**Report**

**Social Network Analysis of Physicians**

**TEAM 3:**

Rohith Reddy Poreddy(811254545), rporedd1@kent.edu

Gopi Cherukuri(811286984), gcheruku@kent.edu

Sai Prasanth Nitta(811297280), snitta@kent.edu

Sumanth Reddy Reddam(811286063),sreddam@kent.edu

Krithika Reddy(811281744), kpabbath@kent.edu

**Introduction:**

In the realm of healthcare, the established norm is for physicians to issue written directives that formally refer patients to other healthcare providers for specialized care or additional medical attention. This customary practice is pivotal in ensuring comprehensive and coordinated patient treatment. Imagine this intricate network as a social graph, where every node symbolizes a distinct physician within the healthcare system. The edges connecting these nodes signify the crucial linkages created through patient referrals. Essentially, each connection encapsulates a pathway through which medical expertise and patient care are extended beyond individual practitioners.

Within this social network of healthcare professionals, the nodes not only represent individual doctors but also embody the wealth of medical knowledge, experience, and specialization that each brings to the collaborative table. The edges, indicative of referrals, serve as conduits for the seamless flow of patient information, facilitating a networked approach to healthcare delivery. This interconnected structure ensures that patients receive a continuum of care, benefiting from the collective expertise of a diverse array of medical practitioners. In essence, this social network of physician nodes and referral edges underscores the collaborative nature of modern healthcare, where effective communication and shared patient responsibility are paramount for delivering optimal medical outcomes.

**Problem Statement:**

To quantify the level of connectedness or influence among physicians, one can employ network analysis techniques on referral data. Building a referral network where physicians represent nodes and referrals represent edges allows for the measurement of centrality metrics. Metrics like degree centrality (number of connections), betweenness centrality (intermediary influence), and eigenvector centrality (connections to well-connected physicians) can quantify the influence of specific physicians.

Identifying the most highly connected physician involves finding the node with the highest centrality scores. This physician is likely a key influencer in the network. Additionally, detecting groups of physicians who refer to each other frequently can be achieved through community detection algorithms. These algorithms identify clusters or communities within the network where nodes have denser connections among themselves than with nodes outside the community. Physicians within the same community are more likely to have frequent referrals, indicating a cohesive group.

By employing these network analysis techniques, healthcare administrators and researchers can gain insights into the referral dynamics among physicians, identify key influencers, and uncover patterns of collaboration within the medical network.

**Related Works:**

1. Social Network Analysis in Healthcare is a comprehensive area of research that explores the application of social network analysis to healthcare networks, including physician networks and patient referral networks.
2. Various community detection algorithms have been developed for identifying groups or clusters of nodes in social networks. This can help in identifying physician communities in the dataset.
3. Various online resources on networkX for graph analysis and python for programming.

**Background:**

**Insights of the dataset:**

The chosen dataset, representing a social network of physicians and their referrals, provides a valuable opportunity to gain insights into the referral patterns and relationships within the medical community.

Dataset: <https://networkrepository.com/soc-physicians.php>

1. Key Influential Physicians: Identification of physicians with the highest degree of centrality, indicating their level of connectedness or influence in the network.
2. Physician Communities: Detection of groups of physicians who refer patients to each other frequently. This can help identify specialized medical communities or clusters within the network.
3. Referral Patterns: Understanding the patterns of referrals between physicians, which could potentially highlight collaboration and specialization in different medical domains.

**Tools & Technologies:**

1. Programming Language: Python
2. Tools: Networkx, Jupyter Notebook.
3. Libraries for Data Analysis & processing: Pandas, Matplotlib
4. Libraries for Visualization: Matplotlib, Seaborn, D3.js

Python, NetworkX, pandas library, and Jupyter Notebook constitute a powerful ensemble of tools widely employed in data analysis, network modeling, and interactive computing. Python, a versatile programming language, serves as the foundation for this stack, providing a user-friendly and expressive syntax.

NetworkX, a dedicated Python library, specializes in the creation, manipulation, and study of complex networks or graphs. It offers an extensive set of functions for network analysis, facilitating tasks such as node and edge creation, centrality computation, and community detection.

Pandas, another essential library, excels in handling structured data through its DataFrame data structure. It enables seamless data manipulation, exploration, and analysis, making it indispensable for organizing and processing datasets in various formats.

Jupyter Notebook, an interactive computing environment, enhances the workflow by allowing the creation and sharing of documents containing live code, equations, visualizations, and narrative text. It supports Python and other languages, providing an ideal platform for data scientists and researchers to conduct analyses collaboratively.

**Basic Programming Procedure:**

1. Data Acquisition: Download the dataset from the provided source and load it into a pandas Data Frame.
2. Data Cleaning and Preprocessing: Prepare the data by handling missing values, if any, and ensuring it's in the appropriate format for network analysis.
3. At this step various types of analysis are performed.
4. Network Creation: Utilize NetworkX to create a graph representing the social network of physicians and their referrals.
5. Centrality Analysis: Calculate degree centrality and betweenness centrality for each physician node to identify influential physicians.
6. Visualization: Create visualizations to represent the network, highlight influential physicians, and illustrate physician communities.

**Algorithms:**

There are many algorithms in social network analysis but then these are the major algorithms which we used are centrality analysis:

* It is a set of measures used to identify the most important nodes in a network.
* Nodes are represented individually such as websites or other units.
* Edges represent connections and relations between them.

These centrality measures are useful in various applications, such as social network analysis, transportation network analysis, and communication network analysis.

Selecting centrality measure depends on the specific characteristics and goals of the network being analyzed.

Most common measures are.

1. Degree centrality

* Degree Centrality: It is one of the simplest measures. It depends on the number of links and edges a node has. Everyone is represented as a node, and connections between individuals are represented as edges. Nodes with a high degree centrality are considered central or influential in the network.
* Degree centrality is a measure used in graph theory to quantify the importance or centrality of a node within a network.
* Degree centrality provides insights into the popularity or prominence of individuals based on the number of connections they have.
* Everyone is represented as a node, and connections between individuals are represented as edges.
* Nodes with a high degree centrality are considered central or influential in the network.

1. Betweeness centrality

* It quantifies the extent to which a node serves as a bridge or intermediary between other nodes in a network.
* Nodes with high betweenness centrality play a crucial role in maintaining connectivity and facilitating communication between different parts of the network.

**Methodology:**

Step 1: Import all the required packages and load the dataset into the current working python script.

A screenshot of a computer

Description automatically generated

Step 2: Build a graph from the dataset using network.

A screenshot of a computer

Description automatically generated

Step 3: Drawing visualizations of the social graphs of physicians using matplotlib.

A screenshot of a computer

Description automatically generated

Improved Visualization with labels.

A screenshot of a computer

Description automatically generated

Improved Visualization with labels and custom color.

A screenshot of a computer screen

Description automatically generated

Step 4: Degree of Centrality

A screenshot of a computer

Description automatically generated

Step 5: Most Influential Physician

A screenshot of a computer

Description automatically generated

Here the most influential physician is node 127.

Step 6: Most Important Connection

A screenshot of a computer

Description automatically generated

The Most Important Connection in physician network is node 127.

Step 7: Betweener Centrality

A screenshot of a computer

Description automatically generated

Node 15 is the best betweener.

Step 8: Building a sub group

A screenshot of a computer

Description automatically generated

**Results:**

Outcome of the project:

1. Improved Visualization: Enhanced visual representation of physician groups and their referral relationships within the network.
2. Identification of Key Physicians: Identification of physicians with the highest centrality, signifying their influence in the network.

We have found that node 127 is the most influential and most important connection and node 15 is the best betweener.

1. Discovery of Physician Communities: Detection of physician communities that frequently refer patients to each other, potentially revealing specialized medical domains.