Math 168: Short Essay 2
Dhruv Chakraborty

Fall 2020: Prof. Mason Porter

October 19, 2020

The Group Project

Data Sets

I'm interested in applying network analysis techniques to a variety of data sets in order to gather insights not only about said data sets but also about the techniques being used. A key data set I'd be interested in looking at to build our analysis off of would be a social network of UCLA students. There are many ways we could go about getting this data by using various social media websites: Twitter, LinkedIn, Facebook etc. Facebook and LinkedIn do not have strong APIs (to fetch the data) that are easy to use or setup, so we will probably work with the Twitter API, which provides us with a number of free requests. If collecting data through such a process proves to be impossible to due some unforeseen blockade or is just generally too cumbersome then we can also continue by analyzing one of the many available Twitter follower-following networks such as this one by Kwak et all (2010)[1].

There are many other interesting data sets I would like to analyze after first solidifying a plan of action with this first data set. In particular, one thing I want to consider is how concrete the 'organizational split' Newman makes in "Networks"[2], i.e. differentiating between social, technological, biological and information networks, actually is. We hope to compare our analysis techniques on networks in each of these categories and see how results differ. It will be especially interesting to see how implicitly defined network structures will compare to more explicit networks.

A potentially separate application that I'm also interested in lies between the intersection of Natural Language Processing and Network Analysis. Specifically I want to look at anomaly detection on an implicitly defined similarity based network structure of tweets or some other body of text. For example, we could assign weighted edges between tweet nodes based on "how similar" the tweets are, presumably getting this score from a separate algorithm. We would then analyze this network with many techniques, with a focus on ones with an anomaly detection bend.

Theoretical Interests

In Chapter 7 of Newman's "Networks"[2], he spends the first few sections talking about various centrality measures in network analysis, how they're calculated and the various features of nodes they end up highlighting and ignoring. He mentions several alternative centrality metrics that are not used enough in practice, or simply haven't been developed enough. In particular there's the version of Closeness Centrality calculated using the Harmonic mean to circumvent issues surrounding how nodes in smaller components have higher closeness centrality than they should compared to more connected nodes in larger components. There are also alternate notions of betweenness centrality, specifically flow betweeness and random-walk betweeness to conteract weighing only shortest paths (as information isn't always necessarily passed through these paths).

Especially interesting to us, however, is the concept of cores, developed by relaxing the strong fully-connected restrictions on cliques. k-cores are defined to be sets of nodes where each is connected to at least k other nodes in the set. Another thing that makes k-cores interesting to look at is that they're easy to find - simply by removing nodes in the set that have degree less than k recursively until all nodes in our set have degree at least k. This process can be applied to a network to discover it's core-periphery structure by increasing k, resulting in an almost onion-like decomposition. This leads to a notion of centrality with nodes in the highest k-cores serving as "core actors" in the network. We wish to understand the notion of core centrality and further develop it by comparing it to other existing versions of centrality.

Teammate Preferences

I would like to work with my friend Daniel Luo on the group project. If more members are required, we're happy to work with anybody else from the class as well.

References

[1] Kwak, Haewoon and Lee, Changhyun and Park, Hosung and Moon, Sue. *What is Twitter, a social network or a news media?* `http://doi.acm.org/10.1145/1772690.1772751`, 2010.

[2] Newman, Mark E.J. *Networks, Second Edition*. Oxford University Press, 2018.