Stability of Centrality Measures in Complex Networks

Daniel Lee, Dr. Ross O'Connell, Dr. Katherine Anderson

Carnegie Mellon University

Abstract

Many real world interactions can be modeled using complex networks, a mathematical structure in which a set of nodes are connected by edges, which represent the relationship between them. The amount of influence a node has in the network is quantified through a wide variety of centrality measures, each of which has a different interpretation. In the context of a particular network, it is unclear which centrality measure is the most meaningful. Here, we explore one method for comparing different centrality measures: their stability. In a given network, we calculate the centrality of the nodes using a particular measure. We then leave each node out from the network, creating an ensemble of similar networks and recomputing the centrality of the nodes. We define a measure using the Kendall-tau distance which counts the number of pairwise disagreements between the recomputed centrality ordering of the nodes and the original ordering. We illustrate this method by comparing the degree, betweenness, closeness, an eigenvector centrality on real and artificial networks. Our empirical networks are canonical examples: the Karate-Club, Jazz, Netscience, and Hep-Th networks. The two artificial networks are the preferential attachment and small world networks. We illustrate that stability can be used to distinguish between centrality measures and identify similarities within ensembles of networks. In future work, we hope to repeat this process on other networks in hopes of finding a definite relationship between the type of network and their most stable centrality measure.

Measuring Stability

Centrality Measures

Four centrality measures were evaluated for each network. The definitions for each measure are defined below.

1. The Degree centrality C_D is defined to be the number of neighbors given a node
2. The Betweenness centrality C_B quantifies the number of times a node serves as a bridge along the geodesic between two other nodes. For a given graph *G*, it is expressed as

$$C_B = \frac{1}{(N-1)(N-2)} \sum_{i \neq k \neq j \in G} \frac{n_{ij}(k)}{n_{ij}}$$

-where *N* is the number of nodes in *G*
-n_{ij} is the number of geodesics between nodes *i* and *j*
-$n_{ij}(k)$ is the number of geodesics between nodes *i* and *j* that pass through *k*

3. The Closeness centrality C_C measures the "closeness" of a node to all others in the network. It is calculated using the reciprocal of the mean distance from a node to all others. For a node *i*, it is defined as

$$C_C = \frac{N-1}{\sum_{j \in G, j \neq i} d_{ij}}$$

-where d_{ij} is the length of the shortest paths between *i* and *j*

4. The Eigenvector centrality C_E, like the Degree centrality, measures a nodes influence based on the number of links it has to other nodes within the network but also takes into account how well connected a node is, and how many links their connections have, and so on through the network. The centrality is computed using an adjacency matrix. Let *a* be the adjacency matrix where $a_{i,j}$ = 1 if node *i* is linked to node *j* and 0 otherwise. The relative influence of a node *i* is

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j \in G} a_{i,j} x_j$$

where *M(i)* is the set of neighbors of *i* and λ a constant. With some rearrangement, the eigenvector centrality is the solution to the eigenvector equation $Ax = \lambda x$.

Kendall Tau Measure

We define a measure using the Kendall-tau distance which counts the number of pairwise disagreements between the recomputed centrality ordering of the nodes and the original ordering. It measures the strength of disassociation between the two orderings. The higher the value, the higher the dissimilarity. For two orderings τ_1 and τ_2, the distance is calculated as

$$\sum_{i,j \in P} K_{i,j}(\tau_1, \tau_2)$$

where
-*P* is the set of unordered pairs of distinct elements in τ_1 and τ_2
-*i* is an element in τ_1 and *j* an element in τ_2
-$K_{i,j}(\tau_1, \tau_2)$ = 0 if *i* and *j* are in the same order in τ_1 and τ_2
-$K_{i,j}(\tau_1, \tau_2)$ = 1 if *i* and *j* are in the opposite order in τ_1 and τ_2
-$K_{i,j}(\tau_1, \tau_2)$ = 0.5 if *i* and *j* went from ordered to tied and vice versa in τ_1 and τ_2

Procedures

In our study, stability is measured by studying the distribution of the Kendall measure of each centrality measure on different datasets. We calculate the centrality of the nodes using a particular measure, and then leave each node out from the network, creating an ensemble of similar networks and recomputing the centrality of the nodes. Using the defined Kendall Tau distance, we count the number of pairwise disagreements to quantify the stability. We apply this procedure to various real and artificial random networks.

Application to Real Networks

The table below summarizes the characteristics of each dataset.

Name	Nodes	Nodes in Giant Component
Karate Club	34	34
Jazz	198	198
Netscience	1589	379
Hep-Th	8361	5835

Table 1: Description of Datasets

In our procedures, the giant component of the networks were considered. The giant component is the largest subset of a graph in the network, and contains a significant proportion of the entire nodes in the network and thus, a large amount of information. Because the number of nodes in the giant component is less than or equal to the number of nodes in the whole network, the computational time is greatly reduced.

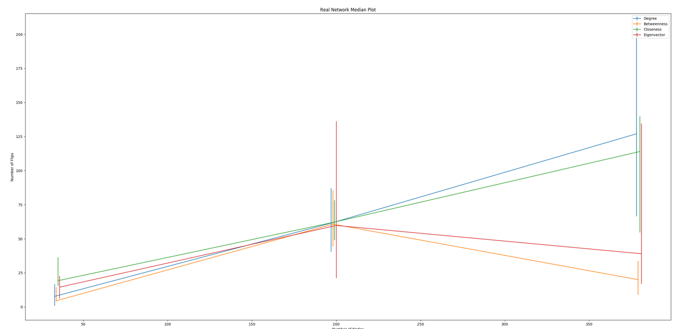

Figure 1: Real Network Kendall Tau Median Plot

Application to Random Networks

Small-World Networks

The small-world network is constructed as follows: first, a ring is created over *N* nodes. Each node is then connected with its *k* nearest neighbors (*k-1* neighbors if *k* is odd). Each edge connecting nodes *i* and *j* such that $i \neq j$, is then rewired to *i* and *u* with some probability *p* with uniformly random choice of *u*. This procedure is the Watts-Strogatz mechanism, and is used to construct the networks. Four different types of small world networks have been constructed. The probability *p* have been modified to study the effects it may have on the distribution. Each network contained 500 nodes sampled 20 times. The results are shown below:

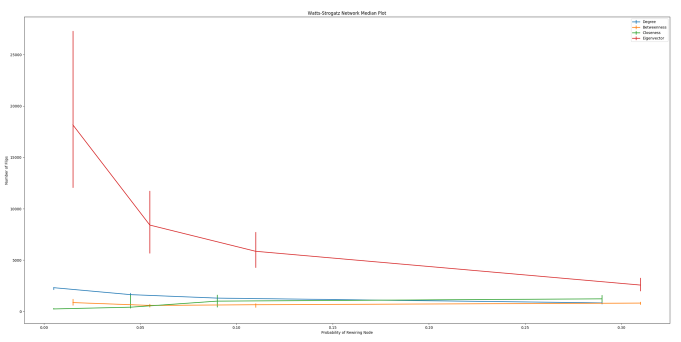

Figure 2: Small World Network Kendall Tau Median Plot

Preferential Attachment Networks

Preferential attachment networks rely on a process that depends on the number of connections a node has. Given a graph with *N* nodes, new nodes are attached with *m* edges that are preferentially attached to nodes with high degree. When new nodes are added to the network, each new node is connected to $N \leq N_0$ existing nodes with a probability that the new node is connected to some node *i* is

$$p_i = \frac{C_D(i)}{\sum_j C_D(j)}$$

where $C_D(i)$ is the degree at node *i* and $\sum_j C_D(j)$ the sum of the degrees over all pre-existing nodes *j*. The Barabasi-Albert mechanism is used to construct these networks. The results are shown below:

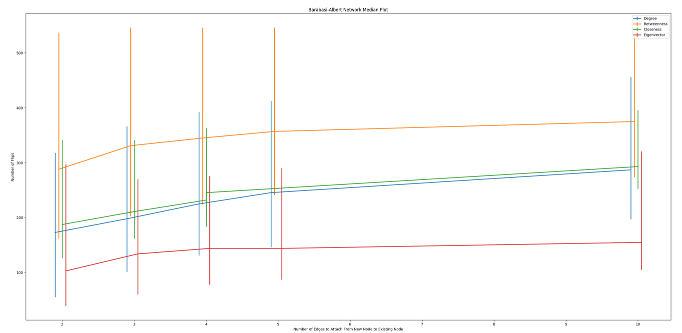

Figure 3: Preferential Attachment Network Kendall Tau Median Plot

Conclusions

- Betweenness Centrality was most stable for most real world networks, but least stable for preferential attachment models
- Eigenvector Centrality was least stable for small world networks, but most stable for preferential attachment models
- Degree and Closeness measures are stable at low number of nodes, but become increasingly unstable at higher number of nodes in real world networks

Acknowledgements

I would like to thank Professor O'Connell and Professor Anderson for giving me this opportunity to explore the realm of complex networks. In addition, I would like to thank the coordinators for allowing me to present my research at the Meeting of the Minds 2018.