**Lab 1b: Descriptive Network Analysis – Local and Global Network Properties**

***CompSci 396-0: Social Networking Analysis* *Win 2022***

Student Name: Jiaqi Guo NetID: JGR9647

* **Responses to Question**
  + **Part III: Individual Network Properties**

1. **(10 points) Provide a table ranking the top 5 nodes in your network on each centrality measure. Each centrality means (a) in-degree, (b) out-degree, (c) betweenness, (d) in-closeness, (e) out-closeness, (f) eigenvector, (g) Burt’s network constraint, (h) hub score, and (i) authority score**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ranking** | **in-degree** | **out-degree** | **betweenness** | **in-closeness** | **out-closeness** | **eigenvector** | **net constraint** | **hub score** | **authority score** |
| **Top 1** | 59 | 1 | 1 | 149 | 131 | 1 | 164 | 1 | 59 |
| **Top 2** | 210 | 193 | 247 | 59 | 185 | 105 | 175 | 119 | 53 |
| **Top 3** | 149 | 247 | 19 | 23 | 160 | 224 | 199 | 57 | 105 |
| **Top 4** | 23 | 53 | 218 | 210 | 46 | 59 | 284 | 247 | 224 |
| **Top 5** | 1 | 19 | 193 | 182 | 226 | 53 | 91 | 251 | 70 |

The **numbers** in the table represent the **indexes of nodes(identity)**. There are 290 nodes in this figure, corresponding to node indexes 1-290

1. **(10 points) Briefly describe each centrality measure. How is each computed and what does its number mean in your network (e.g., a high centrality score means…)?**
2. **in-degree**

In-degree centrality measures the number of edges others have initiated with a vertex.

A high In-degree centrality value indicated other nodes in the network have a high level of engagement with a node.

1. **out-degree**

Out-degree centrality counts the number of edges a vertex has initiated with others.

A high out-degree centrality value indicates a high level of engagement a node initiates with other nodes of the network community.

1. **betweenness**

Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph.

The betweenness of a vertex, v, is given by:

Where is the number of geodesics from to through . Therefore, a high-betweenness vertices usually lies on a large number of non-redundant shortest paths between other vertices. Which can be considered as “bridges” or “boundary spanners”.

1. **in-closeness**

The closeness centrality of a vertex is defined by the inverse of the average length of the shortest paths to/from all the other vertices in the graph.

The in-closeness can be considered as an index of the expected time-until-arrival for information in-flowing from other nodes to a certain node through the network via optimal paths.

1. **out-closeness**

The closeness centrality of a vertex is defined by the inverse of the average length of the shortest paths to/from all the other vertices in the graph.

The out-closeness can be considered as an index of the expected time-until-arrival for information out-flowing from a node to others through the network via optimal paths.

1. **eigenvector**

Eigenvector centrality scores correspond to the values of the first eigenvector of the graph adjacency matrix.

In general, vertices with high eigenvector centralities are those which are connected to many other vertices which are, in turn, connected to many others.

1. **Burt’s network constraint**

Burt’s network constraint is commonly used as a measure of structural holes.

Burt's measure of constraint , of vertex 's ego network , is defined for directed and valued graphs,

for a graph of order , where proportional tie strengths are defined as

are elements of and the latter being the graph adjacency matrix. For isolated vertices, constraint is undefined.

The larger the constraint value, the less structural opportunities a node have for bridging structural holes.

1. **hub score**

Hubs and authorities are a natural generalization of eigenvector centrality.

Let be the adjacency matrix of a directed graph. The hub centrality matrix is given by:

A high hub actor points to many good authories and a high authority actor receives from many good hubs. The hub score is proportional to the authority scores of the vertices on the out-going ties.

1. **authority score**

Hubs and authorities are a natural generalization of eigenvector centrality.

Let be the adjacency matrix of a directed graph. The hub centrality matrix is given by:

A high hub actor points to many good authories and a high authority actor receives from many good hubs. The authority score of a vertex is therefore proportional to the sum of the hub scores of the vertices on the in-coming ties.

1. **(5 points)** **How does the centrality of nodes vary with different types of centrality metrics? Why is this the case? Please offer some potential explanations using certain nodes as examples**

**Example node 1:** This node “[deleted]” has the highest out-degree, betweenness, eigenvector centrality and hub score. And its in-degree centrality is also the top 5 among the network community.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Node index（Name）** | **in-degree** | **out-degree** | **betweenness** | **in-closeness** | **out-closeness** | **eigenvector** | **net constraint** | **hub score** | **authority score** |
| **1** | 21 | 22 | 6378.9199134 | 1.604e-05 | 1. 896e-05 | 1.000000e+00 | 0.08152792 | 1.000000e+00 | 1.0221e-01 |

It is not difficult to find that the in-degree and out-degree centrality of this node makes it play a very important role in the network, which also means that this node can greatly influences the information transmission of network. Therefore, this node also owns a high betweenness centrality. The high overall degree centrality also means a high eigenvector centrality. In addition, as this node has a large out-degree, it is more likely to point out some authoritative information in the network, which makes it more likely to become a hub (a high hub actor points many good authories)

**Example node 2:** This node “Chickenman1964” has the highest in-degree centrality and authority score. And its in-closeness, eigenvector centrality is also the top 5 among the network community.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Node index（Name）** | **in-degree** | **out-degree** | **betweenness** | **in-closeness** | **out-closeness** | **eigenvector** | **net constraint** | **hub score** | **authority score** |
| **59** | 42 | 0 | 0 | 2.112e-05 | 1.193e-05 | 4.288355e-01 | 0.06452707 | 0.0000e+00 | 1.00000e+00 |

This node has the largest in-degree centrality, but its out-degree centrality is 0, that is to say, for this node, the information flow is one-way, which makes this node cannot affect the information transmission of rest parts of the network. In other words, the node’s betweenness centrality is absolutely 0. And, based on the definition of authority score, the highest in-degree centrality makes it capable to receive information form many good hubs and give it the highest authority score. Besides, its high in-closeness centrality also enables it to receive information efficiently.

* + **Part IV:** **Global Network Properties**

1. **(3 points) Briefly describe (a) what k-core is, (b) what insight this k-core decomposition method provides, and (c) what is the highest/maximum level, k, of cores present in your network (e.g., Do any 3-cores exist in your network? Do any 4-cores? 5-cores? etc.)?**
2. **(3 points) Visualize your network using k-core decomposition and include the visualization in your report. In a paragraph, discuss your interpretation of the visualization and whether the results of k-core decomposition make sense based on your expectations of the network.**
3. **(3 points) Pick one of community detection algorithms to run on your network. Which community detection algorithm did you choose and why?**
4. **(3 points) How many communities have been created? For your network, what might a community of nodes potentially have in common?**
5. **(3 points) What is a modularity score? Interpret the modularity score of your results of community detection?**
6. **(3 points) Plot the communities and include the plot image in your report. What information does this layout convey? Are the communities well-separated, or is there a great deal of overlap? Describe the actors between any components and cliques (i.e., brokers). What are common features of these actors?**
7. **(3 points) Present and interpret the in- and out-degree distribution based on your network as well as a log-log plot. Compute and interpret the estimate of the c slope (i.e., alpha value). Note that a p value (KS.p) less than 0.05 indicates the empirical data doesn’t fit with the power-law distribution.**
8. **(3 points) Present in a plot the observed and simulated values for each average path length and clustering coefficient based on the original network and 1,000 randomly shuffled networks.**
9. **(3 points) Based on these data would you conclude that the observed network demonstrates small world properties? If so, why? If not, why not?**
10. **(3 points)** **In two or three paragraphs, discuss your major findings of your network based on all the analyses you’ve done in this exercise and also your own additional analysis if necessary. Your answer here will be evaluated based on depth and comprehensiveness. Thus, you’re encouraged to utilize extra information to answer this question. For instance, you can take a look at your original data (i.e., “twitterData,” “youtubeData,” or “redditData” if you work with the provided R code) in R. These data frames include additional user, text, and time information for your network. Similarly, if you need more insights from your network, feel free to run correlation and regression analysis based on your data collection**