PLSC 508 REPLICATION

Mike Burnham
mlb6496@psu.edu
Github Repository: https://github.com/MLBurnham/networks_replication

May 7, 2020

ABSTRACT

Estimation of ideology via latent space modeling is a powerful tool for researchers. In 2015 Pablo Barbera published a method that accurately estimated mass ideology points using Twitter follower networks. In this research I re-examine his work by replicating and extending portions of his modeling and analysis. I examine a re-specification of his model that is bilinear in form and compare results among political elites. I determine that while the new model shows some promise, additional testing and validation is needed. Additionally, I extend his analysis of the echo chamber effect on Twitter by examining the structure of the retweet network he used in his paper. My findings largely agree with the original research, but provide more depth and context to the findings. To conclude, I discuss potential next steps for this research.

1 Introduction

Due to advances in data access, computing power, and statistical techniques, the measurement of latent ideological positions has made significant strides in recent years. This represents an exciting methodological advancement in political science as it enables researchers to control for and explain ideological dimensions in populations where this was previously not possible. In 2015, Pablo Barbera proposed a network based approach that has proven robust [1]. The method uses Twitter follower networks and applies a Bayesian spatial model to estimate the latent ideological dimensions. This allows researchers to estimate mass ideal points whereas previously research focused more on legislators or political elites.

In this paper I aim to accomplish three things: First, I iterate on Barberas method by respecifying his original model. Second, I expand on his analysis by applying additional descriptive techniques to his original retweet network. Finally, I propose a research agenda to further build upon this approach.

2 Data

Before proceeding to the analysis portion of this paper, a note about the data is warranted. For this research I used Barberas original replication data, which consists of two large networks pulled from Twitter during the 2012 presidential election. The first is a network of political elites and their followers consisting of 317 elites and roughly 300,000 followers. In the replication materials, this network is used to calculate the ideal point estimates of elites by taking a random sample of 10,000 users and subsetting the adjacency matrix to these random users and elites followed by over 200 of these users. In his published paper, Barbera estimates ideal points for all political elites and thus necessarily follows a different sampling procedure than in the replication materials. It is unclear if he added more followers to the adjacency matrix or lowered the threshold by which he filtered elites. While this would not be particularly difficult to triangulate, because the model is expensive to compute and diagnose I simply used the replication sampling procedure rather than trying to reverse engineer the publication sampling procedure. As a result, I work with a set of 165 elites rather than the full set of 317. When I make comparisons with the re-specified model and Barberas original model, I accordingly subset his data to these same elites as well for the sake of symmetry.

The second network is a retweet network Barbera uses to analyze the discourse on Twitter. To analyze this data set I use Barberas original mass ideal point estimates. The reason for this is that at the time of this writing I was able to successfully compute elite ideal point estimates under the re-specified model, but mass ideal point estimates failed to converge. Reasons and implications for this will be discussed further in the following sections. One additional downstream effect of this is that Barbera used the mass estimates estimates to normalize elite ideal point estimates. Because I do not have bilinear mass ideal point estimates I do not normalize either the bilinear elite estimates or Barberas original elite estimates when comparing the two. As a result, some of the numbers and sample sizes in this replication vary slightly from the original research.

3 Bilinear Fit and Validation

3.1 Specification

In this section I examine the impact of changing the functional form of Barberas model from one based on Euclidean distance to a bilinear model. This offers several potential advantages. First, bilinear models tend to separate dimensions more effectively. This may make for more robust estimates given the minutia and nuance to ideological point estimates. Second, bilinear models tend to be more computationally efficient. This is a particularly enticing proposition given that latent space modeling is often computationally intensive. The intrinsic weight of these models is further compounded by the massive scale of the networks they may be applied to. When applied to social media data such as Twitter, small samples of the data may consist of hundreds of thousands of nodes and millions of edges. At this scale many inferential techniques, including latent space models and exponential random graphs, become prohibitively expensive from a computational standpoint.

For his network, Barbera uses a sample of Twitter accounts belonging to political elites and their followers. A follower network is constructed from this sample such that y_{ij}= 1 if Twitter user *i* follows political elite *j* and y_{ij}= 0 otherwise. The decision to follow a particular elite is modeled as a logistic function of both the elite and users political ideology. Additional controls for the popularity of elite *j* and political interest of user *i* are also introduced. The decision of a given user to follow a given elite is thus modeled as such:

$$P(y_{ij} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \phi_j) = logit^{-}1(\alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2) \tag{1}$$

Where α_j is the popularity of a given elite, $\mathcal{B}_i$ is the political interest of a user, θ_i is the ideal point estimate of a given users, ϕ_j is the ideal point estimate of a given elite, and γ is a normalizing constant.

For this research, the most significant assumption in this model is the assumption that a user's decision to follow an elite is a function of the squared Euclidean distance of the ideological point estimates of user i and elite j: $-\gamma||\theta_i - \phi_j||^2$

I re-examine this model by instead assuming a bilinear relationship between ideological point estimates and a users decision to follow a political elite: $(\theta_i \times \phi_j)$. Additionally I remove the γ term from the equation because the term becomes unidentifiable in the bilinear model and has no clear substantive interpretation. The model I test is thus a slight variation on Barbera's:

$$P(y_{ij} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \phi_j) = logit^{-}1(\alpha_j + \beta_i - (\theta_i \times \phi_j)) \tag{2}$$

To test this I used Barberas original network data from the United States with an updated version of his original Stan [2] and R code. The data is a set of political actors with Twitter accounts circa 2012. Political elites are defined as: (1) political representatives in national-level institutions; (2) political parties with Twitter accounts; and (3) political media outlets and journalists. Only political elites with over five thousand followers were used. This results in a total data set of 318 political elites. From these political elites, a sample of 301,537 of their combined followers was taken. This sample was chosen from the entire population of their combined followers by only including those who (1) have sent over one hundred tweets; (2) have sent at least one tweet in the past six months; (3) have more than twenty-five followers; (4) are located inside the borders of the United States; and (5) follow at least three political Twitter accounts. The final subset of the sample used for the latent space model is then acquired by sampling as outlined in the data section to 165 elites and 10,000 followers.

The Euclidean model converged with two chains after 500 iterations. The bi-linear model ran for 3000 iterations and 5 chains. A more detailed list of hyperparameters and convergence plots is found in the appendix.

3.2 Model Fit

To compare the fit of the latent space models I replicated the authors original approach to validation. Figures 1 and 2 plot the ideal point estimates of 106 members of the 112th congress against their DW-NOMINATE scores. Figure 1 shows the bilinear estimates while figure 2 shows the Euclidean estimates. At face, the estimates appear largely the

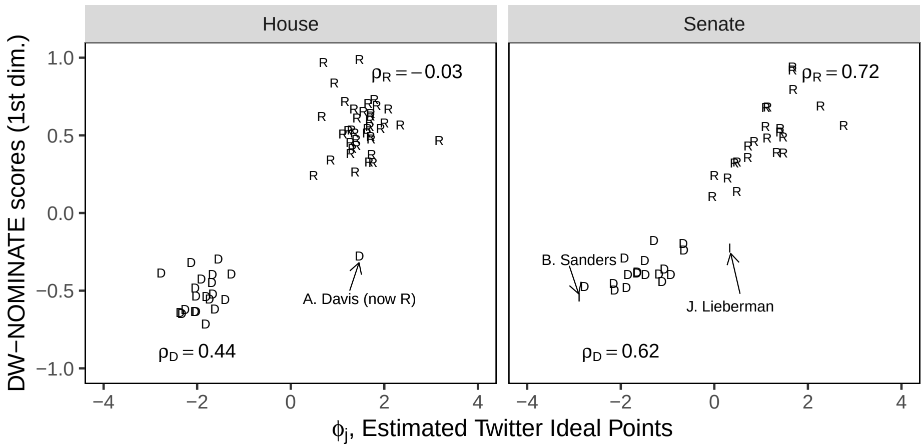

Figure 1: Bilinear ideal point estimates for members of congress

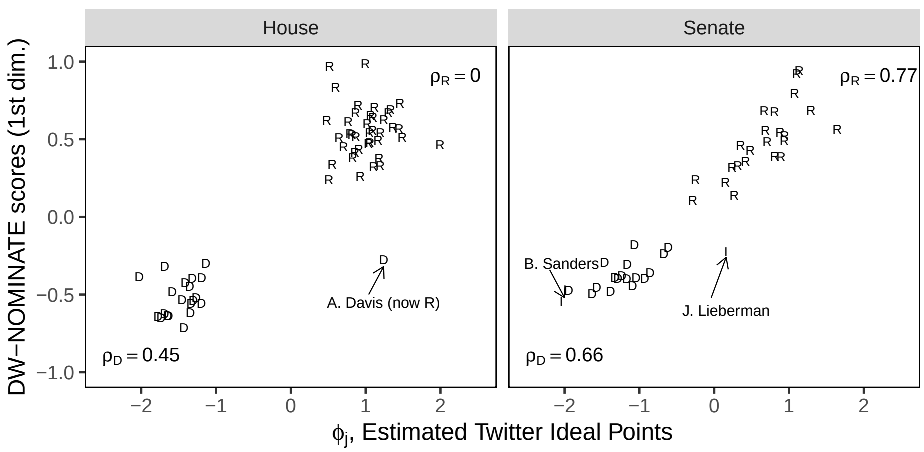

Figure 2: Euclidean ideal point estimates for members of congress

same. Both models struggle with house Republicans and preform best with senate Republicans. The bilinear estimates have a systematically lower correlation with DW-NOMINATE scores by 2-5 points, although it is important to note that this does not imply the bilinear estimates are worse. Assuming DW-NOMINATE scores are not a perfect metric, a higher correlation with these scores may indicate estimates share these flaws to a greater degree. In other words, the correlations are not a definitive measure for the estimates accuracy, but rather a qualitative barometer to judge the face validity of the estimates and if they behave as expected. In this regard, the bilinear estimate preform well. Correlations are largely similar and it correctly classifies outliers within the data such as Joe Lieberman and Artur Davis as the Euclidean model does. One thing to note is that the bilinear model scales slightly differently. While the Euclidean model ranges from roughly -2 to 2, the bilinear model ranges from roughly -3 to 3.

Figures 3 (bilinear) and 4 (Euclidean) compare ideal point estimates for key political actors with the same estimates in the original paper. The placement and ordering of individuals on the spectrum is largely the same. Both place Republican Jon Huntsman in the center of the spectrum with the same media personalities at the extreme end of each spectrum. Both additionally place the median senator for both parties closer to the center than the median house member. The one notable exception to this is President Obama which the bilinear model places further to the left. This may simply be model differences, or it may be a function of the α_j (popularity) term, which President Obama has the largest value of in the data set. Reasons for why this may be are discussed in the discussion of this section.

Finally, I examine the distributions of the estimates in figures 5 (bilinear) and 6 (Euclidean). As expected, both models produce nearly identical distributions for all four political affiliations.

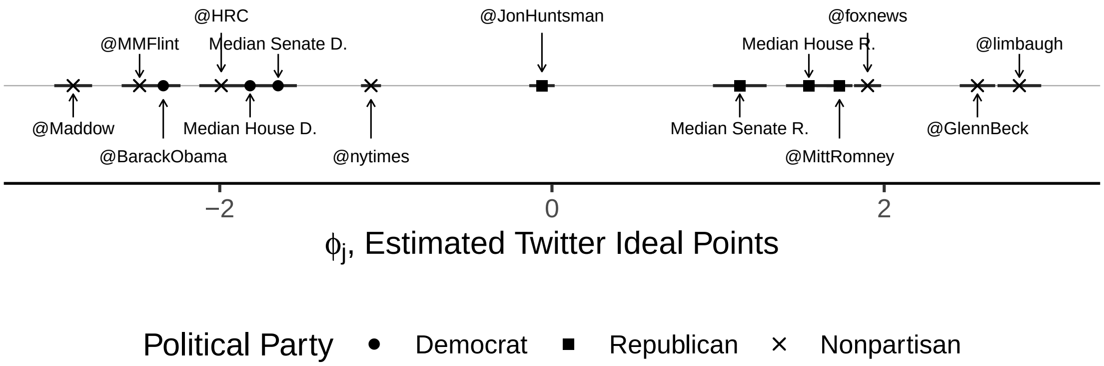

Figure 3: Placement of key elites by the bilinear model

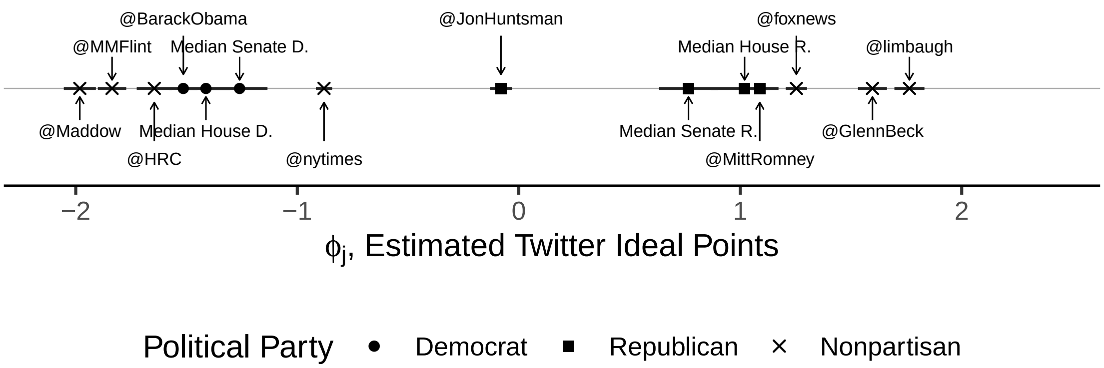

Figure 4: Placement of key elites by the Euclidean model

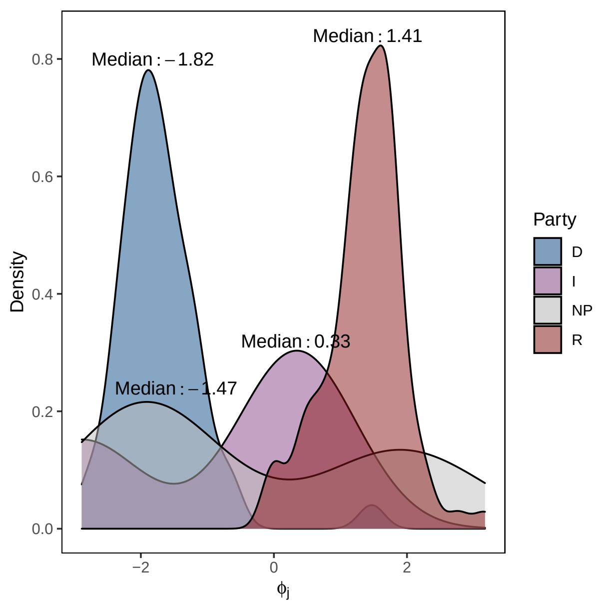

Figure 5: Bilinear ideal point distributions

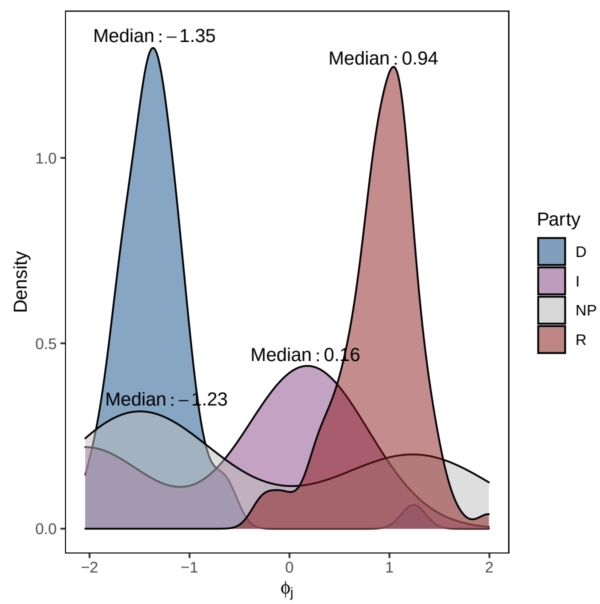

Figure 6: Euclidean ideal point distributions

3.3 Discussion

This analysis shows that the bilinear specification of the model was able to fit the ideological latent space and achieve similar results to the Euclidean model. However, a few questions remain. First, and most obviously, more rigorous testing is needed. While the model appears valid on its face, it remains to be seen if the bilinear estimates are in fact more predictive than the Euclidean ones and if it is a meaningful improvement.

Second, this analysis may have revealed another potential area for improvement. By subsetting the data set to 165 elites rather than using the full sample, correlations with the DW-NOMINATE scores dropped considerably. As noted earlier, this is not necessarily a bad thing but the size of the drop in some populations gives cause for suspicion. For example, in the original study the correlation between the the Euclidean model estiamtes and DW-NOMINATE scores among house Republicans was 0.36. By subsetting the data this same correlation dropped to zero. It is unclear why this occurs, but I suspect it has something to do with the α_j term which is meant to capture a political elite's popularity. There are two reasons I suspect this: First, the sample of elites used in this study was a function of popularity as it was subset by the number of followers. Thus, the data points missing from the sample were the least popular ones. Second, in the bilinear model the α_j term struggled to converge. Currently the model uses the logarithm of the number of followers for an elites initial value of α_j. This seems a sensible starting point for the variable meant to capture preferential attachment in the network. However, perhaps there is a way to improve the starting point, or another variable need be introduced to the model.

Together, these two questions mean it remains unclear how consequential this re-specification of the model is. While this analysis shows there is potential in the bilinear model, it is unknown if it represents a meaningful improvement in compute times or accurately estimating the latent space.

4 Twitter's Echo Chamber

In the analysis portion of Barbera's paper he applied his point estimates to examine the polarization of political discourse on Twitter. In this section I re-examine this analysis by revisiting the original retweet network from the paper. I start with a brief overview of the network size and structure, and then replicate and expand on the original analysis. While Barbera's analysis was largely limited to tweets mentioning one of the two 2012 presidential candidates, I take a deeper dive and examine if the patterns he identified in candidate centered discussion are also reflected in the broader network structure.

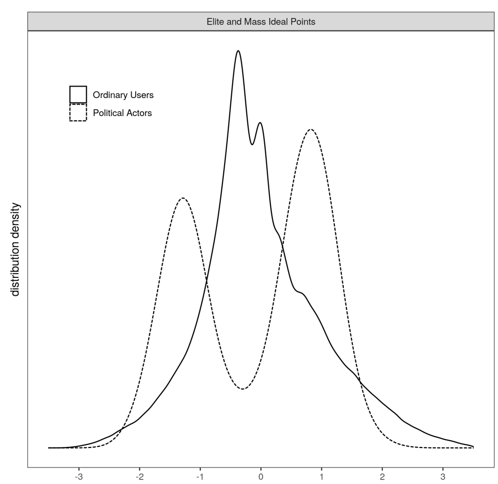

Figure 7: Distribution of ideal points within the retweet network

4.1 Data

The data set consists of 15 million tweets sent by 301,537 users in a retweet network of politically interested users. In total there are over 6 million edges or retweets. A network of this size is simply too large to visualize with conventional methods and sampling down to reasonable sizes simply results in sparse networks due to the fact that the larger network is itself a random sample. A plot of the network is in the appendix, but for purposes of this analysis I rely on multiple other types of visualizations to portray the networks characteristics.

The distribution of ideological groups within the network is shown in figure 7. For ordinary users the distribution is dominated by moderates. As we will see, however, despite making up the clear majority moderates do not drive the Twitter conversation. Figure 8 shows the distribution of degree for each node, or the number of unique users that have retweeted an account, with the left graph keeping all observations and the right graph removing outliers. As we can see, Twitter is a platform dominated by extremes. The vast majority of accounts are retweeted by only one to two users while a tiny minority extend the axis out to over 80,000. Accordingly, the network follows a clear pattern of preferential attachment with a Kolmogorov-Smirnov test p value of 0.85 (see appendix B for plot).

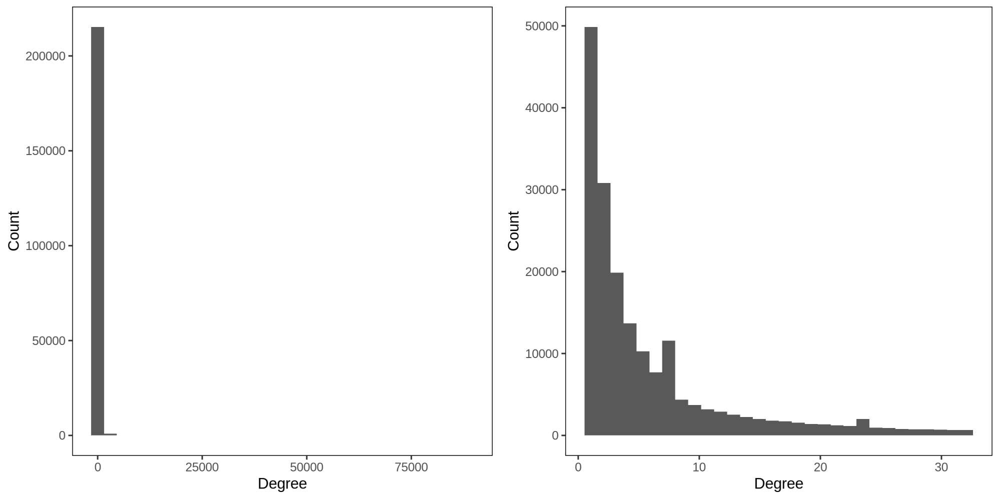

Figure 8: The distribution of degree in the network with all data (left) and with outliers removed (right)

Barberas analysis made two primary claims. First, that the conversation on Twitter is largely driven by partisans criticizing the other party and that this pattern is particularly pronounced among conservatives. Second, that the topology of the conservative network facilitates rapid and broad dissemination of political information. I will examine each claim in turn.

4.2 Partisanship in Retweet Patterns

In figure 9, I replicate Barberas distribution plot of tweets mentioning presidential candidates. The plots exhibit a strongly bimodal distribution around ideal points of -1 and 1. Despite forming the clear majority of users, moderates appear to be less active than partisans in terms of tweet generation about the candidates. In line with the authors claim, tweets mentioning a given candidate are generated more frequently by the opposite party. Partisans, it seems, enjoy talking more about the other candidate than their own.

In figure 10, I replicate the heat map shown in Barberas original analysis. This figure shows that retweet pattern within the network and largely echos the findings in the previous paragraph. Conservatives are more likely to retweet each other about Obama than are liberals, while liberals are more likely to retweet each other about Romney than conservatives. This is strong evidence of an echo chamber effect: When discussing presidential candidates partisans are more likely to discuss opponent candidates with fellow partisans than to engage in cross-ideological debate.

This provides clear evidence of an echo chamber with regards to presidential candidates, however I wanted to examine if this trend is reflected in the broader network structure. I start by looking at reciprocity within the network, or the tendency for users to mutually retweet each other. Reciprocity within the entire network is 0.03. This is unsurprising given that the network is composed of a random sample rather than a comprehensive sample of organically developed networks. To take a deeper, look I calculate the number of reciprocated ties for each node within the network. As shown in figure 11 I find that it follows a bimodal structure similar to that of the heat map. This provides some evidence of a reinforcing characteristic of the echo chamber: Picking sides has social benefits in the form of mutual relationships. Among conservatives ($\theta_i > 0.5$) this seems to be rewarded the most with an average number of reciprocated ties at 0.97, higher than that of liberals ($\theta_i < 0.5$) who on average had 0.59 reciprocated ties. Moderates had the fewest on average at 0.17.

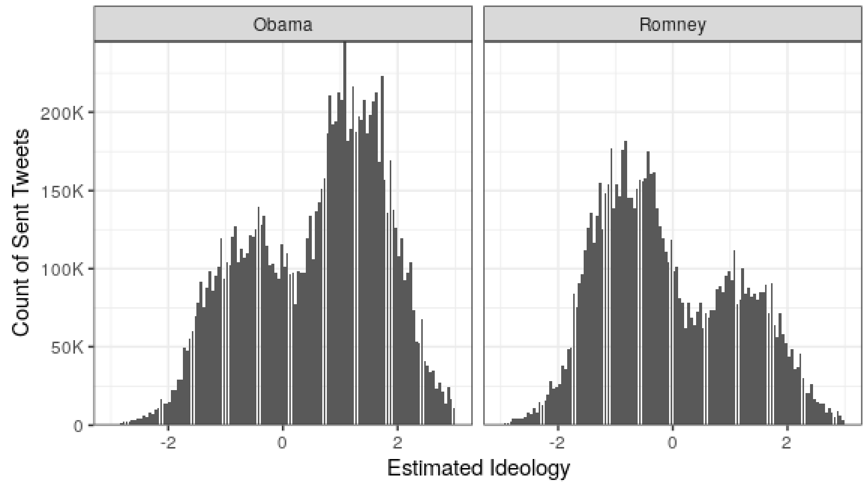

Figure 9: Distribution of tweets mentioning a presidential candidates name by ideal point

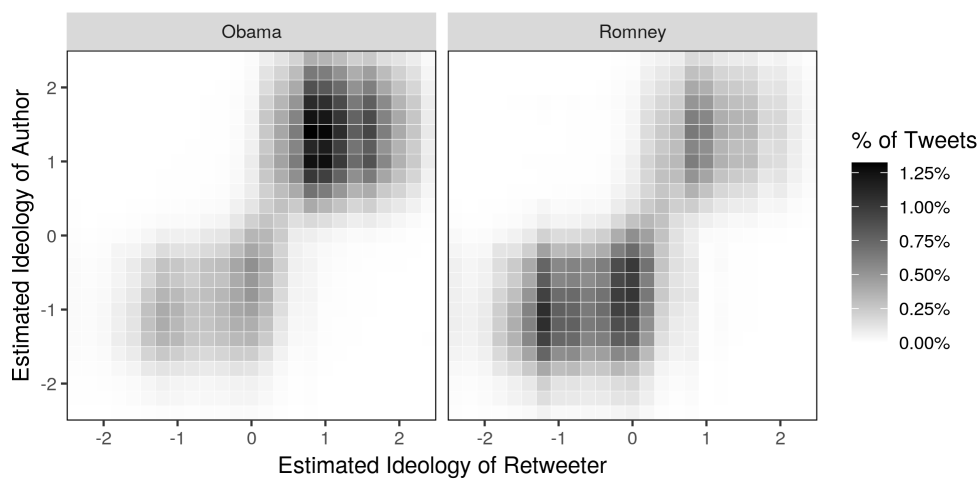

Figure 10: Retweets mentioning a candidates name by ideal point

One of the strongest pieces of evidence for Twitters political community to reinforce its echo chamber is patterns of popularity. In figure 12, I use a two dimensional histogram to plot the popularity of a node against its ideological point score. I use (α_i rather than degree in order to prevent outliers from rendering the data uninterpretable. As shown, the majority of users show low (α_i centered around a θ_i of zero. In other words, the majority of users are centrists of average popularity. Notably the graph tends upwards as ideology moves away from the center. This indicates there is a positive correlation with popularity and ideological polarity. To confirm this I calculated the correlation between the absolute value of θ_i and popularity as measure by both (α_i and degree. Across the entire network there is a positive correlation (α_i and $|\theta_i|$ ρ = 0.179, degree and $|\theta_i|$ ρ = .031). When broken down by ideological group, popularity metrics seem to buck the trend of conservatives (α_i and $|\theta_i|$ ρ = 0.107, degree and $|\theta_i|$ ρ = 0.017) having a stronger echo chamber effect than liberals do (α_i and $|\theta_i|$ ρ = 0.212, degree and $|\theta_i|$ ρ = 0.020). Interestingly, while there is a positive correlation between popularity and partisanship, the range of a user's potential popularity shrinks as they move further towards the ideological extremes. The minimum value of popularity appears to follow an exponential growth curve while the maximum value of popularity seems to follow exponential decay until the two points converge. This indicates a high floor, low ceiling phenomenon where a certain level of extremism guarantees an audience, but also limits the potential size of that audience.

Finally, I look at assortativity within the network. Across the entire network ideological assortativity is 0.75, indicating a strong preference for forming ties with like minded Twitter users. Taken together these pieces of evidence support that the notion that political discourse on Twitter is a self-reinforcing echo chamber. There is some evidence that this

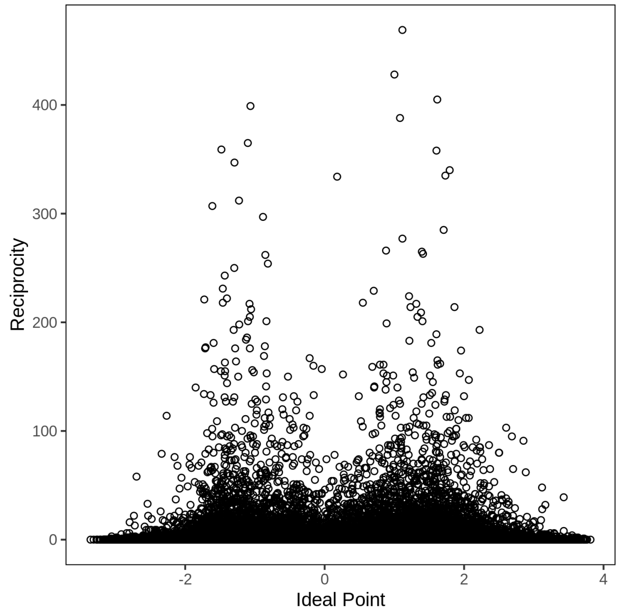

Figure 11: Number of reciprocated ties per account plotted against ideal point

effect is more pronounced among conservatives as conservatives tend to reciprocate ties among their fellow partisans to a greater degree. However, the correlation between popularity and polarity is stronger among liberals than conservatives, casting some doubt on the notion that this effect is demonstrably more pronounced among conservatives.

4.3 Information Flow

I now turn my attention towards Barberras claim that the topology of the conservative network facilitates faster and more broad dissemination across the network. The author did not actually explore this claim, and so I examine betweenness centrality and transitivity within the network to do so. Because the network is too large to calculate exact betweenness centrality in a reasonable time frame, I estimate betweenness centrality with a cutoff of three. This measure estimates the degree to which nodes stand between each other and thus provides some indication as to how efficiently and quickly information can flow throughout the network. However, it is somewhat difficult to contextualize because Twitter is a network of extremes. The median betweenness centrality for the entire network is zero and the max exceeds 90 million. Notably, conservatives are the only group with a non-zero median centrality of 0.177. To account for outliers I use the logged betweenness centrality. Among the three ideological groups, conservatives have the highest mean logged centrality (3.48) while liberals have a slightly lower average (3.11). Moderates have the lowest at 1.80.

Transitivity within the network is the propensity for nodes to cluster and form closed triads. This property, combined with betweenness centrality, allows information to flow more quickly and disseminate broadly throughout the network. For this network I calculate both the global transitivity and the local transitivity for each node. Across the entire network the transitivity is 0.047. Among ideological groups, conservatives again have the highest mean local transitivity (0.153) while liberals have the second highest (0.146). Moderates have the lowest at 0.117.

Given the patterns in centrality and transitivity, this analysis supports the hypothesis that the network of conservatives on Twitter is more conducive to spreading information quickly and broadly. As a whole Barberras hypotheses are supported by this analysis; Twitter does exhibit the traits of an echo chamber and that effect appears to be more pronounced among conservatives. The one counterpoint is the stronger correlation among liberals between popularity and polarity. Barring this, however, the data generally points in the same direction. Conservatives tend to have more reciprocated ties than liberals and the topology of their network is more conducive to rapid information spread, potentially producing a stronger echo.

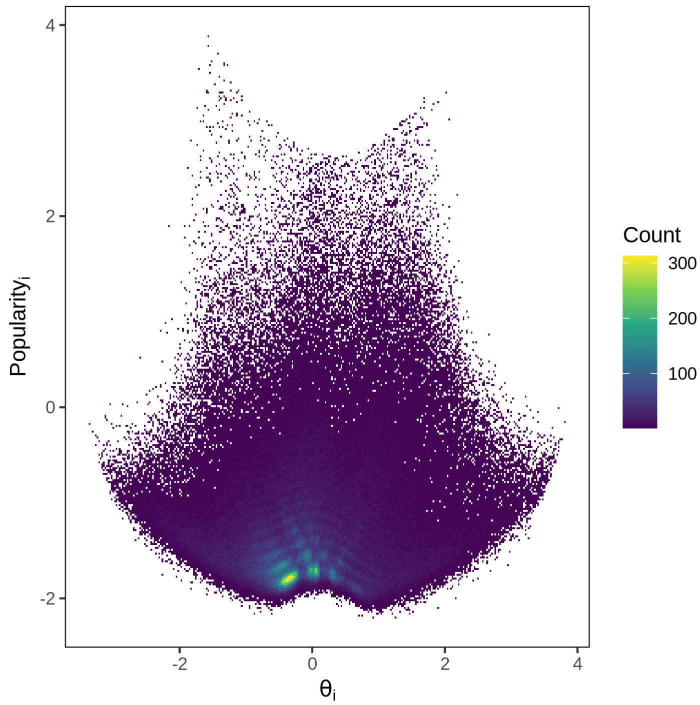

Figure 12: 2d histogram of popularity (α_i) and ideal point

5 Discussion and Research Agenda

In this paper I both explored a re-specification of Barberas original model and re-examined his exploration of political retweet networks in greater depth. I found tentative support for some benefits of changing the functional form of the model from Euclidean to bilinear, and fairly strong evidence confirming the authors original analysis of the retweet network. More than anything, however, this paper identifies a clear direction for future inquiry. To continue this work I have identified three specific areas on which to focus. First is to diagnose and fix any specification issues in the bilinear model. Most notably, more robust testing around the (α_i term should be done. To do so, the first step is to run the bilinear model for more iterations and see if convergence improves. If it does not, potentially new initial values should be tested or additional variables introduced to the model. Additionally, the original Euclidean model should undergo more testing for robustness. Larger samples of data bring forth misspecifications in the model that do not emerge at smaller samples. Within small samples, the (α_i term in the bi-linear model converged quickly. At larger samples, it did not. Currently the Euclidean model is run on a set of 10,000 samples. It would be worthwhile to test if the model continues to hold with more data or if additional information will highlight potential areas for improvement.

Second, once a model has been correctly specified and mass ideal points are estimated, a more robust testing procedure should be executed. One potential approach is to examine the predictive power of each models estimates in regression and classification tasks. Ideally this could be carried out on more current data than was used in this replication.

Finally, it is probably worthwhile to implement this methodology in a more accessible and robust software stack. Whether the model is re-specified or not, this approach to estimating ideal points appears quite robust and has many applications. Its primary downside is that it is fairly inaccessible due to the time it takes to implement both programmatically and computationally. The model requires two stages in which elite ideal points are first estimated, and then mass ideal points are estimated based on the results of the elite estimates. This means that elite estimates would need to be recalculated on a semi-regular basis in order to get reasonably accurate mass estimates. This is fairly cumbersome no matter the infrastructure surrounding the process, but it can still be improved. Since the original publication of Barberas paper, software tools for network analysis and Bayesian inference, as well as computation power, have advanced considerably. With some additional development it would be possible to cut out a number of dependencies and downsize the code base to a more streamlined pipeline. Given this, elite estimates could more readily be calculated and made public at regular intervals, and researchers could draw on these current estimates to calculates ideal points for whatever group they are studying.

References

[1] Pablo Barberá. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, 23(1):76–91, 2015.

[2] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.17.3.

A Elite Ideology Latent Space Model

Table 1: Elite Ideology Latent Space Model Parameters

Parameter	Bilinear	Euclidean
Iterations	3,000	500
Warm up	2,000	100
Thin	1	2
Chains	5	2

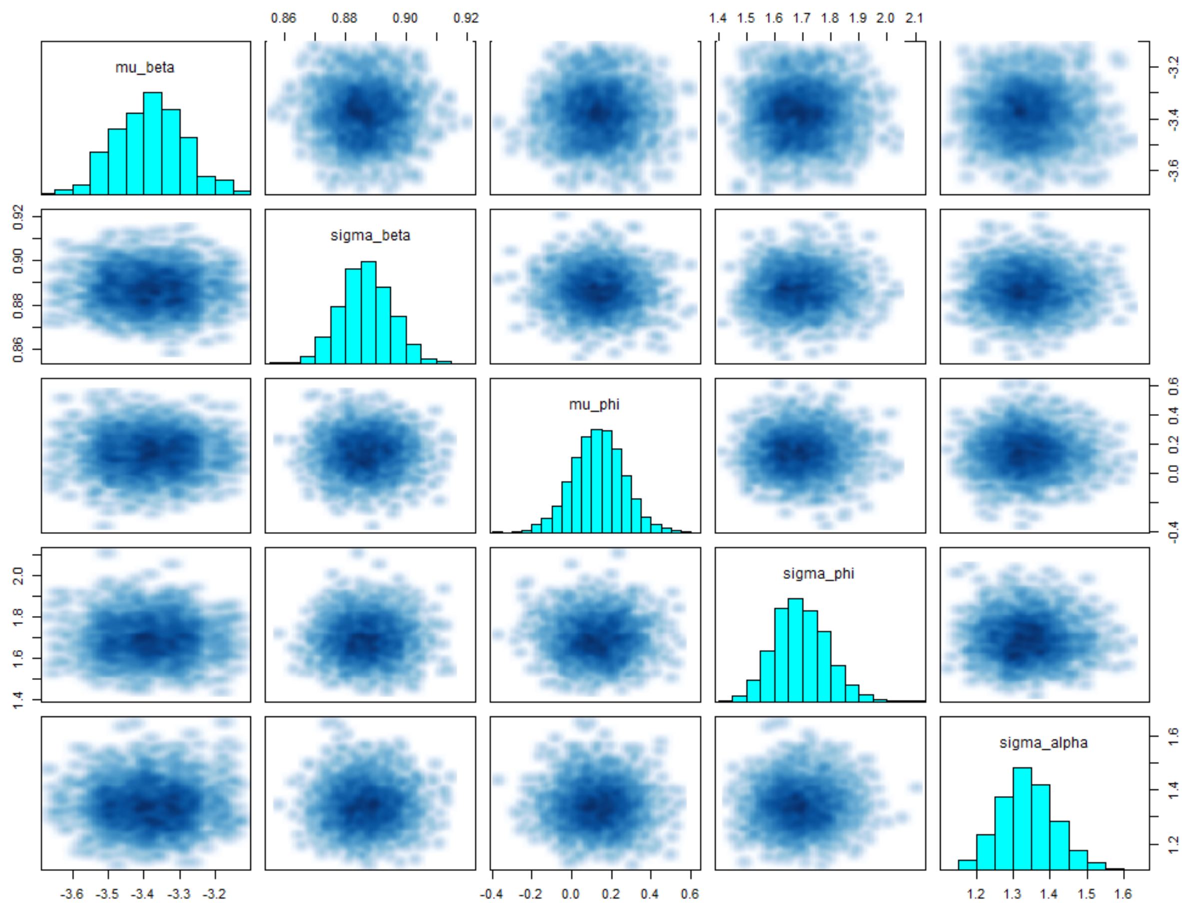

Figure 13: Pairs plot for bilinear latent space model

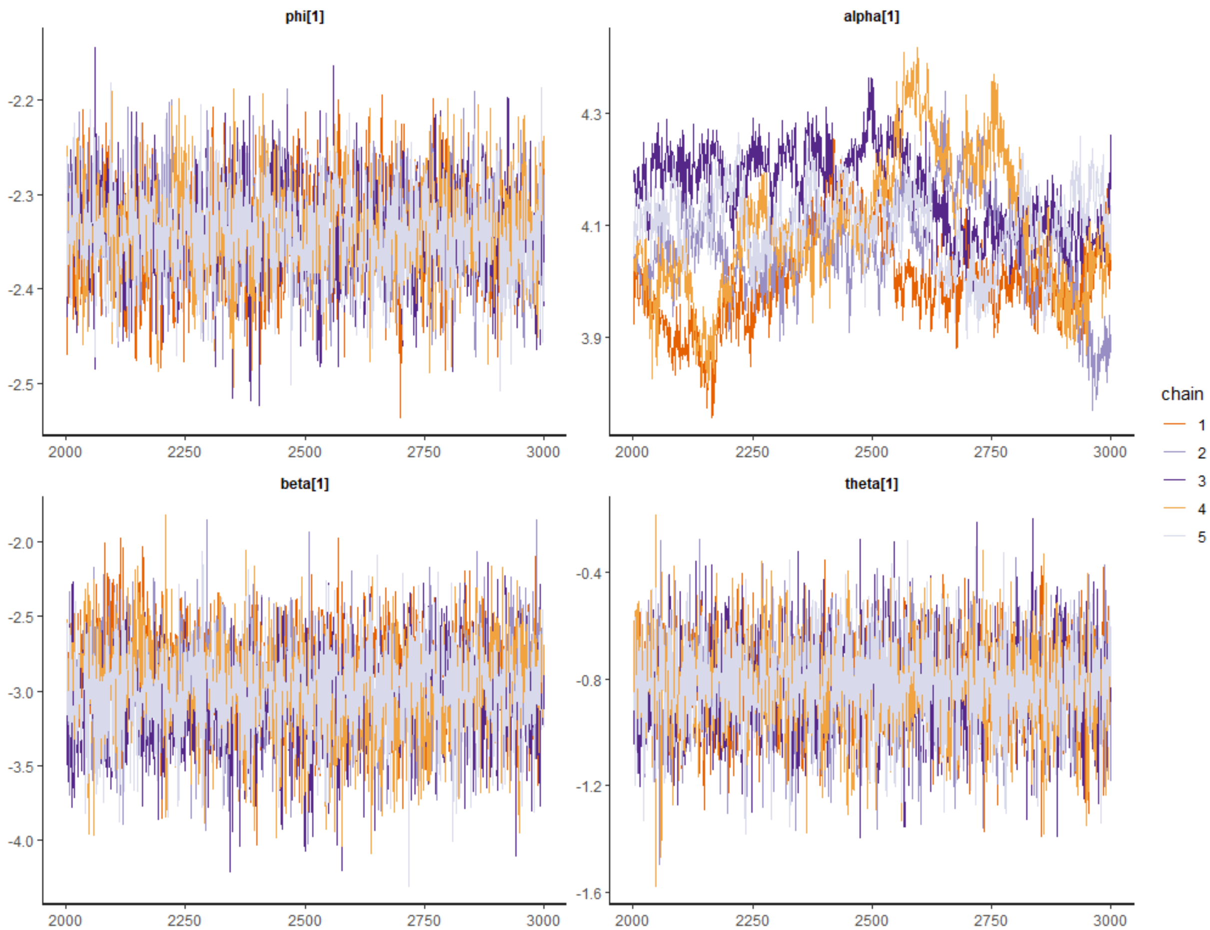

Figure 14: Trace plot for bilinear latent space model

B Additional Plots

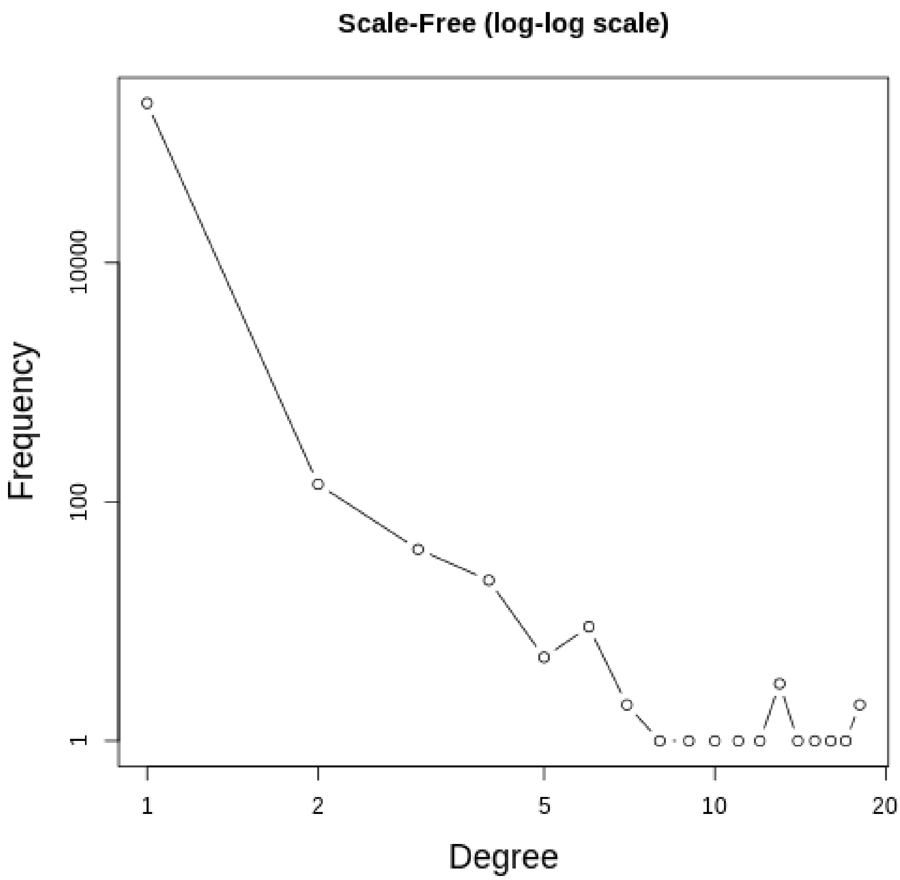

Figure 15: Preferential attachment in the retweet network

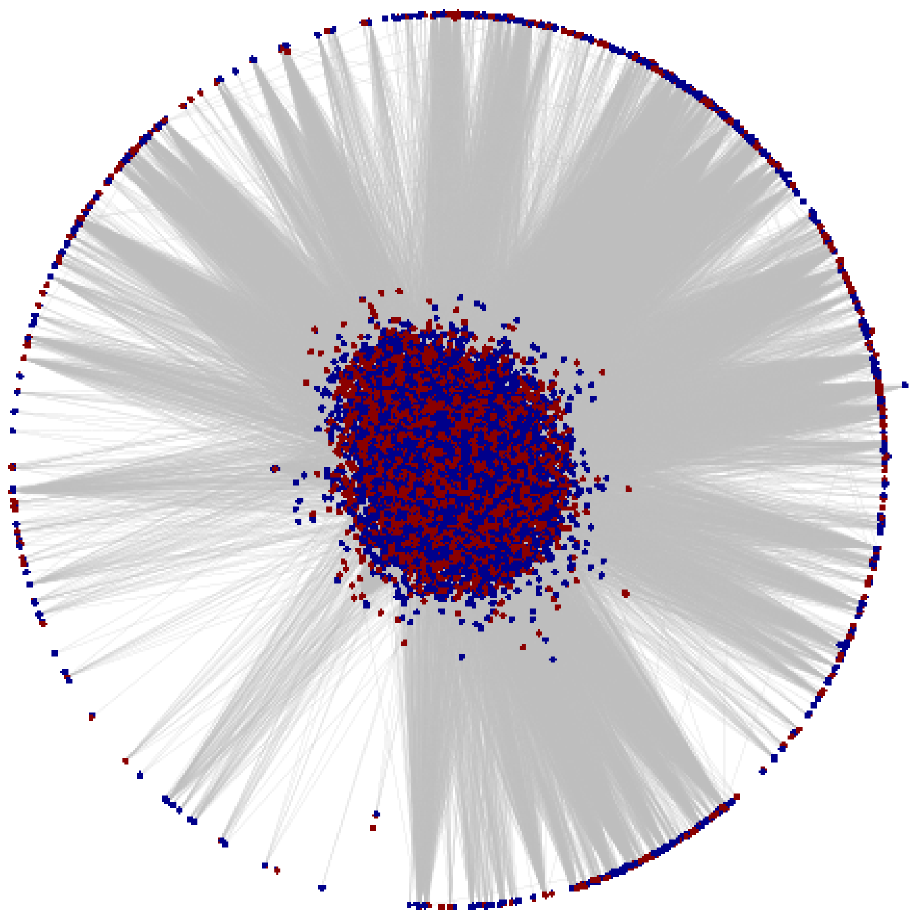

Figure 16: The full retweet network colored by ideology. Theta <= 0 is colored blue and theta > 0 is red