senate, whereas Gulati & Williams (2010) determined that party (Republicans adopt more) and campaign resources were the most important predictors of adoption. Chi & Yang (2010*a*) suggest that adoption is driven by a desire for constituency outreach, rather than a transparency motivation. Adoption may be accelerated by evidence of past users' success with the medium (Chi & Yang 2010*b*), and factors including vote share, funding, usage and influence may help to explain why some congressional users have more followers than their colleagues.

A single study to date has examined Twitter use within the electoral context, in an attempt to predict election outcomes. Tumasjan, Sprenger, Sandner & Welpe (2010) searched for mentions of political candidates and political parties in tweets. Simple word count analysis of this sample of explicitly political tweets revealed that the more frequently a candidate or party was mentioned, the more likely electoral victory for that entity.

Our study hopes to further this literature by incorporating a global understanding of the political Twitterverse and using elements from this understanding as a way to gain insight into political outcomes. This is by design an exploratory study, not aiming at explanation of any explicit outcome, but attempting to describe the shape of a communicative space with regard to a particular subject matter.

1.2 Multi-dimensional Issue Spaces

For decades, scholars have been concerned that issues are too often described merely in terms of "right" and "left." A single dimension of almost any issue space is likely to be flawed, as there are multiple conflicting factors in competition for any given issue. Whether to drill in the Alaskan wilderness, for instance, may pit environmental concerns against fiscal concerns, and also includes issues like security and federalism. A fiscal conservative who is socially or environmentally liberal may be truly conflicted on where she fits on a uni-dimensional spectrum of liberal to conservative. This is apparent empirically, in that some issues simply do not conform to a left-right spectrum (see Norton (1999) for a discussion of the impact of gender, and Anderson (2007) for a discussion of various issues, including agriculture), as well as theoretically, in modeling multi-dimensional issue voting (see for example Downs (1957), Calvert (1985)).

The problem with multi-dimensional issue spaces is the complexity of understanding them. This is the true benefit provided by our analysis of the political twitterverse. In our mapping, we are able to identify users who are more similar to each other by virtue of what they actually say – the elements of speech that they share. At the same time, we are able to identify which articles of political speech are more alike since they share the same users. Thus an understanding of connections between political elites emerges organically, by virtue of their own behavior, rather than by researchers imposing a known spectrum of understanding (most frequently political ideology) upon their actions. In this way, we are able to achieve greater understanding of how active users in the political twitterverse connect to one another and self-identify.

We have two main expectations of what the analysis of the map will produce. First, we expect major clusters to emerge representing the traditional ideological extremes of the left-right spectrum. Much like Adamic and Glance's famous work mapping the political blogosphere (2005), we expect to find a sharply divided Twitterverse along party lines. Those who are on the left will tend to say similar things and use similar hashtags, clustering together and apart from those on the right who will similarly connect with co-hashtagging behavior. Second, we expect to pick up not only a global division based on partisanship, but also to identify local clusters of users who engage in types of political behavior within or between the classic understandings of ideological right and left. Sub-partisan clusters may represent different understanding of ideology, or strategic attempts at communication by political elites. As Twitter is a nascent communication medium, people likely attempt to exploit it for their own purposes. Applying this perspective to the political space, strategy may be to disseminate information as widely and effectively as possible. Therefore, we expect to detect in our data local clusters of users who engage in similar diffusion-maximizing behavior.

2 Data

Data for this project were gathered in two waves. The initial wave began on Labor Day 2010 and was based on a list of 404 candidates in 103 races for seats in the House of Representatives. At the time when this sample was started, the total number of Twitter accounts we could follow was too low for us to be able to follow all candidates for House races along with samples from their follower lists, so we had to select a subset of races to focus on. The strategy employed was to include all candidates in races that were either tossups or leaning to one side or the other (as judged by the New York Times in the last week of August), along with a handful of noncompetitive races chosen at random. 16 of these races had not held their primaries by Labor Day, and for these races all candidates running in the primary were included in the sample. Of a total of 404 candidates, 253 were from one of the two major parties and the rest were independents or third-party candidates. Out of this list of candidates, 233 were found to have Twitter accounts, with 201 of those being major-party candidates.

A random sample of followers was taken for each candidate such that it proportionally decreased as the sample approached the maximum sample size of 50. The size of the sample per candidate was calculated by

$$n_c = \frac{50}{1 + (50 - 1)/F_c} \tag{1}$$

in which F_c is the total number of followers for candidate C at the beginning of the measurement period.

The second wave of data collection was started at the beginning of October and included all gubernatorial candidates and all US Senate candidates as well as a replication of the first wave which resampled the House candidates using the same sampling formula. Among the races for governor, only three included a third-party or independent candidate in the race, and only the two candidates in the Nebraska race had not identifiable Twitter account. The Senate races were very similar, with only occasional races with third-party candidates and few candidates without identifiable Twitter accounts.

A new random sample of Twitter followers of all candidates (N = 409) was added to the existing sample of users being followed, resulting in a total sample of 23,466 followers. Collection of tweets continued for one month after the election on November 2, 2010.

Over this time period (88 days total), nearly 9 million tweets were gathered, either directly tweeted by users in the sample or distributed (retweeted) by those users. The data were collected by using Twitter's Streaming API.

A feature of the collection with the Streaming API is that the data structure returned by the API has a number of different elements included along with the actual tweet. This includes information such as all public user information and geolocation, and most relevant for the current project, what the API labels as "entities", parts of the text which can be identified as either a URL (e.g. http://www.website.com), a hashtag (e.g. `#politics`), or user mention (e.g. `@johndoe`). The current project takes advantage of the distinct enumeration of these "entities" which allow us to parse important information from the data structure itself. For this particular study, we focus on the use of hashtags by users, allowing the data to inform which hashtags are important within this sample, and then further how groups of users coalesce based on similar use of hashtags.

3 Methodology

In order to identify our variable of interest – the clustering of Twitter users together – we performed a multidimensional scaling (MDS) analysis (Kruskal & Wish 1981) for use of hashtags by every unique user in our dataset. First we constructed a two-mode matrix of users by hashtags used. Entries in the matrix were the number of times the user used the hashtag U_{ei} normalized by the user's total usage of hashtags $U_{e\bullet}$, then weighted by the population's total usage of that hashtag $P_{\bullet i}$. Normalizing hashtag use by user helps to distinguish a user who tweets once and uses a particular hashtag from a user who tweets a thousand times, hashtagging each time, but using that same hashtag only once. We believe this is an important distinction when attempting to classify between types of political users of Twitter. We further weight by the population's use of a particular hashtag so that not all hashtags are assumed equal. Because some hashtags are used with much greater frequency than others, it is important to give those hashtags greater importance in our classification scheme as well. Equation 2 expresses this in mathematical notation.

$$M = [\frac{U_{ei}}{U_{e\bullet}} P_{\bullet i}] \tag{2}$$

The MDS analysis was performed using non-metric MDS. Non-metric MDS attempts to retain rank order of entries as ordered by distance while at the same time attempting to minimize the badness-of-fit (stress) iteratively (Kruskal & Wish 1981). We used the Kruskal's Non-metric Multidimensional Scaling function included in the R MASS package (Venables & Ripley 2002). Input to the MDS was a dissimilarity matrix calculated from Euclidean distance between rows of matrix M in equation 2. We allowed for two dimensions in order to most easily interpret the results graphically and substantively (two dimensions map nicely on X and Y axes). Using additional dimensions did not dramatically decrease the stress.

We gain two main insights from this analysis. First we can see how actors cluster together in a two-dimensional space by virtue of what they say. This means we can discern distinct groupings of individuals based on their shared Twitter behavior. Second, we distinguish which entities are substantively closer to each other. Presumably the concordance of entities like hashtags uttered similarly by multiple users suggest some other shared unobservable or latent variable. That is, we imagine that the shared use of language on Twitter is a proxy for other similarities amongst users within clusters. In the context of our study, the most likely latent variable is political sentiment. The clearest understanding of political sentiment is represented by the left-right spectrum of political ideology. Thus we expect clusters to reflect a clear left-right division, with the potential for one or many middle categories as well. Empirically, we expect to see a clustering of hashtags such as `#tcot` (top conservatives on Twitter) and `#teaparty`, and on the other end `#tlot` (Top Liberals on Twitter) and `#p2` (Progressives 2.0).

The local interpretation of the analysis relies on the ability to observe clustering in the MDS output. This lends itself to more nuanced interpretations that do not accord to what may be considered only political leaning. We can also pick up on the variations in the types of political behavior in which users engage. Visually, we can discern clusters and attempt to assign meaning to them based on our knowledge of the cases. We also can identify clusters systematically using hierarchical cluster analysis (HCA). We used the output of the MDS analysis to generate a fitted distance matrix based on Euclidean distance between rows and used this as the input to HCA using Ward's method. We use this method to minimize the loss of information we get from the clustering process. This results in compact, spherical clusters of actors.

4 Making and Interpreting the Map

To generate the MDS, we used the 4979 users who used the most hashtags and the 4016 top hashtags. We found that using more users and hashtags did not change the analysis dramatically. Because the same essential dynamics underly Twitter use as do those behind political blogs, we expect similar polarization in the political Twitterverse (based on the left/right or Democrat/Republican dichotomy) to that of the political blogosphere (Adamic & Glance 2005). We can achieve an understanding of this potential polarization both visually and computationally.

First, we need to be able to understand the map in a meaningful way and say more about its variance. There are two ways we can interpret the mapping, the global and local interpretation. The global interpretation lends itself to interpretation upon the axes. To assess the significance of any global interpretation, we created a variable based on the number of candidates the user followed, separated into five categories: *Democrat*, *Republican*, *Independent*, *Third Party Left*, and *Third Party Right*. The first three categories are self-explanatory, while last two were generated by categorizing various third-party groups according to their political leaning (i.e. Green for 3rd Party Left, and Libertarian and Tea Party for 3rd Party Right). We regressed the coordinates from the MDS using generalized linear modeling (GLM) on this variable. Essentially, we are interested in seeing how the slopes of the various lines generated by this regression vary - the further the distance between lines, the more distinct the follow patterns of the users. In addition, we created separate variables from the two most popular hashtags in the political Twitterverse: `#tcot` and `#p2`. This allows us to see how hashtagging behavior falls in terms of the political divide generated by candidate following behavior, and allows us to answer a number of interesting questions. Do users employing conservative hashtags follow conservative candidates? Do third party followers use traditionally party-specific hashtags? Again, we used GLM to assess the direction to which elements in the map lean.

Figure 1 displays the results for the the map based upon hashtags. We see the `#p2` and `#tcot` curves approaching the point of being orthogonal to each other, which is expected given the way that the map is constructed (remember we imposed two dimensions upon the graph, so it is not surprising that a conservative hashtag occupies one dimension while a liberal hashtag occupies another). It seems as though the `#p2` line accords with the users on the left of the map (convenient since it is the hashtag that is supposed to represent the left of the political spectrum) and the `#tcot` with those on the bottom and the right of the map. Again, this is not terribly surprising. We also see that the lines of the Democrats and the Third Party Left are nearly identical, and that they are the closest to `#p2`.

We would expect that the lines for the Republicans and Third Party

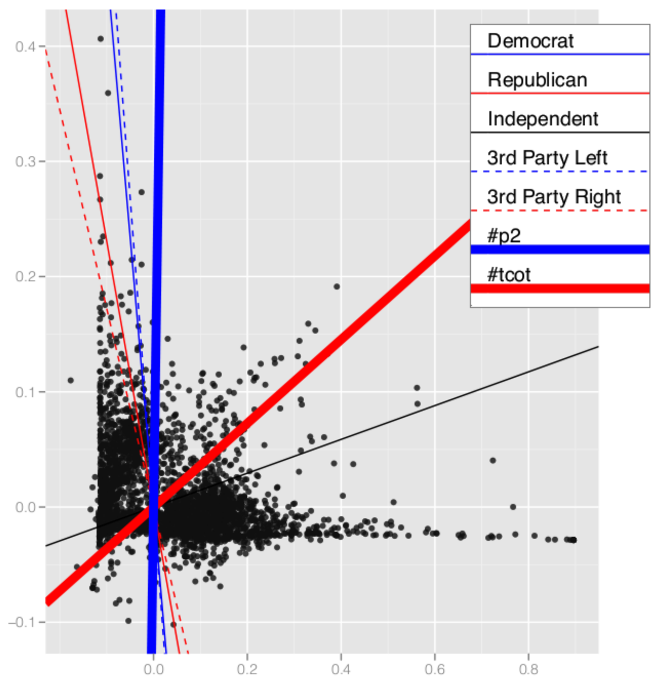

Figure 1: Multidimensional scaling with trend lines. Stress: 2.55

Right approach the `#tcot` line. Oddly enough, however, their slopes actually have the *opposite sign* of their respective hashtag. This could be an artifact of the estimation procedure, or could reflect unexpected behavior amongst these Twitter users. Our cluster analysis will speak to this possibility to some extent, but future research should also consider more explicitly the relationship between party and hashtag use.

Interestingly, the line for Independents seems to be far removed from all the other plots. This may reflect a number of possibilities. First, there were not a large number of independent candidates in our sample. Thus this particular slope suffers from a relatively small N, and as a result we are less sure of its slope than we are for some other lines (particularly those for the major two parties). Additionally, it is quite possible that independents choose not to engage in use of the two major hashtags we plot, as they are commonly associated with the two ends of the ideological spectrum, whereas independents by definition choose a third path. As such their slope in relation to the two lines reflecting hashtag use holds little true meaning.

The global analysis is revealing in two ways. It confirms that there is polarization in the political Twitterverse, in that conservatives and liberals tend to align separately, and also describes how disparate users are with regard to hashtag usage. Secondly, and perhaps more importantly to understanding political behavior, the global analysis indicates that a traditional partisan dichotomy is not sufficient to explain political behavior (at least in terms of hashtag use) on Twitter. There seem to be other important mechanisms at

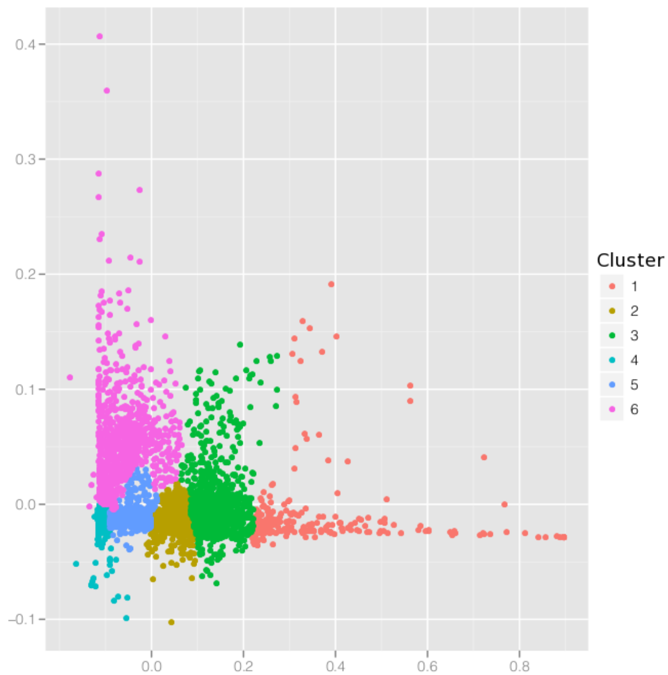

Figure 2: MDS with clusters

work with regard to this behavior. Thus we turn to cluster analysis to reveal other potentially non-partisan breakdowns among political tweeters.

The results of the hierarchical cluster analysis are reported in Figure 2. We chose to separate the map into six clusters, although using five or seven clusters would not have changed the analysis dramatically. Figure 2 demonstrates these distinct clusters, which change consistently with the X axis, but change much less on the Y dimension. Again, these clusters are based on hashtagging behavior on Twitter, suggesting that different types of such behavior occur within each of the six clusters.

In order to better understand the details of this behavior, we conducted

Cluster	1	2	3	4	5	6
N	223	708	1011	1603	616	817
Dem	-0.32	-0.25	-0.30	0.08	-0.05	0.06
Rep	0.01	0.04	0.03	-0.06	0.03	-0.11
Indep.	0.16 (ns)*	-0.29	-0.05	0.25 (ns)	0.42	0.01 (ns)
3rd R†	0.30	0.05 (ns)	0.05 (ns)	0.14 (ns)	0.20 (ns)	-0.12 (ns)
tcot	0.01	0.01	0.01	-0.10	-0.01	-0.06
p2	-0.01	-0.01	0.01	-0.15	-0.01	0.07

(ns) denotes non-significance at the $p \leq 0.05$ level
Logistic analysis performed for inclusion in each cluster individually. † All coefficients for third party left were insignificant and so are not shown.

Table 1: Local interpretation of clusters

a series of regression analyses, to determine what behaviors predicted a user falling in any given cluster. Because there are six clusters, six separate regressions were estimated.

The regression analysis is shown in Table 1. Clusters 1 and 6 are the most straightforward to interpret. Cluster 1 is positively related to all conservative concepts – following Republicans or members of a right-leaning third party, and using the hashtag `#tcot` – and negatively correlated with each liberal concepts – following Democratic politicians and using the hashtag `#p2` (note that the left-leaning third party following is not shown, though it was included in each model, as no coefficients representing this behavior were significant). Cluster 6 represents essentially the opposite set of behaviors – following Democrats and using `#p2`, but less likely to follow Republicans or use the hashtag `#tcot`. These clusters fit clearly within the left-right dichotomy into which American politics is most frequently divided. However, there are four other clusters of political Twitter users that fall somewhere outside of that clear spectrum. This is where a local interpretation is of most use to our understanding of the politics on Twitter.

We will address each remaining cluster in turn, and then offer some overarching observations regarding the clustering in its entirety. Clusters 2 and 3 are nearly identical – users in both groups are less likely to follow Democratic and Independent candidates, more likely to follow Republican candidates, and more likely to use the hashtag `#tcot`. The notable difference between the two clusters is their use or avoidance of the hashtag `#p2`, with cluster 2 avoiding use and cluster 3 positively associated with use. This suggests that users in cluster 2 resemble cluster 1 users (classic conservatives, avoiding liberal entities), whereas users in cluster 3 offer a new type of political behavior on Twitter. Cluster 3 users seem to employ *both* of the two top political hashtags, even though in terms of following behavior, they seem to lean right. This may suggest strategic hashtagging, in that Twitter users following a hashtag either on the left or the right (rather than another individual Twitter user) would still be likely to see a tweet, thus disseminating the information farther than it would otherwise travel, and to a different audience.

Similarly, clusters 4 and 5 somewhat resemble one another. Users in both clusters are less likely to use the top political hashtags from either ideological persuasion. This could be strategic, in that independent-minded users might choose to avoid classifying their tweets into one camp or the other. These users might also be employing other, more specialized political hashtags. Future work within this framework should further determine which hashtags are most common within each cluster, but particularly in those less likely to use #tcot and #p2. Alternatively, this could represent a cluster of users that is less sophisticated in Twitter use, and thus less likely to employ hashtags at all. The main difference between clusters 4 and 5 is their opposite affiliations in terms of following Democrats and Republicans for office. Cluster 4 users are more likely to follow Democrats and less likely to follow Republicans, whereas cluster 5 users are more likely to follow Republicans and less likely to follow Democrats. This may suggest a slight leaning toward or away from each of the main parties. Interestingly, cluster 5 users also have strong likelihood of following Independent candidates (the largest coefficient seen in any of our models). Thus, in terms of a classic understanding of political ideology, cluster 5 seems to best represent the true Independents.

How are we to characterize these clusters, then? The major insight provided by our local analysis is that we cannot assume all political activity on Twitter falls neatly into the left-right dichotomy to which political scientists are accustomed. Unlike the blogosphere, which has very little political middle ground (Adamic & Glance 2005), many of the users on Twitter have mixed following patterns and hashtagging behavior, suggesting greater nuance in the political behaviors and discussion occurring within Twitter. At the very least, this may serve as a call for greater research into the burgeoning political Twitterverse.

5 Conclusion

Through this project we have developed a method of creating a map of the political Twitterverse, using the built-in functionality of Twitter. We found that solely left/right distinctions, while useful in some ways, inadequately describe political behavior on the platform. Rather, we find it much more fruitful to discuss how users employ Twitter for political purposes in more nuanced ways, including how they interact with one another and how they self-affiliate and self-identify using the tools available to them within Twitter. Most notably, we think it fruitful to consider the strategic nature of political action and conversation on Twitter, particularly in terms of strategic hashtagging, such as encroaching on others' keywords. Moreover, the construction of this map may ultimately be useful in attempting to explain political outcomes such as elections, referenda, protests, and the like.

This analysis only chose to look at the use of hashtags in mapping the political Twitterverse. We hope that this method can be generally extrapolated to mapping any bounded space of discourses in the social media sphere, and attempting to explain outcomes in that space by virtue of elements of the map. By the same token, we could have used other entities used within the realm of Twitter, such as URLs and user mentions. Future research should consider whether similar rules apply in alternative areas of Twitter, or whether the more directed user mention is employed differently. Additionally, with computer-aided content analysis software we can create an entirely new set of attributes from which to categorize tweets, using political tweeters own language to inform us.

While the potential for research within this subfield of study is enormous, our project represents an important step in understanding the political interactions and connections happening every day on Twitter.

Works Cited

Adamic, Lada A. and Natalie Glance. 2005. "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog." WWW-2005 Workshop on the Weblogging Ecosystem, May 10-14, 2005, Chiba, Japan.

Anderson, Sarah E. 2007. "Are We Missing Something? Assessing Measures of Congressional Ideology." Unpublished manuscript.

Calvert, Randall L. 1985. "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence." *American Journal of Political Science* Vol 29:1. 69-95.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. 2010. "Measuring User Influence in Twitter: The Million Follower Fallacy." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Chi, Feng, and Nathan Yang. 2010a. "Twitter in Congress: Outreach vs. Transparency." Available at SSRN: http://ssrn.com/abstract=1630943.

Chi, Feng, and Nathan Yang. 2010b. "Twitter Adoption in Congress." *Review of Network Economics.*

Downs, Anthony. 1957. *An Economic Theory of Democracy.* Harper and Row.

Gulati, Girish J. and Christine B. Williams. 2010. "Communicating with Constituents in 140 Characters or Less: Twitter and the Diffusion of Technology Innovation in the United States Congress." Presented at the annual convention of the Midwest Political Science Association.

Kruskal, Joseph B. and Myron Wish. 1981. *Multidimensional Scaling.* Sage Publications.

Lassen, David S. and Adam R. Brown. 2010. "Twitter: The Electoral Connection?" *Social Science Computer Review.*

Lerman, Kristina and Rumi Ghosh. 2010. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

Norton, Noelle H. 1999. "Uncovering the Dimensionality of Gender Voting in Congress." *Legislative Studies Quarterly* Vol 24:1. 65-86.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Tumasjan, Andranik, Timm O. Sprenger, Phillipp G. Sandner, and Isabell M. Welpe. 2010. "Predicting Elections with Twitter: What

140 Characters Reveal about Political Sentiment." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Venables, W. N., and B.D. Ripley. 2002. *Modern Applied Statistics with S, Fourth Edition.* New York: Springer.

Yang, Jiang, and Scott Counts. 2010a. "Comparing Information Diffusion Structure in Weblogs and Microblogs." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Yang, Jiang, and Scott Counts. 2010b. "Predicting the Speed, Scale, and Range of Information Diffusion in Twitter." Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.