Introduction

In the wake of the Big Data movement, network analysis was able to expand beyond small world networks and begin studying networks with millions of nodes and billions of edges. With the increase in popularity and complexity of network analysis, the outcomes (i.e. the visualizations) have also correspondingly increased in use and complexity. The following research intends to study the inherent interpretability of networks, and whether or not researchers who utilize this medium need to rethink how they approach the presentation of network graphs.

Network analysis provides a unique approach to studying the dynamics and relations between objects or entities. The corresponding visualizations provide a seemingly robust, yet complex interpretation of said relationship. In Cairo’s *The Functional Art* (2013, p. 120), he outlines the hierarchy of perceptual tasks according to accurate versus generic judgements. With the amount of complexity in network visualizations, most network visualizations focus on the more generic portion of the hierarchy with color or node size. However, it may be worth thinking through methods for more exact comparisons without sacrificing complexity or intuitiveness. We hypothesize that networks are not inherently explanatory, and further visualizations should focus on determining methods for more exact comparisons.

Methodology

Using Gephi’s random network generator, we constructed a random 50 node network with a wiring probability of 0.07. This entails that there is a seven percent chance for a node to be connected to any other node. This produce a network that was small enough to make quick interpretations, yet large enough to not inherently self-explanatory. We then modify node color and size depending upon one of three network metrics of centrality: degree centrality, betweenness centrality, and closeness centrality. Degree centrality is the number of ties or lines that connect this node or dot (high degree refers to a node with many ties; low degree refers a node with few ties). Betweenness centrality is the number of times a node bridges two other nodes (high betweenness refers to a node that is a broker to many other nodes; low betweenness refers to a node that is rarely a broker). Closeness centrality is the average distance or path any node is from any other node (high closeness refers to a node that has the shortest path to any other node; low closeness refers to a long path to any other node). Each node in the network was given an identification number, and each network was identical except for the color or size of the nodes depending upon which metric it corresponded with.

Participants began with an unmodified version of the network and asked to select the node that corresponded with the highest and lowest value of the given metrics. They were presented with the above descriptions of the network metrics. Depending upon the answer participants gave for the first question, they either received the color version of the networks or the node size version of the networks. I was under the assumption that this would be a method that would allow for randomization without succumbing to selection bias, however, I was wrong which will be discussed further down. Participants were then shown a network with either the color or node size to represent the given network metric for a grand total of four networks.

Results/Discussion

Before reporting the results of this study, it would seem essential to review the limitations and shortcomings of the following study, which there seem to be a lot of. First, it would seem that the instructions and networks may not have been as intuitive as I initially thought. My networks were missing legends which could easily affect the interpretability, and the task that I was having participants engage in could easily be misunderstood if the person does not completely understand the network terminology used. The next obstacle that hinders the results is the population on mTurk. This population does not seem to seriously regard the tasks, at least the tasks that I presented them. A number of individuals simply gave one answer or two answers throughout the course of the study. My favorite response came from a participant who in the last two questions typed “FUC YOU” rather than the three-digit node ID. There were a few participants who took the survey seriously, but it would seem that a good percentage did not. With only a sample size of 35, each result is going to sway the end analysis. The final limitation of this study comes in the form of my randomization technique. There were only three participants who were shown the color version of the questionnaire. All other participants selected the node size version. This severely hurts the interpretation and analysis of these results.

Due to the limitations, as cited above, we decided to refrain from doing a complete analysis of comparing the three different interpretations across the three-network metrics. Instead, we focused on comparing the network metric of high and low degree centrality between the unmodified network and the node size network. Due to the inherent repeated measures design, we treated the unmodified network as the pre-test and the node size network as the post-test. We utilized signal detection theory to code participants’ responses (1 if hit, 0 if miss). From here, we compared the proportion of hits between the pre- and post- measures by using two paired t-tests (pre-high degree centrality and post-high degree centrality and pre-low and post-low) and two independent samples t-test (pre-high and pre-low and post-high and post-low). There were no significant t-values, and the proportion of hits were the exact same for both pre-test and the same for both post-test. Although, from pre- to post-, we do see about a 5% increase in hits.

One potential interpretation is that these channels that we use provide no additional value to readers. Networks seem to be naturally incoherent, and the channels we use to provide insights and extrapolations only provide a slight benefit. However, due to the complex nature of network analysis, the limitations of the study, and the respondents of mTurk, these generalizations are weak at best. Among those who did take the survey seriously, they amounted to highest proportion of hits. The strongest claim that we may be able to pull out of this study is simply that those who pay serious attention to visualizations make the most accurate interpretations. Future analysis of this data could run more robust analyses on the different factors and dynamics of the intended study. However, the limitations heavily demotivate from any further use of this data outside of this assignment.