Network Analysis

Digital Humanitites Tutorials

A tutorial presented for the
Willson Center Digital Humanities Lab

The University of Georgia
www.digi.uga.edu

Caleb Crumley B.S.
crumleyc@uga.edu

1 Introduction

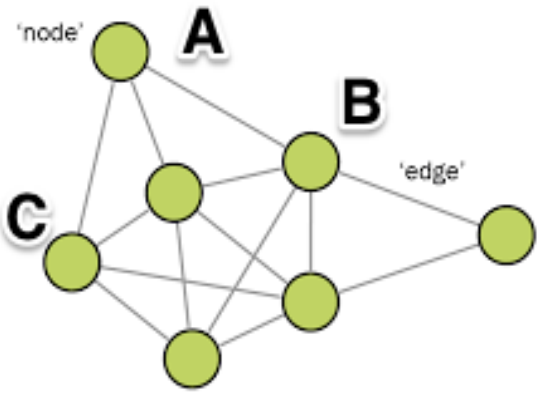
Visualizations can bring a new perspective to one's research. They are good for a plethora of reasons, which range from giving your presentation a nice face-lift to being a key tool in research. Graph Theory is the overall study of mathematical structures that are used to model reciprocal relations between objects. This article will specify on the use and benefits of the analysis of Networks, which is a graph that illustrates the flow of interactions and relationships. Using this resource with your research will allow you to answer certain questions like who, or what, is the most influential character? Who knows whom? How are particular characters are related? An object can be related to multiple objects in different ways. Where did this particular strain, gossip or disease, originate from? How artistic and social movements grow and develop. These are just some of the questions that could be answered through network analysis.

I highly suggest you look up "network analysis" in any academic studies archive. In doing that, I have found research on modeling the brain activity while sleeping and how is it processed, tracking the flow of knowledge in economic geography research, and the most common one I have come across, tracking STDs and modeling the path through networks. Look into more specific research using network analysis and it is apparent that network's applications are widespread, and by doing this, it will provide other outlooks and applications of network analysis.

2 Theorems

Diving into the terminology of networks, there exist two abstract terms, nodes and edges. I simply mean that these two terms can take on multiple forms and represent the same nodes in different ways. Nodes are connected by edges, which illustrates the relationship between the two of them. Nodes can represent anything, in a biological process or a stack of literature. Anything that could have an interaction with a similar type of node would be appropriate for a network. The edges could also be represented in different ways. They could be assigned a different colors or weights that would show a particular type of collaboration between the two. For example, the edges' color can be coordinated by the social movements or other significant time periods. The weight illustrates how much interaction there exist between the two nodes, where interaction can be any form of communication depending on the application of the network. The heavier the weight directly correlates with the frequency of communication.

Figure 1: A basic visual representation of a network

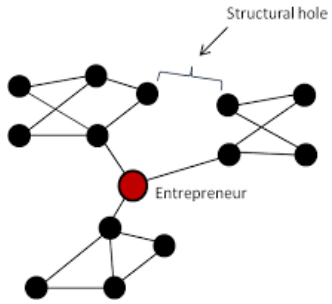


2.1 Weak Tie Theorem

Now to dive in deeper into the more complexities of a network graph. It is significant to observe certain types of behavior between the nodes. The weak tie theorem states that the heavier the weight between the two nodes will

directly correspond to overlapping between adjacent nodes. This theorem makes sense because having a stronger tie with another node suggest a more frequent connection between them. Thus the nodes are bound to have one or more related nodes as well. For example, let us say that I represent node A and my sister represents node B in Figure 1. Since we are family it is safe to assume that the weight between us will be fairly strong compared to an additional node C that could represent someone I hardly interact with. Given that the edge between A and B are strong, the theory suggest that A will most likely interact with some of B's interactions and vice versa.

Figure 2:

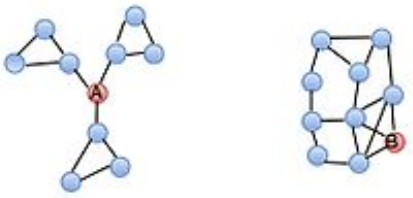


2.2 Structural Nodes Theorem

It is also significant to observe the inactivity between two or more groups of nodes, which are called structural nodes. This hole is shown by the white space, or background, between the two groups, as shown in Figure 2. The node that bridges smaller groups, such as in the figure, is called the entrepreneur. There are some structural advantages to being the entrepreneur. The main one being the sole connection between the groups, thus communication travels through it. To contextualize the entrepreneur, let us think of him as a district manager who coordinates over three different branches within his district. The branches can be seen as the individual groups of nodes that do not communicate. Thus the manager has complete control over the communication between branches. In certain contexts, that can be a powerful position.

Illustrated in Figure 3, it is obvious based upon structure that A has the advantage when compared to B, although both networks have the same number of edges. If B were to attempt to keep information from a particular node or group of nodes, there still exist a potential connection containing that bit of information. Behaviors such as these are useful to look into while analyzing specific regions of the overall network. Entrepreneurs are useful to examine due to the success in forming organizations. Does the entrepreneur have any distinct characteristics that allow him to be successful? What types of organizations has he formed, and how does that relate to his overall character? Refereeing back to the district manager, it says a lot about him and his ability to manage based upon how his individual braches are preforming. If they are all successful branches, the manager may possess leadership skills. Not only do the node/entrepreneur's direct actions dictate how it can be perceived, interactions and associations can too.

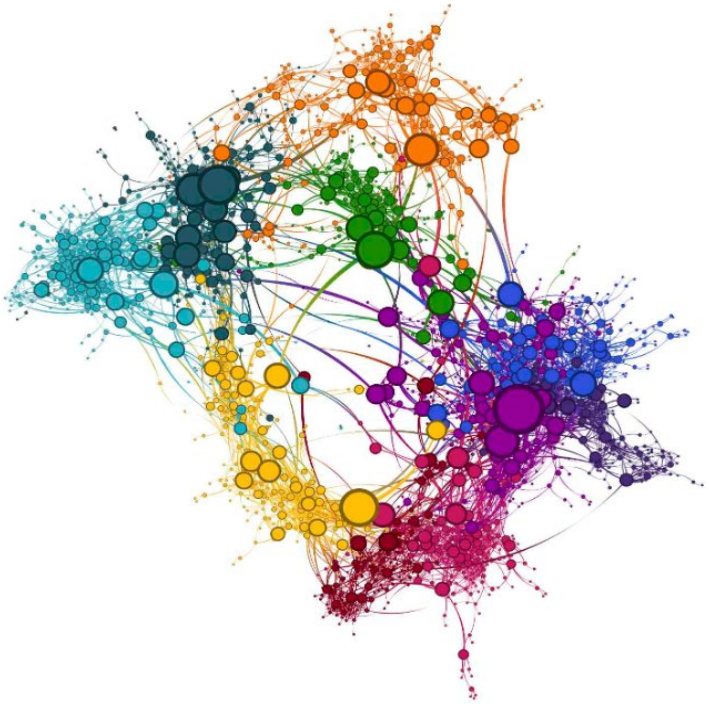
Figure 3:



3 Centrality

It is important to see who or what is the most important and influential in the entire network, this is called centrality. This is why network analysis is so applicable in humanities research. They can help in tracking the origin of certain relationships and social developments. In respect to network analysis, it is necessary to differentiate the two meanings of the word 'important.' The first definition is the transfer or flow throughout the network. Within this approach, it is good to record the behaviors , such as information control and what the node's affiliations says about itself, that can relate to the structural nodes. Studying the flow of communication can be helpful to study gossip, something that is passed from one node to the next, and so on. Gossip can be referred to as serial duplication. Parallel duplication is the broadcast, or duplication, to multiple nodes at one time, example being a radio broadcast. The other definition is to analyze the cohesiveness of the network. To examine the weak ties and study how the nodes interact collectively and independently. The group can be connected by weak ties thus meaning a group that hardly communicates, or vice versa.

Figure 4:



While observing the importance of nodes, it is good to note the ability to color-coordinate the nodes and edges. This option can add another visual tool to network analysis. Without color, the main take away is the flow of relationships and a map of the communication. Implementing color, whether on the nodes, edges, or both, can show a particular era or process that the communication occurred. Looking at Figure 4, it is evident that the green section is the most influential since it is the most central and connects to most, if not all, subnetworks. All nodes in that color have a similar relationship that is associated in that color. One could also study the structural nodes and other types of behavior within each color. Along with the interpersonal connections, it is also significant to study the bridges between to separate colors. It might be useful to study the significance of the each group's placement. Is it significant that the blue and purple groups are practically overlapping or that the pink and orange are on opposite sides. What can the placement say to the group it's self and the surrounding groups?

4 Conclusion

The unique characteristic about networks are that they can be examined as an entire graph, or a specific region that can be treated as it's own network. The theorems can apply to any scale of the network itself. It is good to answer the questions depending on the scale of the network. Looking into the flow of knowledge, gossip, behaviors, and how social movements grow and their development are important to dive into for research. Network analysis is an applicable research method to grasp all aspects, even the ones not mentioned.

4.1 Suggested Readings

- How to use Palladio and Gephi, two network analysis tools, and how to extract the neccessary data to implement them.

  https://programminghistorian.org/lessons/creating-network-diagrams-from-historical-sources

  https://historyblogger.net/2013/08/17/getting-started-with-gephi/

- More applications and resources on network analysis

  http://science.sciencemag.org/content/323/5916/892.full

  https://github.com/briatte/awesome-network-analysis

- For a more advanced reading "Networks, Crowds, and Markets: Reasoning About a Highly Connected World" by David Easley and Jon Kleinberg. It probably goes into more detail than needed, but it has some good illustrations and dives deeper into the theorems mentioned in this tutorial. The pdfs of each chapter are all posted online on Cornell's computer science's website.