

Math22 Final Project

Andy Li
`andy_li@college.harvard.edu`

Ethan Tran
`etran@college.harvard.edu`

Fall 2021

Contents

1 Overview

Networks are a common graph structure in mathematics with vertices connected by edges, usually with the purpose of visualizing relationships between objects. The use of network graphs have been extended into other fields, and this project will focus on the use of networks in sociology. Here, networks have nodes that represent agents, which are connected by links (edges) that signify some social interaction or tie between the two agents. Further detail comes in the form of node degrees (number of connections) and strength (weight of connections) to represent networks (Porter, 1082).

A large component of networks lies in communities, which are densely connected subgroups of nodes that in the context of sociology represent extensive human interaction. These communities are discovered through community detection algorithms, which have important real world uses such connecting people with common interests over social media. Even with the wondrous applications of community networks in sociology, they are still built on the fundamentals of graphs which require linear algebra for construction. Thus begs the guiding question: How does linear algebra assist in community detection and community network graphing?

2 Graph Theory

2.1 What is Graph Theory

Graph theory is the study of graphs, which are mathematical structures composed of nodes and edges. In the most common definition of a graph $G(V, E)$, there is a set V of nodes/vertices/points, and set of edges E such that $E \subseteq \{\{x, y\} | x, y \in V \text{ and } x \neq y\}$ (Bender and Williamson 7). Various modifications can be applied to this definition to create different types of graphs. For example, directed graphs, in which edges have orientation to and from certain nodes, have different definitions of edges E compared to undirected graphs, which follow the definition above.

2.2 Adjacency Matrices

We will define adjacency matrices, which are common algebraic representations of graphs in Graph Theory (Biggs 7).

Definition 1: The adjacency matrix of a network is the $n \times n$ matrix A whose entries a_{ij} are given by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent, or connected by an edge from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$$

An example adjacency matrix is shown here:

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Adjacency matrices can be easily used to construct graphs (both directed and undirected) and vice versa. We can note a few characteristics of adjacency matrices. Firstly, for undirected graphs, the adjacency matrix will be symmetric. This is because if two nodes are connected, then both a_{ij} entries in the matrix will be 1 as direction doesn't matter, and 0 otherwise if they are not connected. On the other hand, for directed graphs, adjacency matrices are not necessarily symmetric. As the a_{ij} entry will only be 1 if there exists an edge from i to j , if there is not a edge going in reverse then the "symmetric" value in the matrix will have a value of 0. Usually, the diagonal values of an adjacency matrix will all have the value 0, as most graphs will not have nodes that are connected with themselves.

Adjacency matrices become more complicated in the environment of community networks. Instead of using 1 to represent the presence of a edge between two nodes in the matrix, the nonzero value can be greater or less than 1 to represent the weight of edge. In community networks, this weight represents how close two agents relate to or interact with each other, and is critical in community detection algorithms, which will be elaborated on below (Porter 1086).

An example adjacency matrix is shown here:

$$\begin{bmatrix} 0 & 3 & 2 \\ 1 & 0 & 1 \\ 2 & 3 & 0 \end{bmatrix}$$

3 Community Detection

3.1 What is Community Detection?

Community Detection is the process by which researchers can identify groups within a network in which individuals' membership in particular groups are yet to be ascertained. Community Detection is performed by researchers to map out the relationships between isolated nodes, members, or objects within the community. By gathering and ordering network members by the strength of their relationships to each other one can discover communities—otherwise known as groups, clusters, cohesive subgroups, or modules. This information can yield valuable insights. For example, community detection can reveal the key figures and leaders within a group, around which other members gather. The results of community detection are often represented through graphs of nodes and edges, and are generated through the application of algorithms to networks as represented by adjacency matrices.

3.2 CD Algorithm

There are many different algorithms used for community detection on graphs, each with their own approaches and required data structures. For the purposes of this project, we will focus on an algorithm that utilizes the eigenvectors of matrices. Specifically, this algorithm tries to find densely connected subgraphs in a graph by calculating the leading non-negative eigenvector of the modularity matrix of the graph (Chandal, Kabre, Aggarwal 66).

This algorithm focuses on modularity, which is a measure the strength of division of a network into modules (also called groups, clusters or communities). The modularity matrix of the graph is an important aspect of this algorithm to detect communities.

4 Sample Community Network

In this simple example, we have interviewed five people about who they eat with most frequently. They will rank their peers from 1-4, with 1 representing the person they would be least likely to eat with and 4 representing the person they would most be likely to eat with. On a five by five matrix, with zeroes or ones on the diagonal depending on if a person eats with themselves or not, we will convert these values to an adjacency matrix. This matrix will display the strength of connection according to how often they eat with someone. We will proceed to generate a directed graph with the use of this data and the CD Algorithm.

References

“Communities in Networks” by Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha

”An Introduction to Community Detection” by Lei Tang and Huan Liu

”Community Detection Algorithms” by Thamindu Dilshan Jayawickrama

”Algebraic Graph Theory” by Norman Biggs

”Lists, Decisions and Graphs. With an Introduction to Probability” by Edward A. Bender and S. Gill Williamson

”Community Structure Detection” by Amar Chandole, Ameya Kabre, and Atishay Aggarwal