Centrality of Congress Twitter Network

Tim Ballard, Ruite Guo & Liam Johnston

1. Introduction

The idea of centrality and human communication was first introduced at M.I.T. in the late 1940s. These first studies revolved around analyzing a group of people and "concluded that centrality was related to group efficiency in problem-solving, perception of leadership and the personal satisfaction of participants" (Freeman 1978). Since then, graph theory has developed many ways to identify the most "central" or influential people in a network. This study examines multiple centrality algorithms and attempts to identify which politicians in the U.S. congress are most influential via the twitter network formulated from current congress members.

2. Methods and Data

In network analysis a natural objective is to identify the "central" nodes. With this objective a logical question would be to ask what defines a "central" node (active, important, non-redundant, etc.) (Hoff, n.d.). In a study in 2008, Koschutzki suggested classifying centrality measures under the following four constructs:

1. Reach: A node is central if it reaches many other nodes

2. Flow: A node is central if many edges pass through it

3. Vitality: A node is central if its removal drastically disrupts the network

4. Feedback: A node is central if it is connected to other central nodes

(D. Koschutzki 2008).

Following Koschutzki's definitions of centrality, we will compare four contemporary approaches: degree centrality, closeness centrality, betweenness centrality and eigenvector centrality. Each measure aims to capture a particular construct suggested by Koschutzki.

The data used for this work was compiled through Twitter's REST API. The current members of Congress' Twitter handles were obtained from github (Gaber 2017). Provided with handles, Twitter's API provides access to that user's entire public Twitter history. For our purposes, we recorded details of every instance where a member of our user base retweeted any other member of our user base. Once this data was collected, a Python script was employed to construct an adjacency matrix, with retweets serving as the edge connecting two nodes (politicians).

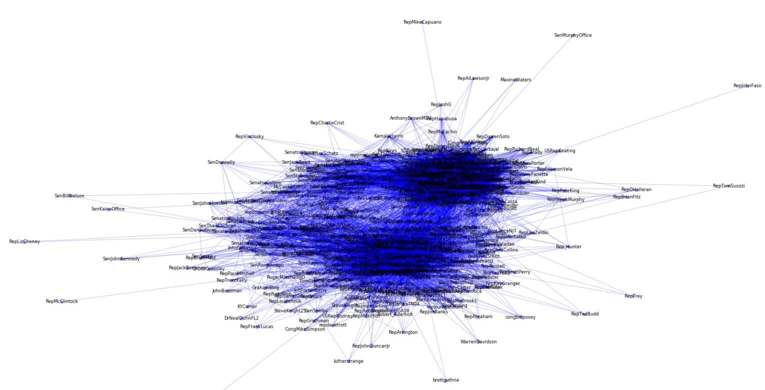

Figure 1: Graphical model (strongly connected) of the U.S. Congress Twitter network where nodes are the individual members of congress and the edges are an indicator determining if a retweet took place between the two corresponding nodes.

Throughout this paper, we will be using the following notation, defining a given graph/network $G := (V, E)$ with $|V|$ vertices and $|E|$ edges, a centrality measure at node v as $C(v)$, the number of neighbors of v_i as $deg(v_i)$, the distance between nodes v_i and v_j as $d(i, j)$, the number of shortest paths between ending nodes v_i and v_j as $b_{i,j}$, the number of shortest paths between ending nodes v_i and v_j that include node v_k as $b_{i,j}(k)$ and the adjacency matrix $A = (a_{i,j})$ as the matrix of indicators determining if an edge exists between node v_i and v_j.

3. Centrality Measures

In the hope of being able to accurately depict the socially central players in the U.S. congress, multiple measures of centrality will be defined and implemented in the following sections. Since there are over 500 members in the U.S. congress, for the sake of brevity we will be reporting the top ten most central congress members under each centrality construct.

3.1: Degree Centrality

Degree centrality is a mechanism that measures the reach of a node. It's defined as a sum of the number of neighbors a node has. For example, a node with a large number of neighbors would have a large degree centrality.

Definition 3.1. (Degree Centrality). $C_D(v_i) = deg(v_i)$

Table 1: Ten most central politicians in U.S. congress computed via degree centrality.

Politician	Degree Centrality
SpeakerRyan	207
SteveScalise	172
GOPLeader	168
WhipHoyer	161
NancyPelosi	140
repjohnlewis	137
repjoecrowley	120
RepSwalwell	117
rosadelauro	113
cathymcmorris	109

3.2 Closeness Centrality

One of the measures used to measure the flow of a network, and identify nodes that are central in concordance with flow is closeness centrality. Closeness centrality is defined as the inverted distance away from all other nodes in the network. Under this measure, nodes that are separated from others by many nodes/edges inherently exhibit a large distance measure, and thus a low closeness centrality. Nodes with the shortest sum of paths to every other node in the network will be ranked as most central.

Defintion 3.2 (Closeness Centrality). $C_{Close}(v_i) = (\sum_{j=1}^{N} d(i, j))^{-1}$

Table 2: Ten most central politicians in U.S. congress computed via closeness centrality.

Politician	Closeness Centrality
RepJayapal	3.58397e-05
RepRaskin	3.58179e-05
RepEspaillat	3.57577e-05
RepKihuen	3.57488e-05
RepCarbajal	3.57334e-05
CongressmanRaja	3.57245e-05
RepJimmyPanetta	3.57219e-05
RepStephMurphy	3.57207e-05

Politician	Closeness Centrality
RepJimBanks	3.56913e-05
RepDarrenSoto	3.56812e-05

3.3 Betweeness Centrality

Koschutzki's third definition of centrality, vitality, is a measure of the extant that the removal of a node would effect the network. Betweenness centrality is able to capture this centrality principle as it is a measure of the extant that a node bridges two other non-connected nodes.

Defintion 3.3 (Betweenness Centrality). $C_B(v_i) = \sum_{i \neq j \neq k} \frac{b_{j,k}(i)}{b_{j,k}}$

Table 3: Ten most central politicians in U.S. congress computed via betweenness centrality.

Politician	Betweenness Centrality
SpeakerRyan	21053.492
SteveScalise	13018.571
GOPLeader	10362.329
WhipHoyer	5621.354
NancyPelosi	4505.782
RepKevinYoder	4270.130
cathymcmorris	3898.104
RepTerriSewell	2987.263
RepDonBeyer	2787.523
JohnCornyn	2673.580

3.4 Eigenvector Centrality

The last of the four definitions of centrality is highly related to PageRank. The idea behind eigenvector centrality (and also PageRank) is that a node is important/central if it is connected to other important nodes (Sabidussi 1966). This centrality can sometimes be misleading, as it can drastically differ from the degree of the node (Bonacich 1987). For example, a node with a high degree may have a low eigenvector centrality if all of the linked nodes have a low eigenvector centrality. The reverse is also true, a node can have few neighbors but a high eigenvector centrality if its neighbors are highly important. The mathematical definition of eigenvector centrality is defined as the sum of the node's neighbors' eigenvector centralities (Freeman 1978).

Defintion 3.4 (Eigenvector Centrality). $C_\lambda(v_i) = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} C_\lambda(v_j)$

where λ is a constant.

In matrix form this is equivalent to:

$$\lambda \mathrm{x} = \mathrm{x} A$$

From this notation it is easy to see that the centrality vector, x, is the left-hand eigenvector of the adjacency matrix, A, associated with the corresponding eigenvalue, λ.

Choosing λ is usually done following the Perron-Frobenius theorem which we introduce below. To clarify why we can apply Perron-Frobenius theorem, we must introduce another term, *primitive*. Because our adjacency matrix is not real positive valued, as it contains 0 entries, we must check if our network is *primitive*.

Definition 3.5 (Primitive). Matrix A is *primitive* if the following properties hold:

1) For any $1 \leq i, j \leq n$, there is an integer $k \in \mathbb{Z}$, s.t. $A_{ij}^k > 0$

2) Any node pair $(i, j) \in E$ are connected with a path of length $\leq k$

3) A has unique $\lambda^* = \max|\lambda|$

Our network is strongly connected and is thus *primitive.* Since our adjacency matrix fulfills this criteria, we are able to apply the Perron-Frobenius theorem for non-negative matrices as well as the extension provided through the Collatz-Wielandt Forumla to ensure convergence.

Theorem 3.1(Non-negative Matrix, Perron-Frobenius). Assume that $A \geq 0$ and A is primitive. Then

1) $\exists \lambda^* > 0, v^* > 0, ||v^*||_2, s.t. Av^* = \lambda^* v^*$ and $\exists w > 0, ||w||_2 = 1, s.t. (w^T)A = \lambda^* w^T$

2) $\forall$ other eigenvalue λ of A, $|\lambda| < \lambda^*$

3) v^* is unique

The Perron-Frobenius Theorem was extended to:

Lemma 1. (Collatz-Wielandt Formula).

$\lambda^* = \max\ \min_{x>0} \frac{[Ax]_i}{x_i} = \min\ \max_{x>0} \frac{[Ax]_i}{x_i}$

(Weisstien, n.d.)

These results guarantee that if our graph is strongly connected, then the eigenvector solution x is both positive and unique, both desirable properties. The power method is the technique used to compute this eigenvector centrality problem. From Perron-Frobenius theorem we can conclude that x_k converges to the eigenvector associated with the dominant eigenvalue (Boyle, n.d.).

Algorithm 1.(Power Method). For $k \geq 0$:

1. Initialize random vector x_0
2. $x_{k+1} = \frac{Ax_k}{||Ax_k||}$

After applying the power method to the adjacency matrix we can find each nodes' corresponding eigenvector centrality (results are below).

Table 4: Ten most central politicians in U.S. congress computed via eigenvector centrality.

Politician	Eigenvector Centrality
SpeakerRyan	0.01310
SenJohnMcCain	0.01019
SteveScalise	0.01006
GOPLeader	0.00963
WhipHoyer	0.00876
repjohnlewis	0.00828
NancyPelosi	0.00770
SenSchumer	0.00648
senrobportman	0.00648
cathymcmorris	0.00623

4. Discussion

Our experiments, under the assumption that the network formed from congress' twitter interactions is representative of congress' social influence, have successfully identified which politicians (nodes) are the most influential. Interestingly, over the four measures of centrality, we observed different politicians (nodes) being deemed more central depending on the mode of centrality being employed. This isn't to say that some of the measures of centrality are incorrect, but merely the aspect that you define centrality in will determine the structure of the network that is defined as important. In fact, we would argue that the discrepancies of rank in the four centralities actually speak to the diversity of graph structure that is measured and is also more symbolic of the underlying structure of the social network. To gather a holistic portrayal of which politicians are most socially influential, a single centrality construct really isn't appropriate. Each respective definition is a facet of network centrality, but we'd argue no one measure provides a comprehensive definition of network centrality. This being our view point, a natural extension of this would be to assign a weighted value to each centrality measure in order to get a holistic view of the U.S. congress' most central players.

Definition 4.1(Comprehensive Centrality) Ranking each node from $1, ..., n$ under each centrality measure and denoting the rank of node j under centrality measure i as R_{ij}, |centrality measures| as the number of centrality measures and w_j as the weight corresponding to each centrality measure, we can define comprehensive centrality as:

$$C_{comprehensive}(v_i) = (\sum_{j=1}^{|\text{centrality measures}|} w_j R_{ij})^{-1}$$

We inverted this measure so that we can keep consistency across centrality measures, in that nodes with high centrality have a higher comprehensive centrality, $C_{comprehensive}$. Applying this comprehensive metric to our centrality measures with equal weights summing to 1 and $j = 1, ..., 4$ and $i = 1, ..., 527$ we get the following results:

Table 5: Ten most central politicians in U.S. congress computed via comprehensive centrality.

Politician	Comprehensive Centrality
SpeakerRyan	0.1739130
SteveScalise	0.1250000
GOPLeader	0.1025641
WhipHoyer	0.0975610
NancyPelosi	0.0816327
cathymcmorris	0.0606061
RosLehtinen	0.0526316
RepSwalwell	0.0380952
rosadelauro	0.0370370
repjoecrowley	0.0363636

From the comprehensive centrality results, we see some interesting results. Comprehensive centrality reveals that its top ten socially influential politicians overlap with the top ten central politicians for centrality measures corresponding to degree, betweenness and eigenvector, but not for closeness. In fact, the first politician that ranked in the top ten for closeness centrality doesn't show up in our comprehensive measure until the 155^{th} politician. This result leads to two possible explanations, both being areas that future research could explore. The first being that degree, betweenness and eigenvector centrality are very correlated centrality measures, thus making our comprehensive centrality measure biased away from closeness centrality. If this is the case, our comprehensive centrality measure could be adjusted so that the weights corresponding to each centrality measure were weighted in a way such that the bias introduced by highly correlated centrality measures was reduced/eliminated. The second being that closeness centrality is measuring an aspect of centrality that is nearly disjoint from the aspects the other three centrality measures are grading. If this is indeed the case, and closeness is assigning importance to a disjoint network architecture, we would expect our results to appear as they are.

References

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." The University of Chicago Pres.

Boyle, Mike. n.d. "Notes on the Perron-Frobenius Theory of Nonnegative Matrices." Department of Mathematics, University of Maryland. www.math.umd.edu/~mmb.

D. Koschutzki, Falk Schreiber. 2008. "Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks." Gene Regulation; Systems Biology.

Freeman, Linton C. 1978. "Centrality in Social Networks." New York: Social Networks. https://www.bebr.ufl.edu/sites/default/files/Centrality%20in%20Social%20Networks.pdf.

Gaber, Yonah Bromberg. 2017. "A List of Twitter Handles for Members of Congress."

Hoff, P. n.d. "Centrality." University of Washington.

Sabidussi, Gert. 1966. "The Centrality Index of a Graph." Psychometrika.

Weisstien, Eric W. n.d. "Perron-Frobenius Theorem." MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Perron-FrobeniusTheorem.html.