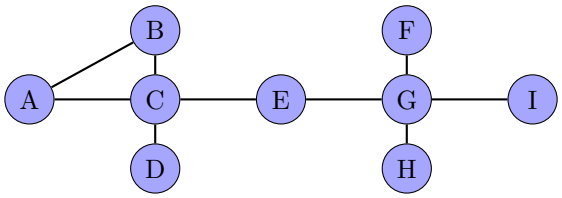
Modeling With Networks

Philip Hossu, Paolo Ratti Tamayo, Jiateng Sun

April 6, 2018

1. Introduction

Graphs are some of the most versatile and interesting structures in mathematics. A graph G is characterized as an ordered pair of vertices and edges, $G = (V(G), E(G))$. The edges in a graph imply the existence of a relationship between two vertices or objects. They can be directed (implying a relationship based on ordering) or undirected where the edges have no orientation. Edges can also be weighted, having several real world applications like distance, cost, etc. These structures evidently can represent countless applications which have discrete and related objects. A simple graph which will operate as our running example is shown below.



Influence is the capacity to produce an effect in indirect or intangible ways. If one wants to be successful in their field, they must share their ideas and persuade their peers to accept them. This idea is present in all types of industries and specializations. A significant portion of this project focuses on the widespread influence of a famous mathematician, Paul Erdos, and his large co-author network.

In the following project report, we analyze how influence can be determined in a graph and describe our exploration into mathematical modeling with networks. There exist several known measures to explore the importance of a vertex in a graph, including degree centrality, betweenness centrality, etc. We will model various networks, analyze them, find who/what has the most influence in the network, and the reasons behind this result. We seek to combine real world scenarios with some of these mathematical ideas to accurately describe the concept of influence in graphs. Furthermore we will argue how these methods can show steps to rapidly boost ones own influence.

The prompt asked us to consider a number of specific tasks, briefly outlined below:

- Build and analyze the Erdos1 co-author network using the data from https://files.oakland.edu/users/grossman/enp/Erdos1.html. The Erdos1 co-author network is characterized by the authors who have collaborated with the famous mathematician, Paul Erdos. Rather than including every entry in this data file, we were asked to consider a network where each author has collaborated directly with Erdos (omitting Erdos himself). After building this network, we are asked to explore the graph and find some interesting features and properties.
- Define and study two critical measurements by which to determine the influence of authors in the network we created in part a.
- Gather data, build, and analyze a network showing the relationship between some foundational papers in the emerging field of network science. Apply the influence measures used for parts a,b and discuss their effectiveness. Also, discuss methodology and other factors surrounding this network.
- Gather data, build, and analyze a real life scenario which can be modeled as a network. Again apply influence measures and discuss the results, shortcomings, external factors, etc..
- Discuss how influence and impact can be used in real life situations. Consider business decisions, improving influence, selecting a graduate school etc.

To address the above tasks, we utilized the R programming language and our knowledge from various math and computer science courses.

2. Statement and Analysis of the Problem

While our project does have a number of specific requirements discussed in the introduction, essentially the problem we are trying to approach is how to measure the influence of a node within a network. We define two influence measures to address this problem, discussed in depth in the latter sections. Once these are defined, we are tasked with collecting and cleaning data, creating multiple networks, and analyzing these networks under our two metrics.

While the field of graph theory is certainly not young, we noticed a wealth of more recent works geared towards our topic and general social network analysis. We began by finding a modern survey of various graph centrality measures, notably the work by Bloch, et al. titled *Centrality Measures in Networks* [NUMBER]. Rather than focusing on a concrete example, this work provides a much needed overview of several fundamental graph measurements and their properties.

Another work which we used to help shape our understanding of the problem was the comment released by Vishesh Karwa and Sonja Petrovic titled *Coauthorship and citation networks for statisticians* [NUMBER]. This paper looks directly at how co-author networks and citation networks can be analyzed which is very relevant to our project. Regrettably, the networks which Karwa and Petrovic are commenting on are somewhat more advanced than the networks we were able to make. Consequently, their analysis is more advanced. As we will discuss, our Erdos1 co-author network is slightly oversimplified. An edge in our network simply implies the presence of these two authors collaborating at least once, but the exact number is unknown. Regardless, we were able to take a number of pointers from this paper. For a complex network, the paper suggests that a metric like counting the degree of each node is far too simple. For our simple network, such a measure may be more appropriate.

A number of publications also exist which detail analyses of specific co-author networks. One such work is *Co-authorship and citation networks in Spanish history of science research* by Osca-Lluch, et al [NUMBER]. The authors data allows them to look at papers published by various journals, number of papers published by Spanish scientists over time, types of publications, etc. While we did not have this type of data readily accessible, the paper also looks at co-author collaboration and the impact of these publications. The paper interestingly notes that a very high percentage of authors have no collaboration whatsoever. They create a visualization of their graph where the edge thicknesses represent number of collaborations. To analyze the publication impact, the authors weigh in on using a quantative measurement vs a qualitative one. They argue that a qualitative approach can be effective due to its ability to take into consideration the expertise, reach, and prestige apart from the publishing journal.

Despite that our networks were based on much more simple data sets, it was interesting and useful to see how others have approached similar analyses of co-author and citation networks.

3. Description of the Model

4. Analysis and Testing of the Model

5. Results and Quality of Model

6. References

7. Appendix