

A Work Project, presented as part of the requirements for the Award of a Master's degree in
Business Analytics from the Nova School of Business and Economics.

DISSECTING THE PRIMARY TO SPECIALIST REFERRALS USING GRAPH NEURAL
NETWORKS: Exploring the Relationship Between Physician Referral Patterns, Primary Care
Access and Healthcare Spending

Isabel Maria Mora Labarca

Work project carried out under the supervision of:

Rongjiao Ji

Qi Shi

25-01-2023

Abstract

The healthcare referral system affects all points of the healthcare ecosystem – access to care, patient satisfaction, physician utilization and healthcare costs. The state of these variables plays a critical role in determining healthcare efficiency. In this paper we dissect the medical referrals from primary to secondary care in Florida in 2015 and tackle them from three perspectives – influence of physician experience in referral choice, relationship between physician referral choice and Medicare spending, and pattern detection given different referral windows. To accomplish our goal of identifying patterns in primary to secondary referral mechanisms, we use Graph Neural Networks (GNN) unsupervised model to learn the vectoral representation of our physician nodes and their properties in the network. This work provides new discoveries on factors that influence the referral patterns and can be used to make better decisions when aiming to improve the efficiency of referrals.

Keywords: Graph Neural Networks, National Provider Identifier, Physician Referral Network, Medicare

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209), POR Norte (Social Sciences DataLab, Project 22209) and CMU Portugal program explorative research project MD2TRUST (CMU/TIC/0016/2021)

Table of Contents

1. Introduction	8
2. Literature Review	10
2.1 Previous work on healthcare referrals	10
2.2 Graph Embeddings and Graph Neural Networks.....	14
2.3 Applied GNN in Healthcare	16
3. Data	17
3.1 Data Collection and Understanding	17
3.2 Data Curation	18
3.3 Exploratory Data Analysis	19
3.3.1 Overview of physician's attributes.....	19
3.3.2 Referral Analysis.....	21
3.3.3 Network Analysis.....	22
4. Methodology	31
4.1 Feature Selection	33
4.2 GraphSAGE and Node Embeddings	33
4.3 Dimensionality Reduction.....	34
4.4 Model iteration and Fine Tuning.....	35
4.5 Clustering methods.....	36
4.6 Example of Model Output.....	38
5. Exploring the Relationship Between Physician Referral Patterns Primary Care Access and Healthcare Spending.....	40

6.	Discussion	52
7.	References	54
8.	Appendix	61

List of Figures

Figure 3-1. Top 10 specialties ordered by average services provided by secondary care.....	20
Figure 3-2. Average services provided by specialty and gender.....	20
Figure 3-3. Number of Beneficiaries per Specialty	21
Figure 3-4. Nodes and edges from the 97.5th quantile of the top 50 In-Degree Scores (right) and Out-Degree Scores (left).....	23
Figure 3-5. Nodes and edges from the 97.5th quantile of the Eigenvector Scores	24
Figure 3-6. The degree distribution of nodes. Degree distribution of physicians' network (x axis limited to 70) (left) and degree distribution of ER random network (right).....	27
Figure 3-7. Distribution of Node Linkages. Log-log plot of scale-free network	29
Figure 4-1. K-Means algorithm explained	36
Figure 4-2. TSNE visualization of GraphSAGE embeddings. Dimensionality reduced using TSNE (left) and UMAP (right) methods.....	39
Figure 5-1. Node embeddings visualization using TSNE (left) and UMAP (right), where yellow represents the top 10% of physicians with highest average submitted charges per services. ..	46
Figure 5-2. Node embeddings visualization using TSNE (left) and UMAP (right), where blue represents “Cluster 0” and coral represents “Cluster 1” as labeled by K-Medoids clustering. 46	
Figure 5-3. Boxplot distributions of Average Submitted Charges per Service and Total Medicare Payment Amount each Provider, by Cluster	47
Figure 5-4. Log-scale histograms for Total Medicare Payment Amount and Average Submitted Charges per Service provided, by cluster.....	48
Figure 5-5. Count of Primary Care providers and specialists, by cluster.....	49
Figure 5-6. Percentage of Physicians for that region that were assigned to cluster 1	49

List of Tables

Table 5-1. Summary Statistics of Financial Variables	44
Table 5-2. Summary Statistics of Referral-Based Variables	44
Table 5-3. Price, age, sex & race-adjusted Medicare reimbursement rates 2015 Miami & Ocala	49
Table 5-4. Primary Care Access and Quality Measures for Medicare Enrollees, 2015 (Miami & Ocala)	50
Table 5-5. Referral Measures by Cluster	50
Table 5-6. Results and Statistical Significance Summary	51

List of Abbreviations/Acronyms

CMS – Centers for Medicare & Medicaid Services

DBSCAN - Density-based spatial clustering of applications with noise

EHR – Electronic Health Records

ER – Erdos-Renyi

ETG – Episode Treatment Group

GCN – Graph Convolutional Neural Networks

GNN – Graph Neural Network

GRAM – Graph-Based Attention Model

HRR – Hospital Referral Regions

NPI – National Provider Identifier

PC – Primary Care

ReLU - Rectified Linear Unit

RNN – Recurrent Neural Network

RVU – Relative Value Unit

SC – Secondary Care

t-SNE - t-Distributed Stochastic Neighbor Embedding

UMAP - Uniform Manifold Approximation and Projection

WHO – World Health Organization

WWW - World Wide Web

1. Introduction

For as long as humankind exists, people have been highly focused on enhancing the quality of life and increasing life expectancy by improving the quality of healthcare services, meaning “any services provided by a healthcare professional, or by any individual working under the supervision of a health care professional, that relate to— (A) the diagnosis, prevention, or treatment of any human disease or impairment; or (B) the assessment or care of the health of human beings.” (*Legal Information Institute* n.d.).

According to the World Health Organization (hereafter, WHO), to attain quality healthcare, the services must “reduce waiting times and harmful delays”, “provide care that does not vary in geographic location”, and “maximize the benefits of available resources and avoid waste” (WHO n.d.). Thus, the referral system is one of the most important pillars in the healthcare service chain that determines whether these aspects are implemented.

The healthcare system is built in a way to encourage beneficiaries to first reach out to primary care providers and then, if needed, seek out specialist care, in this way maintaining lower costs for the patients. The conventional paper-based referral systems are designed to optimize the doctor's workload, maintain enough time to address patient problems, and connect the patient to a specialist that is best equipped to approach his needs on time. Through GNN and clustering, our results showed that lower submitted charges are related to a higher number of interactions with different physicians, thus with a wider network. Furthermore, the region that was more representative of this cluster showed lower healthcare spending and higher care quality measures. In conclusion, it appears that higher quality, lower submitted charges and lower healthcare spending are related to having a wider network.

Regretfully, the real-world referral system bypasses the optimal scenario, with certain physicians being overthrown by too many patients, and having little time to address their needs, resulting in untimely and inefficient care. A crucial aspect in the referral system and for primary

care providers is the possibility that they disregard the staff's availability and, instead, refer the patient to a specialist that they are familiar with. We believe manual-based referrals may be prone to biases, and through this work project we will analyze this hypothesis by considering the healthcare outcomes and physicians' referrals. Given that physicians seem to make more referrals to physicians of similar experience level, the experience seems to influence physician referral choice. Nevertheless, the extent to which experience has an impact is unclear, because, given information of experience, the extracted embeddings do not seem to have a clear structure.

Physician cooperation is the practice of doctors working together to offer the best possible treatment for their patients. This can take numerous forms, such as consulting with other doctors on a case, exchanging information and skills, or collaborating on research. Collaboration is an important component of the medical sector because it allows doctors to combine their talents and knowledge to deliver the best possible treatment for their patients. This physician collaboration can take place both in the short and long term. In the short term, doctors may collaborate on a specific case or patient, working together to diagnose and treat a particular medical condition, thus it is important. Our results depict the existence of pattern differences when comparing 30-day referrals and 90-day referrals.

This work project is structured as follows: Section 2 consists of a literature review regarding previous studies on the topic. Section 3 will provide details on the data considered for this research. Section 4 will specify the methodology followed by this study. The following sections focus on addressing business questions. Section 5 will compare the patterns between different referrals windows given physician specialties. Section 6 will attempt to explain the relationship between Medicare spending, primary care quality and physician referrals. Section 7 will aim to infer the experience significance in primary care physician referral choice. Finally, sections 8 and 9 will provide final remarks, conclusion and discussion.

2. Literature Review

2.1 Previous work on healthcare referrals

According to Md Abu Bashar et al. (2019), referrals in healthcare is an ever-changing process in which a healthcare representative, driven by a lack of resources such as skills, passes on the management of a certain clinical condition to a better-equipped healthcare worker. The mentioned referral process, oftentimes held as a measure to determine the performance of the health system, is deemed effective when it is constructed to ensure that individuals receive the best possible care and that all levels of health care maintain a close relationship (Prof. Ali Akbari-Sari 2021).

Beyond theoretical definitions of how a sound healthcare system is supposed to function, society is struck with grim health statistics. According to 2020 Health OECD data, the United States had 238 deaths per 100,000 thousand population that could have been avoided if the patient had received proper care. Behind these numbers lie overworked and exhausted doctors and patients experiencing long waits resulting in curable diseases becoming untreatable. All these factors are a result of a faulty functioning referral system and can be tuned to best serve the giving and receiving ends of healthcare services.

The basis of this work is to understand the driving forces of the referral system by examining previous research, conducting novel analysis to obtain insights, and be able to provide competent recommendations on how this imperative process could be improved to optimize healthcare operations.

A handful of research has been conducted to examine the United States referral system and the components that influence it. Chuankai An et al. (2017) in their work emphasize the importance of a well-designed referral system for resources to flow efficiently in the healthcare ecosystem. The author highlights three causes driving primary care doctors to refer a patient to a specialist, which, sub sequentially, determine the quality and price of care: “(1) seeking advice

on a diagnosis or treatment (52.1%), (2) requesting surgical management (37.8%), and (3) asking the specialist to directly manage the patient (25.1%)” (Chuankai An et al. 2017, 2). Employing the 2009 – 2015 data from Medicare, the authors measure their network structure by comparing it to three traditional baseline models: Erdos-Renyi (ER) random network (the null model), small-world network characterized by a higher clustering coefficient and network homophily, and core-periphery structure (measured using the Gini coefficient) that contains a “core” of closely connected nodes and a “periphery”, which includes vertices linked loosely to the core and one another. The authors found that the physician’s networks exhibit both, core-periphery and small network structures; the latter suggests that doctor networks are suitable for spreading innovations and passing on knowledge and information.

Conventionally social networks unveil the so-called friendship paradox, which implies that people are more likely to form friendships with individuals that are already friends with more people. These networks seem to follow a power-degree distribution, which means that some individuals are “hubs” of the network and have a vast number of connections, while most individuals in the network have very few connections. Intuitively, the doctor’s network could adhere to a similar structure, as there might be physicians that are in a sense more “popular” and, hence, receive and make more referrals. Chuankai An et al. (2017) found that indeed the doctor’s network follows a power law, with an outdegree distribution having a more robust inclination than the indegree, meaning that certain doctors conduct more referrals than others.

The inclination of some doctors to conduct more referrals is a key factor to understand. A study by Peter Franks et al. (2000) of United States and United Kingdom found that family care to specialist referral rates ranged from 5% to 60% per year, which affects patient access to specialists. They examined the deviations of primary care to specialist’s referral rates and factors that affect it using 1995 claims data that indicated whether a patient was referred or not. Although factors, such as reimbursement, time pressure, and clinical problems were found to

account for minor variations, physicians with a bigger fear of malpractice or higher specialization were more likely to refer their patients.

Extending beyond the decision of whether to refer a patient, Kraig S. Kinchen et al. (2004) studied how PC doctors choose the specialist. By running a cross-sectional survey, the authors report that respondents find medical skill, previous experience with the specialist, and patient's insurance coverage to be of high relevancy. The latter often leads to family doctors having little information on whether the specialist would provide the best care for the patient.

“Birds of a feather flock together” - an idiom used to express the nature of social networks where beings tend to form connections with others of similar type and personality represents an idea that might resemble the physician network. This is what is known under Network Analysis as homophily. Although ideally, physicians build ties with other physicians for patient referrals and clinical advice (Michael L Barnett et al. 2011), associations between doctors tend to be affected by other aspects as well (Bruce E. Landon et al. 2012).

Bruce E. Landon et al. (2012) in their work employs U.S. Medicare data (2006) on shared patients among physicians and draw attention to doctor traits, such as sex, age, location of practice, the intensity of care (using Episode Treatment Group (ETG) software), and overall differences in network features across 51 hospital referral regions (HRRs). After employing a multivariable regression model, the authors' findings suggest that doctors are used to share patients with other providers with similar personal traits; for instance, more than 65% of pairs between doctors were male-male, the average difference in age between those with ties (11.5 years) was smaller than those without (12.5 years), and 96% of unconnected physician pairs did not work in the same hospital (c. 31% of connected physician pairs were from the same hospitals). Additionally, physicians with ties had closer geographic proximity (mean of 13.2 miles for connected pairs versus 24.2 miles for unconnected pairs) and alike practice intensity

estimated by ETGs. Similar patient characteristics were also shared between connected physicians, including patient age, race, and complexity of the clinical issues.

Previous work in the field of network analysis has been developed for the healthcare industry, leading to successful applications to investigate physician referral networks, “advice networks, and the diffusion of information among physicians” (Barnett et al. 2011). The further motivation behind previous work in this area has been focused on healthcare efficiency, including hospital costs, treatment quality, and patient needs. By analyzing the network, these studies extracted information about possible gaps in the system's efficacy. For example, Barnett et al (2011) evaluated how the patient-sharing networks of doctors contributed to the expenses and intensity of care delivered by United States hospitals by studying physician-based networks. They concluded that the network structure was strongly associated with Medicare spending and care patterns. Another finding was that hospitals with doctors who have a higher number of connections have also higher costs and more intensive care; hospitals with primary care-centered networks have lower costs and care intensity.

An et al. (2018) decided to carry out an in-depth study of the United States network by considering only patients with cardiological conditions. Through the analysis of metrics such as local clustering coefficient, betweenness, closeness, eigenvector, and PageRank centralities and the core-periphery score, as well as node embeddings features such as their position in the network, they concluded that physicians send patients to other physicians that have many “common connected neighbors in the national referral network” (An et al. 2018, 22). This supports the hypothesis that physician position within the network influences their level of popularity, being that those with more common neighbors would be reached out more often. Through their research, An et al. (2018) considered that it would be beneficial for hospitals to grasp the discoveries for several reasons. First, whenever a new treatment is approved, if hospitals are aware of the first physicians and hospitals that adopt it, they could extract key

metrics from their network that could possibly contribute to a similar level of success and popularity. Moreover, by implementing Machine Learning algorithms on their data, they were able to isolate the key factors in predicting referral paths characteristics, “such as the time gap between two visits on the referral path and the total RVUs of all physicians’ endeavors.” (An et al. 2018, 22).

Yet, other studies give more emphasis on the patient’s development to understand physicians’ social networks. Herrin et al. (2019) focused on claim data from breast cancer patients to “develop an empiric approach for evaluating the performance of physician peer groups” (Herrin 2019). This approach was based on social network analyses to understand the existing relationships between physicians and how that could impact the patient’s health.

“Physician-to-physician referrals are the currency of day-to-day transactions in medicine” (*New York Times* 2009). Although instinctively we think of referrals as a process driven to best serve the patient’s need or influenced by the doctor’s skills, studies show that referrals are often made based on friendships. This work will extend the existing research by incorporating physician features such as location and specialty whilst studying the latest (2015) Medicare data available with the aim of understanding the physicians’ network structure and how could it be shaped to best serve the healthcare system.

2.2. Graph Embeddings and Graph Neural Networks

Regarding the Artificial Intelligence area, some advances have been made in the field of graphs and Machine Learning. Neural networks have been one of the most efficient types of models for prediction and classification tasks. Similarly, GNN can repeat these tasks on nodes and edges as primary subjects. Especially in the healthcare field, GNNs would allow us to deeply understand the physician referral network, extract network metrics and predict links, classify nodes, or determine the efficiency of the network within the healthcare system. This is done with the generation of graph embeddings. The methodology of graph embeddings has

been explored by previous studies (Grover and Leskovec 2016; Perozzi et al. 2014; Ribeiro et al. 2017; Tang et al. 2015), where they aimed to extract each node's information and position within the network as a low-dimensional vector. Thus, they proposed embedding mechanisms where this information gets transformed and used afterward as input features for other Machine Learning models – logistic regression, and decision trees, among others.

As expected, different graph embedding mechanisms have been proposed. Random-walk based methods were proposed with the goal of generating random node paths and, with such, learning more about the node's position regarding the overall network. (Yue et al. 2020). More specifically, given a starting node, this mechanism will move from the starting point to a random neighbor, and then repeat the process until it has a node sequence. One of the first algorithms proposed for this operation was DeepWalk (Perozzi et al. 2014), which provides information regarding the local structure by truncating the random walks. Node2vec (Grover and Leskovec 2016) accommodates a biased random walk that is considered more flexible because it incorporates several types of samplings to generate the node sequences. Moreover, Struc2vec (Ribeiro, Savarese, and Figueiredo 2017) is another framework proposed that differs from the Node2vec regarding the meaning of the produced representations. While Node2vec tries to represent nearby nodes in a similar way, Struc2vec focuses on the role that each node plays within a network. As a result, nodes that are far apart may have similar representations. This is done by applying DeepWalk to a multi-layer weighted graph, where layer k is parameterized considering the k -hop neighborhoods of the nodes. Attri2vec (Zhang et al. 2019) implements DeepWalk and Struc2vec to learn node representations, with the difference that it performs either linear or non-linear mapping on the nodes' attributes.

Hamilton, Ying, and Leskovec (2017) introduced the GraphSAGE framework, where they developed a function to extract node attributes - node degrees, features - by sampling a neighborhood and then aggregating feature information from neighbors. This methodology

primarily considers the local node neighborhood information, besides the node features, and transforms it into a new, lower-level space. The main difference of this framework from the rest is that it is inductive, which means that it has a higher probability to be generalizable and achieve better performance on previously unseen data.

2.3.Applied GNN in Healthcare

Some studies have included GNN-based methodologies to apply solutions to the healthcare sector. Liu et al. (2020) implemented GNN algorithms to predict patients' prescriptions for the next period. Because this was considered both a temporary and spatial problem, they implemented a novel hybrid method of GNN and Recurrent Neural Networks (RNN), where the RNN considers the patient sequence representation, and the GNN the graph that matches different medical events to their according prescriptions.

Other studies have considered the GCN for the medical domain. Choi et al. (2017) proposed a Graph-based Attention Model (GRAM), where they used electronic health records (EHR) to predict sequential diagnoses. When compared to RNN, GRAM outperformed accuracy by 10% in the task of predicting rarely observed diseases, helping the medical community to extract “medical concept representations from the graph of medical ontology knowledge” (Choi et al. 2017). Ma et al. (2018) classified the drug-drug interaction (DDI) problem as a graph task, where they were able to classify each graph using GCN.

GNNs have the potential to not only state that the current healthcare system is not efficient, but to model what links might be best to improve either the patient's medical services quality or healthcare efficiency, reducing costs. However, this area of research is still relatively recent and has not been deeply explored. Thus, little research has been focused on physician referral patterns represented through GNNs. To improve healthcare efficiency, mostly network analyses considering social network theory have been considered. Hence, our goal is to extract the

benefits of GNNs for their implementation in this field and provide insights at the patient, physician, and hospital levels.

3. Data

3.1.Data Collection and Understanding

In this paper, we study the components that form the network links between physicians and depict the rationale behind medical referrals in the US healthcare system by employing the Centers for Medicare & Medicaid Services (hereafter, CMS) patient referral dataset. This Medicare data includes information on when the same patient is reported as receiving care from two distinct doctors within a specific timeframe, between the years 2009 and 2015, spaced out across 30-, 60-, 90-, and 180-day periods. This project utilizes the latest data available (2015) for the 30-day period, which has about 34 million record count (CMS 2021).

According to the data dictionary provided by the CMS, the National Provider Identifier (NPI) is "a unique identification number for covered health care providers" (Centers for Medicare & Medicaid Services n.d.). The data contains two columns of NPIs, the first of which is labeled as the "Initial Physician NPI" (NPI 1) and the second as the "Secondary Physician NPI" (NPI 2). The referral dataset does not assure that doctor A referred to doctor B, meaning that for each pairing of NPI 1 and NPI 2, the same patient visited both doctors in that order within the 30-day period. For the analysis, we selected the referrals from primary care (NPI 1) (family practice, internal medicine, pediatric medicine, geriatric medicine, general practice, obstetrics/gynecology, and preventive medicine) to specialists (NPI 2) because patients usually go through primary care provider as their first option, and as such this person might determine how fast the patients reach their final specialist required for their diagnosis. Furthermore, constructing a GNN with fewer connections will allow an easier extraction of information from the nodes and the reduction of noise and edges. The data also includes a shared count of the patients for each pair, allowing for a deeper understanding of the providers with the greatest

number of connections. This relationship is extracted from all the Medicare claims filed between each of the periods. For instance, if a patient is listed under two claims from different physicians within each of the periods mentioned, these physicians will have an additional patient summed to the shared count in the referral dataset.

Besides this data set, we extracted additional data that contains physicians' features from the 2015 Medicare Physician & Other Practitioners - by Provider report from the Centers for Medicare & Medicaid Service (Centers for Medicare & Medicaid Services Data). This way, we would be able to study the factors that affect referrals. The NPI column, which serves as the connection between the two data sources, is included in this dataset along with numerous doctor profiling features – making a total of 73 columns in the data set –, including first and last name, gender, city, specialty, and others, as well as the characteristics of their beneficiaries (e.g., average age, the proportion of men and women, number of distinct races, etc.).

3.2. Data Curation

Regarding the subset of the data that we chose, we considered only those physicians with practice locations set in Florida. This way, we will only study intra-state referrals. Moreover, we considered only those referrals who shared over 100 patients within that month to enable the measurement of the shared patients among healthcare providers by focusing on the group that interacts most frequently. The two main reasons why we chose Florida as our study set is:

1. Medicare data include patients that are 65+ and “Florida had the highest percentage of its population age 65 and over among states in 2015 (19.4 percent)” (United States Census Bureau 2016).
2. Geographically, Florida is a unique state given that is only contiguous to Georgia and Alabama in the north, thus we would be able to minimize the interactions of patients from out-of-state, those being more common in areas such as the Midwest.

After considering this subset, we disregarded features that are not relevant or whose data was missing. There is a difference in the number of distinct healthcare providers, where the features' dataset as 15,448 unique NPIs and the referral dataset 15,439 unique NPIs. Only those physicians that are present in both datasets (15,439) will be used. This resulted in a directed referral network where each node will be represented by a healthcare expert and the edges will be their referrals, hence there will be 15,439 nodes and 67,480 edges.

Depending on the business question that each study described before aims to answer, different features will be curated and explained accordingly. To illustrate the output of the model, we will use gender as the input feature and visualize the embeddings considering this attribute. Therefore, to satisfy the requirement for the algorithm to have a numerical input, we transformed the gender feature `Rndrng_Privr_Gndr` into a binary column.

3.3.Exploratory Data Analysis

3.3.1. Overview of physician's attributes

Some further analysis was done to help interpret the data, such as calculating the proportion of male and female users per amount of service. The results revealed that male doctors outnumbered female doctors by a difference of 75% to 25%. To obtain a general overview of the doctor's attributes, in the following part we analyzed:

1. Frequency of specialties per primary and secondary care;
2. The secondary care specialties that provided the highest average of services
3. Average services provided by specialty and gender;
4. Number of beneficiaries per specialty;
5. Total submitted charges and total Medicare paid amount by specialty.

Family practice, internal medicine, and general practice are the most frequent primary care physician in the data set, whereas cardiology, nurse practitioner, and diagnostic radiology are the specialties with the highest number of appearances. The secondary care specialties with the

highest average services provided are hematology/oncology followed by medical oncology (Figure 3.1).

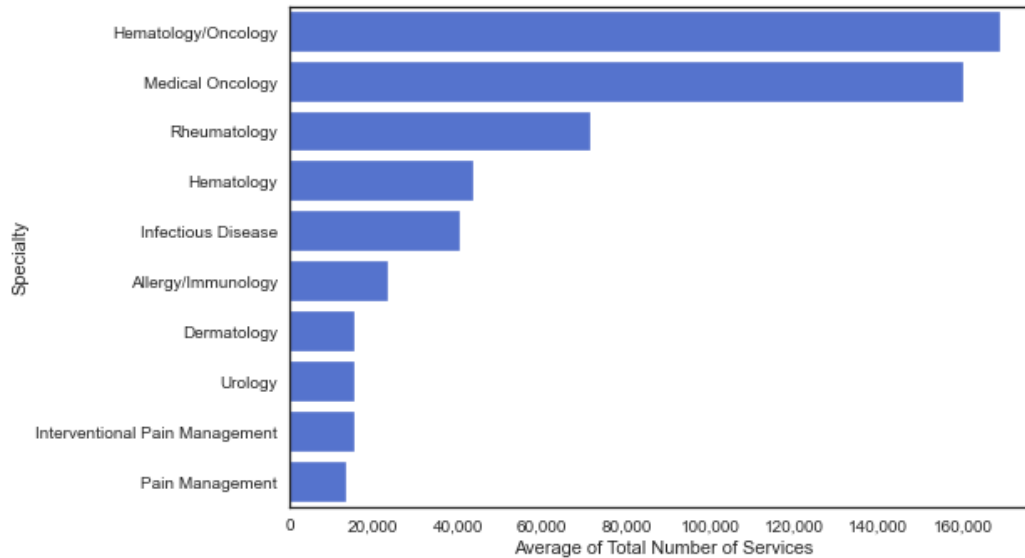


Figure 3-1. Top 10 specialties ordered by average services provided by secondary care

By observing the top 10 specialties with highest average of services provided by specialty and gender (Figure 3.2) we can see that females seem to, on average, provide a wider amount of services. However, we might need to interpret this result cautiously given that the number of female physicians per specialties is frequently lower, making the average higher.

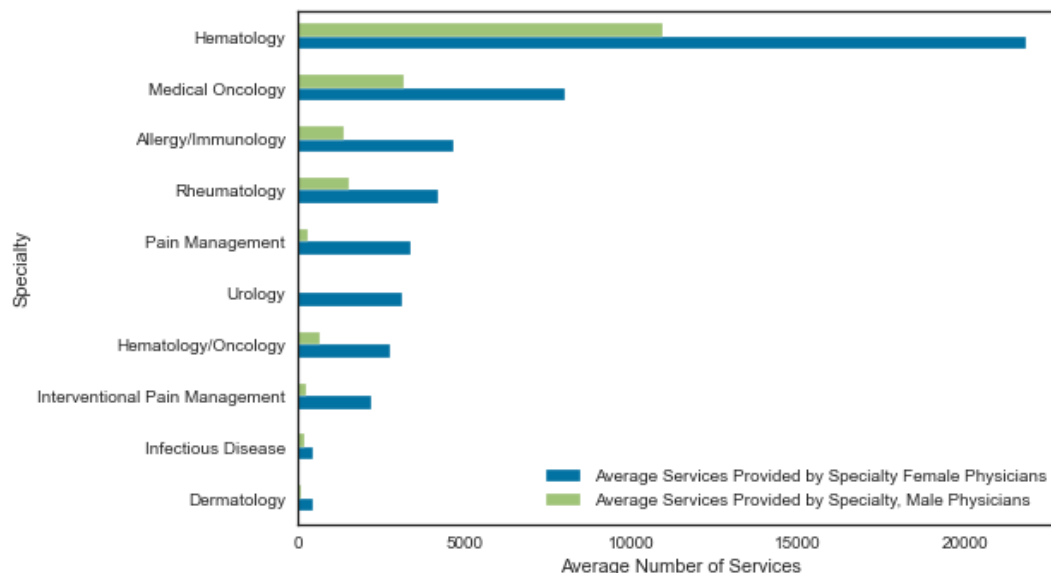


Figure 3-2. Average services provided by specialty and gender

When filtering specialties by the total number of beneficiaries, there seems to be a stronger demand for diagnostic radiology, cardiology, and pulmonary disease (Figure 3.3).

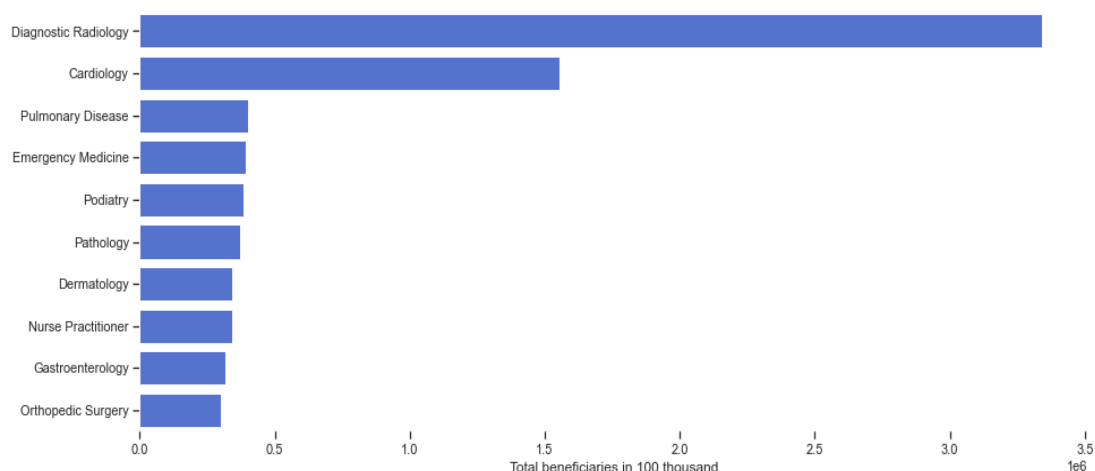


Figure 3-3. Number of Beneficiaries per Specialty

Looking closer at the specialties and the finance variables, the top 3 specialties that had the highest submitted charges, as well as the highest Medicare paid amount were hematology/oncology, medical oncology and radiation oncology.

3.3.2. Referral Analysis

The referral dataset from 2015 contains the National Provider Identifier (NPI) of the physician that refers and the physician that receives those referrals. However, to represent the number of times that this instance occurs between two distinct physicians, the “Shared Count” column is provided. We calculated that 75% of these connections between physicians have no more than 248 referrals within the timeframe under study (30 days). The pair of physicians with the most shared count referrals have 11,085, where the primary care provider is a general practice physician, and the specialist is a psychiatrist. The second pair had 9,998 referrals shared from internal medicine to nephrology, and the third pair had 9,863 referrals shared from internal medicine to pathology. Yet, to look from a more general perspective, we grouped physicians by specialty, and calculated the average referral count. Some findings include that geriatric medicine, pediatric medicine, and preventive medicine are among the specialties that make more referrals. Occupational therapists, hematology, and nurse practitioners are among the

specialties that receive the most referrals on average. However, given the way that we subset our dataset, it is important to note that referrals made from Secondary care specialties were not considered, as well as those made to primary care providers. Lastly, the primary care provider's gender with the highest referral count was male. On the contrary, the secondary care physician's gender with most referrals, on average, was female.

3.3.3. Network Analysis

3.3.3.1. Centrality Measures

The centrality measures are also referred to as social network analysis since they are fundamental in depicting how the graphs work by evaluating the importance of the physicians on the overall network. Thus, this paper will use degree and eigenvector centralities to study the doctors' influence on the rest of the community. Because our network only consists of primary care referrals to specialists, and all the referrals within secondary care specialists were excluded, we will not analyze the closeness or betweenness centralities.

Closeness centrality aims to measure how close the nodes are to one another. A node is deemed key in this scenario if its distances from other nodes are shorter, which means that doctors may more eagerly refer patients to these nearby nodes than to the ones further away. However, because we have a directed graph, the only paths that exist contain two nodes – the primary care provider and the specialist. As a result, there are no distances longer than one referral in our graph. This impacts directly the insights that we could obtain from the closeness centralities; thus, we will not consider it.

The betweenness centrality is another well-known centrality metric. This metric, rather than just counting the number of edges a particular node has, measures how many times a node appears in the shortest path since a node would only be deemed important in this scenario. This measure is used to determine who affects the network's flow the most, showing which nodes

are acting as "bridges" in the process. As mentioned before, because there are no differences between our shortest paths, this measure would be irrelevant to our network analysis.

For the rest of the analysis, we used *networkx* library with the following functions: *in_degree_centrality*, *out_degree_centrality*, and *eigenvector_centrality*.

One of the social network metrics investigated in this research is **degree centrality**, which is used to identify the most popular nodes. This centrality purely considers the number of referrals held by each physician and assigns a significance score to them. Because we are utilizing a directed graph in this case, it is vital to distinguish between **in-degree** and **out-degree**, and as the names imply, they assess the number of recommendations received by or supplied by a doctor, respectively. These scores will then translate how many ties there are from node to node, with higher values corresponding to doctors who have more connections than the norm. This measure is typically used to detect highly connected nodes as well as the most popular ones – “individuals who are likely to hold most information or individuals who can quickly connect with the wider network” (Disney 2022).

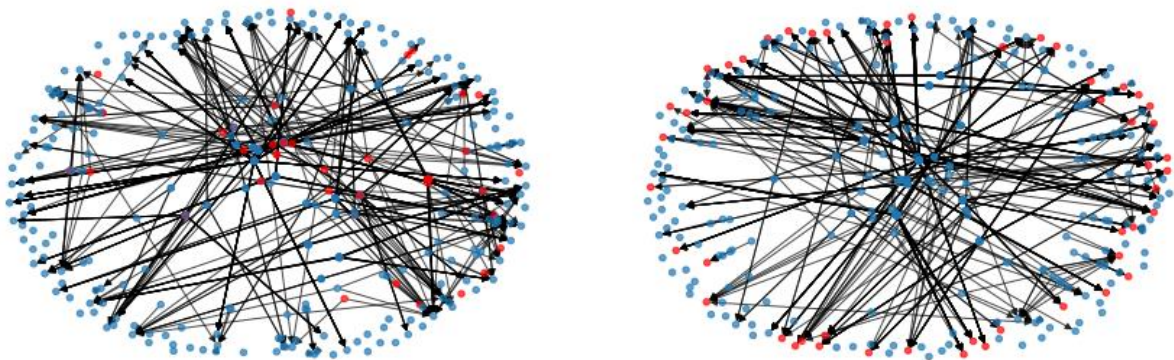


Figure 3-4. Nodes and edges from the 97.5th quantile of the top 50 In-Degree Scores (right) and Out-Degree Scores (left)

Figure 3.4 shows the 97.5th quantile of the top 50 scorers' nodes and edges - this constraint in the data will be present in the following figures concerning the centrality measures. This allows us to detect a denser presence of nodes in the center of the plots, most likely the primary care physicians, and dispersed nodes around, that might represent the secondary care

physicians. We noticed that the in-degree and out-degree bulk of red dots (top physicians' scorers) are different, meaning that the physicians with the highest in-degree score will have 0 out-degree scores, and vice-versa. Since the secondary care physicians are the ones that receive the referral, it would only make sense for them to have the highest in-degree scores; so when we look at the specialties scorers, we see that the prevalent ones are diagnostic radiology, cardiology, and pulmonary disease. On the other hand, internal medicine, and family practice show a clearer commonality on the out-degree top scorers. Finally, most of the greater out-degree scorers are located in either Orlando, or Winter Park, whereas the in-degree scorers, are most common in Fort Myers, Panama City, and Tampa.

The **eigenvector** measure shows the significance of a particular node given its links to other significant nodes. As a result, it takes into account how many connections a node has overall in the network as well as how highly rated its neighbors are (meaning, how many connections its connections have). Therefore, it is clear that doctors with higher eigen scores contribute more to the network as a whole and have a greater influence on the referral system, and that their connections likewise have higher scores (since the main high-degree node has a strong influence on them). Furthermore, this social network measure is excellent for depicting distances between nodes depending on their similarity, allowing for greater flexibility over how much effect the features will have when assigning weights. As a result, we used the number of referrals between each physician pairing as the weights to determine their relationships and contributions.

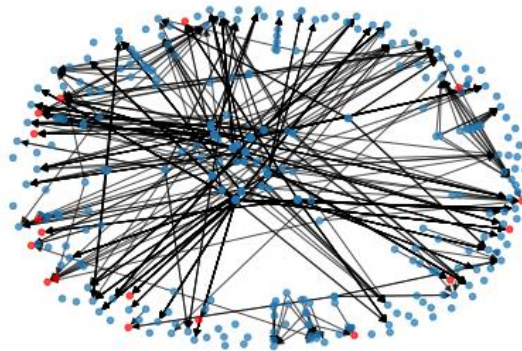


Figure 3-5. Nodes and edges from the 97.5th quantile of the Eigenvector Scores

The doctors with higher scores are dispersed in the graph, as we can observe from the representation in Figure 3.5 above. The specialties shown in red are the ones that have a bigger impact on the network and are similar to the in-degree centrality top scorers – diagnostic radiology, pulmonary disease, and infectious disease, however, the highest scorer is a nephrologist. In this case, the new NPIs with better ratings have an overall greater number of female beneficiaries than male beneficiaries. Furthermore, as expected, these specialties with higher eigenvector scores would have a 0 out-degree score and a higher in-degree score.

3.3.3.2. Shortest Path

For our network analysis, some of the relevant features to calculate will be the average path length and number of connected components. This allow us to study the graph structure before we apply our GNN model. The shortest path from node A to node B is defined as the path of minimal length between these two nodes. Because of our graph structure, we decided to make our graph undirected for the following part of the analysis since this could provide information regarding how long the chain of connections is between primary care providers and specialists.

To calculate the shortest paths of a graph, different algorithms may be implemented. We used the *networkx* library *shortest_path* function, along with its default method, which implements Dijkstra's algorithm (Dijkstra 1959). Similarly, a network's average path length is the average shortest distance between two nodes of the graph. In other words, the average path length is the average of the previously mentioned shortest paths of the graph. This algorithm will consider every node of the graph as a source node. However, because our graph contains more than one connected component, which will be reviewed in the Connected Components and Connectivity section, the distance between two nodes from different components will be infinite and will not be defined. As a result, we calculated the average path length on the largest connected component. The distance found was 7.58 as the average shortest path length.

According to Perez and Germon (2016), a low average path length l is where $l \sim \ln n$. In our graph, $\ln n = 9.64$. Thus, our average path length is low. This metric is considered a metric of efficiency regarding how information flows through the network, where a low average path length indicates information is not spread throughout the network, and a high one represents that even “outliers” receive information. However, our graph provides more information about the possibility of physicians knowing one another based on referrals. Another metric used to evaluate distances in the graph is the diameter, defined as the maximum eccentricity. The eccentricity is the maximum distance from source node A to all other target nodes B . The diameter of our network, undirected, is 19. This indicates that there is one chain of connections where node A and node B have, at least, 18 other physicians that they could potentially reach in social circumstances. This shows how long the connection is between specialists referred by the same primary care provider or between those that have referred their patients to the same specialist. Thus, specialists that share the same primary care providers would be connected by the primary care provider, and primary care providers that have been referred to the same specialist would also be connected between them.

3.3.3.3 Connected Components and Connectivity

In a directed graph, we can find either weakly and/or strongly connected components. A weakly connected component is a subgraph where each node can be reached by another in any direction, that is, the source node A can reach the target node B , or vice versa, but not necessarily source node B could reach source node A , or vice versa. Thus, as long as the nodes are part of a component in any direction, they are part of a weakly connected component of the graph.

Intuitively, a strongly connected component is a subgraph where every source node A can reach target node B , but also every source node B can reach target node A . Thus, they are connected in both directions. It is unlikely, but possible, that a directed graph itself is strongly

connected - meaning the graph has one strongly connected component, conformed by all nodes. In direct acyclic graphs, there are no strongly connected components, given the nature of their cycle. Because our data does not contain any edge where a specialist would refer patients to any physician, our graph would not contain any strongly connected component.

In an undirected graph, a connected component is a subgraph where each node can be reached by another - they are connected. Hence, a graph where this structure is presented has only one connected component. However, if there are cases where a pair of connected nodes cannot be reached by another, then there is more than one connected component within the network. In real-life networks, it is common to find one considerably big, connected component, and many other small ones. In our network, we have a total of 188 connected components, with the largest one having a size of 14,987 nodes - approximately 97% of the total nodes in the network.

3.3.3.4. Type of Network

3.3.3.4.1. Erdős–Rényi

Comparing the physician's network to other baseline graph models allows us to infer the properties of the studied network. We begin by comparing our network to Erdos-Renyi (ER) random graph model, which possesses properties such as binomial degree distribution, low variation in node degrees, and low clustering coefficient. The main idea behind the ER model is that each connection, i.e., edge, is equally likely to be present or absent, regardless of other

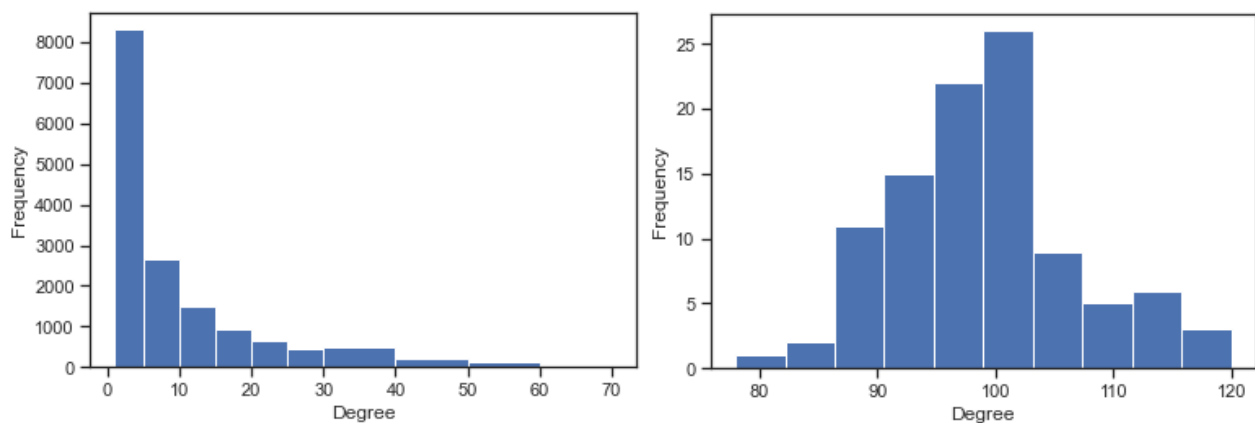


Figure 3-6. The degree distribution of nodes. Degree distribution of physicians' network (x axis limited to 70) (left) and degree distribution of ER random network (right)

edges (Fienberg, Stephen E. 2012). Nevertheless, real-work networks rarely form at random and usually, the presence or absence of one connection does impact the state of the other.

First, when observing the network degree distribution plot, the shape of the histogram resembles the power law distribution with an average degree of 9, in contrast to ER graph that follows a binomial distribution (Figure 3.6). This implies that a small number of doctors give and receive many more referrals, while the majority of doctors have fewer connections (Figure 3.6). Nevertheless, degrees might follow different types of distribution, which we will test in the following parts of the paper.

Second, the physician network node degree variance is c. 1.4, which is a low variation of node degrees. Last, the average clustering coefficient of the physician network is 0; a clustering coefficient of 1 would mean that all doctors and their neighbors know each other, however, since our study comprises primary to secondary care referrals, there are no physician connections that form “triangles”, i.e. physician’s neighbors are connected.

The physician network structure seems to follow the properties of the Erdos Renyi graph and therefore, it has a random network structure. The power-law degree distribution is a characteristic of a scale-free structure.

3.3.3.4.2. Scale Free Network Analysis

Scale-free networks are characterized by having a node degree distribution that follows a power law. That is, networks with a power-law degree distribution will be composed of a majority of nodes with very few connections (degree), and with only a few nodes that have a high number of connections - known as hubs. These types of networks have been studied and real-life examples include the World Wide Web (hereafter, WWW) and the research citation network. Other social networks are considered to be scale-free as well, but there has been some debate regarding whether they are truly scale-free (Broido, A. D., & Clauset, A. 2019).

Barabási and Albert (1999) studied for the first time the power-law degree distribution in networks. They argue that the scale-free structure of real-life networks was based on two distinct characteristics that were being missed by other studies on random networks. These networks were always considering a fixed number of nodes N for the random graph generator. As we know, the WWW or the citation network are graphs that grow constantly, a condition of these types of networks that had not been included within network science studies beforehand.

As mentioned earlier, random graph models assumed that the probability of a node connecting to another node was independent of the node degree for the graph to be "random". However, Barabási and Albert argue that most networks display what is known as preferential attachment - the phenomenon where nodes would have a higher probability to connect to other nodes based on their degree. For example, it is common in the research citation network that papers cite already known and popular papers, usually cited already by many others, rather than those that are barely known and/or recently published. Similarly, on the WWW, a blog is more likely to contain hyperlinks to well-known websites rather than less popular ones.

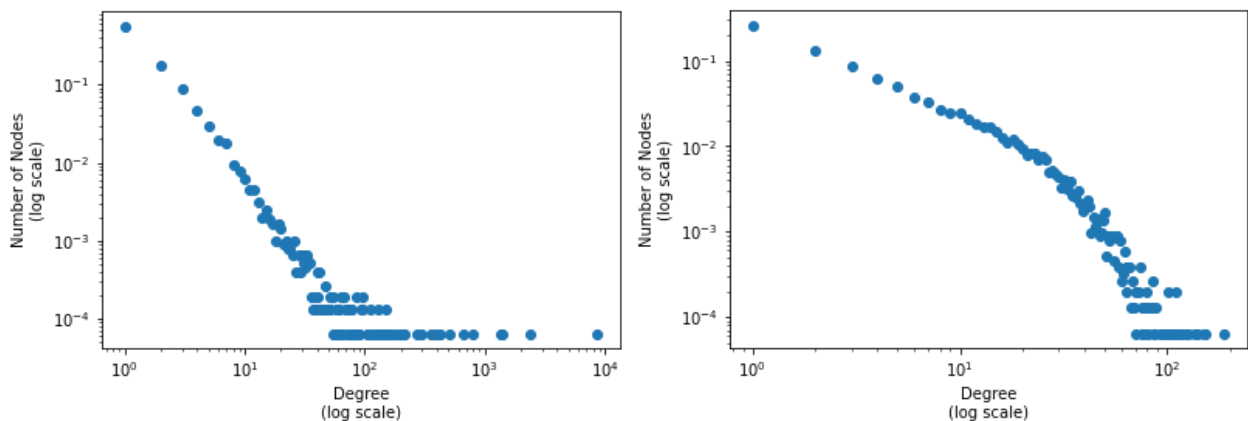


Figure 3-7. Distribution of Node Linkages. Log-log plot of Physician referral network (right) and random network (left)

These two distinct characteristics are known as growth and preferential attachment, and the Barabási-Albert model displays a scale-free network containing these two characteristics.

For the purposes of our study, we plotted a histogram with the node degree distribution, along with a log-log plot of the number of nodes by their degree to evaluate whether it follows

a power-law distribution based on the linearity of the plot. This intuition is based on the idea that, if the probability P that a node will connect to another one depending on the degree k ; hence P_k , follows a power law, by plotting P_k as a function of degree k on a log-log scale, a straight line of slope $-\alpha$ should be seen on the plot. While the node degree histogram seems to follow a power-law distribution, the log-log plot does not, given that there is no straight line as expected. However, because of the imprecise nature of estimating by visualization, we decided to implement the python package *power law* to statistically evaluate whether our distribution falls under a power law. We compared a random scale-free generated graph by *networkx*'s *scale_free_graph* function with our graph to test if it was considered a power law or exponential distribution. For this test, the log-likelihood ratio of the two distributions' fit to the data is considered. While our graph showed more similarity to a power law distribution, the significance of this test was not high enough, with a p-value of 0.07. In contrast, the random scale-free graph showed a significance value of 0.001.

In conclusion, there is debate on whether social networks are truly scale-free, and that might be the case in ours. This means that while there are very few nodes with many connections and most nodes with very few connections, our node degree distribution might not be all explained by a power-law distribution. In other words, the physician referral pattern is not entirely scale-free.

3.3.3.4.3 Small world

Small-world networks were first designed to address and construct networks akin to real-world networks. This form of the graph allows non-related physicians to be linked in a social world setting since the majority of nodes in its graphs are not neighbors but are easily reachable by a minimal number of hops. The Watts-Strogatz mechanism is the primary process for constructing small-world networks. The clustering coefficient and the average shortest path length are the two structural properties considered in this sort of network, with Watts-Strogatz

arguing that the model should have a high coefficient as well as a short average path. Thus, if we take one physician as an example, a high average of the clustering coefficient indicates that its neighbors are connected, implying that its contacts are acquainted with each other.

Chuankai An et al. (2017) in their work emphasize that the small world network "is defined as a network with greater than expected local connectivity and average path length smaller than expected in a comparable ER random network". To determine whether this referral system is a small world network, we can compare the clustering coefficient and the shortest path length of the data being used with a random Erdos-Renyi graph. To generate the random graph the same number of nodes as the current network (15,439) and a probability of 0.0006 are used. If we consider μ as the average degree of the physicians in the network and n as the number of nodes, we can calculate the probability as follows:

$$p = \frac{\mu}{n - 1}$$

After computing both graphs, we notice a significant difference in the clustering coefficient, with the random graph one being about 0.006 and the referrals network one being 0. The results found for the shortest path were 7.58 for the random Erdos-Renyi and 2.56 for the patient referral. Hence, it is possible to say that the network explored in this paper does not present characteristics of a small world network.

4. Methodology

Given that our goal is to extract the network structure and to explore whether any feature shows relevance over one another when it comes to patient referrals, we decided to implement a Graph Neural Network (GNN) unsupervised method to extract node embeddings. In short, a GNN follows the same steps as a Neural Network, with the exception that each layer uses a multilayer perceptron (MLP) on each one of the components of a graph. The goal of this process is to retrieve a learned node vector, known as node embeddings. Thus, for each node, we have

a vector with its representation. These layers are then stacked together, following the same Neural Network structure.

In terms of GNN predictive power, there are three distinct tasks that it can accomplish: node classification, link prediction, and graph classification. As a regular data classification task, if we wanted to predict a node label from a graph, this could be accomplished through a GNN. If needed for the problem, a regression task could also be used by adjusting properly the activation functions used.

Given that we want to extract meaningful insights that explain the referral patterns from Medicare patients in the US, we decided to implement a GNN unsupervised algorithm - GraphSAGE. We are implementing an unsupervised method because we are investigating whether there are clearly defined physician clusters that might affect the healthcare system's efficiency. This hypothesis would be confirmed if, after feeding the model, physicians that have similar embeddings with each other also have similar features.

Thus, we followed some steps to best achieve this project's purpose results. Firstly, we will define the features that would be used to answer the proposed business question in each specific question. After understanding the hypothesis and the features choices, we train an unsupervised GraphSAGE network as a way of obtaining the physicians' network simply through their features. Afterwards, we train a model to extract node embeddings as a method of capturing network topology, implying that doctors may be encoded as vectors using similarity. Furthermore, we reduce node embeddings dimensionality through t-SNE and UMAP to facilitate the 2D visualization of the patterns on the doctors' characteristics. The following steps included finetuning the model to get a perception of the best hyperparameters for it. Finally, with the k-means algorithm from the *sci-kit learn* library we defined clusters to extract insight results for the hypothesis.

4.1. Feature Selection

As referred above, the paper uses Medicare Physician & Other Practitioners - by Provider (2015) dataset that contains 73 different physician features to infer any patterns among connected doctors. Each section dedicated to solving a certain business question employs features that are suitable to approach the problem. Some of these features in subsequent sections are not included in the physicians features data set and are obtained from additional sources which will more deeply explained in the respective sections.

Besides the 30-day referral dataset, section 5, will also make use of the 90-day referrals time window dataset from the Medicare data, as a resource for the pattern comparison. Features such as the physician's specialties and percentages per disease will be key assets to answer this hypothesis. Section 6 will consider the total charges that the provider submitted for all services, the Medicare allowed amount for all provider services, and the total amount that Medicare paid after deductible and coinsurance amounts have been deducted for all the provider's line-item services (Centers for Medicare & Medicaid Services Data 2020). Moreover, it will factor into the analysis the average submitted charges per beneficiary, average Medicare paid amount after deductions, per beneficiary, and average submitted charges per service provided.

The primary goal of Section 7 is to infer how referral patterns are impacted by physician experience. Thus, it also includes physician experience, which is represented by Provider Enumeration Date from CMS NPI Files (NPI files)

4.2. GraphSAGE and Node Embeddings

The GraphSAGE algorithm was introduced by Hamilton, Ying, and Leskovec (2017), and is considered a distinctive algorithm from the rest of deep-walk-based algorithms because it is an inductive framework. In other words, it does not only consider the graph structure but is also capable of learning from rich attribute networks. GraphSAGE generates low-dimensional representations for each node of the graph, and after training, it can also generate these

embeddings on not seen data (Leskovec n.d.). As mentioned in the StellarGraph documentation, GraphSAGE's objective is: "Given a graph, learn embeddings of the nodes using only the graph structure and the node features, without using any known node class labels" (*Node representation learning with GraphSAGE and UnsupervisedSampler*). For the unsupervised GraphSAGE model, both types of nodes are generated, "positive" and "negative." These are generated based on the random walks that are considered by the algorithm and adjusted by the user. The two main aspects that would define these random walks are the number of random walks that the model would learn from each node, and the length of the walk, hence, the number of "hops" from the root node. Thus, the "positive" nodes are those that co-occur in the same random walk of the graph, and the "negative" ones are randomly selected from the graph. As a result, whenever a randomly selected node co-occurs within the same random walk, it is labeled as positive. This way, the model learns the nodes' information of its attributes and local neighborhood, retrieving a vector representation of the node characteristics within the graph. The hyperparameters of the algorithm that we considered for analysis were the number of random walks, the walk length, the minibatch size, the layer size, and the number of samples considered for each layer. The extracted embeddings are then plotted with different visualization techniques discussed in the next subsection.

After training and fitting the model, we generate the mapping of the nodes pairing, meaning that each physician node was turned into a low-dimensional space, allowing us (when visualizing) to extract meaningful insights based on their similarities.

4.3. Dimensionality Reduction

Regarding visualizations, before plotting the networks we used both t-SNE and UMAP to help with the high dimensionality of the data set and to reduce the number of random features to a 2D array of main variables. The dimensionality reduction is regularly used to better understand and interpret the data by simplifying it, and given that the data used is multi-

dimensional, it is wiser to reduce it to a lower dimension, to make possible the visualization and insights extraction.

The t-SNE algorithm allows to separate data that is nonlinear – cannot be separated by a straight line – and so it “models the probability distribution of neighbors around each point.” (Hoare, *How t-SNE works and Dimensionality Reduction*). As a result, it is an effective tool for interpreting high-dimensional sets such as this one. This study, on the other hand, leverages the UMAP technique, which uses the same ideas as t-SNE and is a powerful tool for visualizing large data sets. They both compute the distances between nodes and their neighbors and ensure that these distances are similar when the data is changed into a 2D space. In contrast, t-SNE converts a high-dimensional graph to low-dimensional space, whereas UMAP condenses the graph, which means that the UMAP tool rather than measuring it point by point, does not make a thorough estimation of the graph. This is valuable given that it will translate into a more accurate exhibition of the overall network.

In the end, when plotting both of the tools previously explained, the nodes of the same color will be expectantly clustered together, indicating a higher similarity of the embeddings.

4.4. Model iteration and Fine Tuning

Even though this research relies on the Unsupervised Sampler for modeling, finetuning its hyperparameters is still indispensable due to the lack of a target variable. As a consequence, based on the findings of the first model, the hyperparameters will be manually adjusted to improve accuracy and reduce loss. Several iterations must be tried and minor adjustments need to be made to some parameters to better understand their impact on the overall referral network. However, it is crucial to note that the learning rate will be constant - Adam rate of $1e-3$ - as well as the regularization - L2, and always including a bias term. This L2 regularization is chosen over L1 regularization because, in this situation, we require a parameter that works with codependent features rather than performing feature selection.

The Sigmoid function entails exponential computing, implying that it is more advantageous for large networks, and because accuracy is the measure of predictions when the true value equals the predicted, it is expected to be consistently high. The loss function, which accounts for the sum of all errors produced for each sample, is also supposed to decrease as the binary accuracy increases. ReLU differs from Sigmoid in that it does not activate all of its neurons at the same time, making this function less time-demanding and computationally easier.

4.5. Clustering methods

The K-Means was first mentioned by Hugo Steinhaus (1950), which evolved into an iterative process that is made up of “partitioning a set of n objects into $K \geq 2$ clusters, such that the objects in a cluster are similar to each other and are different from those in other clusters” (Ortega et al 2019). The steps for the k-means algorithm can be found in Figure 4.1.

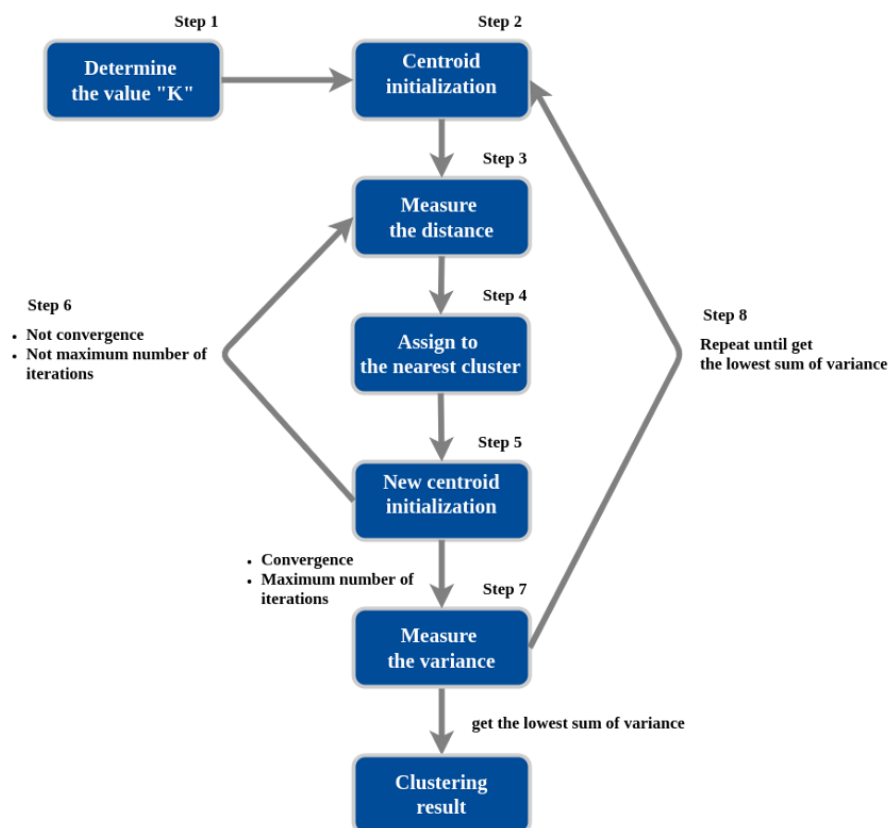


Figure 4-1. K-Means algorithm explained¹

¹ R, Arif. 2020. *K-means Clustering Overview*. Image. October 4, 2020. Medium. <https://medium.com/data-folks-indonesia/step-by-step-to-understanding-k-means-clustering-and-implementation-with-sklearn-b55803f519d6>

This clustering approach is based on a centroid-based algorithm, where the purpose is to discover data points that are closer together, in this case, to better classify and understand physicians. Thus, rather than pre-defined categories, K-Means classifies unlabeled data based on their features. It is crucial to understand what is the ideal number of clusters given the business problem. This number can be found using the elbow method, which works "as a cutoff point in mathematical optimization to decide at which point the diminishing returns are no longer worth the additional cost" (Sharma 2022). Thus, the number of clusters where the elbow bends is the optimal K. On another note, we can use the silhouette score that accounts for both inter and intra-cluster distances to evaluate the quality of the clusters created. This score ranges between -1 and 1, where 1 is the best value, representing that the data points within the cluster are close to one another and distant from the other clusters, and 0 indicates that there is an overlap of the clusters.

Two of the main disadvantages of K-Means clustering are that it is not robust against outliers given that it is based on the average, a metric highly affected when data contains too much noise and outliers; and that the center of the cluster does not necessarily represent a real data point from the dataset. As will be seen in the individual sections, some distributions of the data used as the input for the model follow right-skewed distribution, where many data points are closer to 0 and few data points contain high values. This difference is reflected when looking at the averages vs. the medians of each of the variables in consideration.

As such, some business questions would need to implement a different clustering algorithm that is more robust to outliers so that the output results are more precise. For this, an option is the K-Medoids clustering algorithm. This name was first given by Leonard Kaufman and Peter Rousseeuw through the Partitioning Around Medoids algorithm (Kaufman and Rousseeuw 1990). We will implement K-Medoids by using the *sci-kit learn* extra cluster library. As explained in *sci-kit learn* documentation, "KMedoids tries to minimize the sum of distances

between each point and the medoid of its cluster. The medoid is a data point (unlike the centroid) which has the least total distance to the other members of its cluster” (“Clustering with K-Medoids and Common-Nearest-Neighbors”). Thus, K-Medoids selects a real data point as the cluster center from the cluster based on who is the closest to the other members of the shared cluster. This is a particularly useful feature since, in practical terms, this data point represents the leading characteristics of this cluster and is worth considering for further investigation.

In terms of the algorithm used to determine cluster assignment, we will consider the default method from *sci-kit learn*, the "Alternate Method" (“Clustering with K-Medoids and Common-Nearest-Neighbors”). The alternate method works as follows. It will initialize with a K number of medoids, a method that depends on the approach selected in the *init* parameter (heuristic, random, or kmedoids++). Then, it will assign each data point to the closest medoid. This is followed by the update step, where it is reconsidered the medoid of the cluster and selected a new one. Lastly, this process iterates until there are no more changes regarding what data points are the medoids or until a user-specified maximum number of iterations is attained.

4.6.Example of Model Output

To get a glimpse of the output achieved through node embedding extraction, purely for the sake of illustration, we fitted the model with the feature that classified whether the physician was male or female (model loss: 0.6958, binary accuracy: 0.5593). Reducing the dimensionality using t-SNE and UMAP allowed us to obtain a visualization of our node embeddings (Figure 4.2). As depicted in Figure 4.2, the red embeddings represent female physicians, and the blue embedding represents male physicians. At first glance it seems uncomplicated to conclude that red embeddings cluster on the top left and are closer to each other, indicating that females are either close to each other within the network or represent similar roles regarding overall graph structure. To be precise, representing a similar role within the graph structure can be explained

with a similar example: if doctor A has made a total of 10 referrals, the referrals were made to 5 different doctors, resulting in an average number of referrals per patient of 2, and doctor B has the same average referral per patient count, then these doctors are said to represent a similar role within the graph structure. Similarly, blue embeddings cluster together that forces the same inference about male physicians.

Nevertheless, it is difficult to infer any actual clusters from the node embeddings, therefore, we use a K-means clustering method that groups or divides the features into clusters and minimizes cluster variance. With a silhouette score of 0.5, we obtain that our optimal number of clusters is 2. The proportion of male-to-female physicians was higher in cluster 0 - 90% male and 10% female. Cluster 1 comprised 52% male and 48% female physicians. Cluster 0 hints that male physicians in that cluster prefer to refer to male physicians.

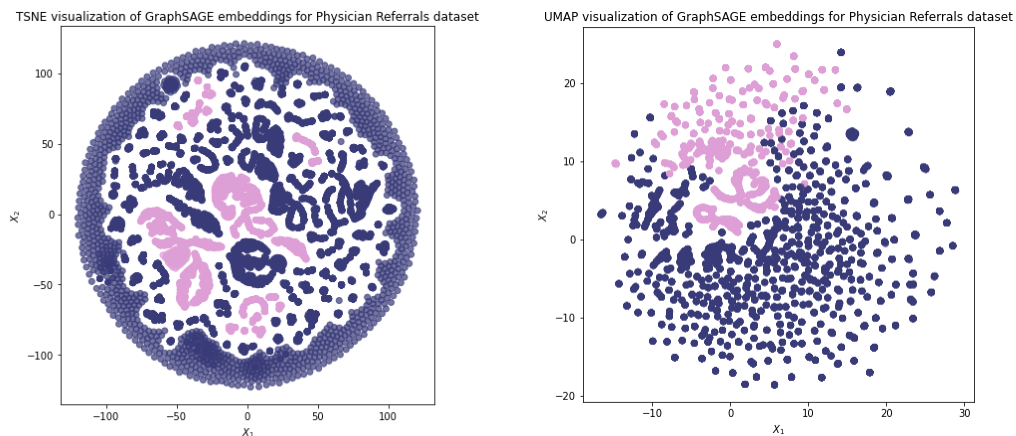


Figure 4-2. TSNE visualization of GraphSAGE embeddings. Dimensionality reduced using TSNE (left) and UMAP (right) methods

5. Exploring the Relationship Between Physician Referral Patterns Primary Care Access and Healthcare Spending

This section will explore the relationship that exists, if any, between the physicians' referral network and the CMS corresponding financial data available for 2015. Specifically, we will study the significance that the referral network embeddings provide us about financial variables between physicians. Our goal is to understand whether the referral network affects Medicare spending for a physician's sub-network. Our analysis will consider the total charges submitted by providers for all services, the Medicare allowed amount for all services, and the total amount that Medicare paid after deductible and coinsurance amounts have been deducted for all provider's line item services. (Centers for Medicare & Medicaid Services Data 2020). Moreover, we aggregated some of these features to provide richer information to our model: average submitted charges and average Medicare paid amount, per beneficiary; and average submitted charges per service provided.

Furthermore, after extracting the node embeddings from the model, we will use a clustering technique to analyze features based on each cluster produced. To account for geographic-based price variations, we will explore the physicians' Hospital Referral Regions' (HRRs) spending rates, rural-urban commuting area (RUCA) codes and healthcare quality measures at the HRR-level. In the United States, HRRs "represent regional health care markets for tertiary medical care. Each HRR contains at least one hospital that performs major cardiovascular procedures and neurosurgery" (Dartmouth Atlas of Health Care n.d., under "Research Methods FAQ"). RUCA codes "classify U.S. census tracts using measures of population density, urbanization, and daily commuting" (USDA Economic Research Service (ERS) 2020). All data used within this analysis is publicly available by the Dartmouth Atlas of Health Care (DAHC), the CMS and the USDA ERS. Only 2015-level data was considered for the study to guarantee consistency within the results. The HRR-level data will serve as a baseline comparison of our results and may provide information regarding these regions' delimitations and levels of healthcare quality, considering physicians' referrals in that area. Finally,

we will test for statistically significant differences across these clusters' spending and referral-based variables.

Our motivation to investigate this area arises from the problem of increasing Medicare spending over the last few years. Medicare accounted for 3% of the federal spending in 1972, while as of now accounts for 13% of the total. By 2052 Medicare spending could account for 19% of the federal spending, *ceteris paribus* (Peter G. Peterson Foundation 2022). Our insights from the physicians' referral network may provide valuable information for future policymaking regarding Medicare. If referrals influence Medicare spending, solutions to optimize spending without affecting healthcare quality should be a priority, such as an AI-based referral system (Han 2018).

5.1. Relevant Work

Previous studies have examined the relationship between Medicare spending and physicians' referrals through different methods. Recently, Skinner et al. (2022) researched the relationship between rurality, healthcare spending, and quality. Using the HRRs as a unit of observation, they concluded that rural areas might have a lower Medicare reimbursement rate because there are fewer specialists than PC providers. Likewise, the more frequent use of lower-value services and a higher ratio of SC to PC potentially caused the higher per capita spending in HRRs.

Some studies have explored community detection algorithms to improve healthcare services based on physicians, hospitals, and patients in the area (Landon et al. 2013; Wang and Wang 2020). Landon et al. (2013) tried to identify physician communities that could be the basis for defining Accountable Care Organizations (ACOs), a voluntarily formed group by doctors to attend their communities in a coordinated manner (CMS 2021). They showed that, for some regions, a physician-based network would provide better outcomes than hospital-based networks.

Jia et al. (2020) conducted a pilot study in Florida to analyze whether HRRs are still the best unit to consider in the health markets or if any other delimitations should be made to guarantee better outcomes. Fisher et al. (2009) investigated the Medicare provider's spending as their key variable to

increase savings in this area for the federal budget. Fisher's study shares our motivation in uncovering local referral patterns that could translate to better policymaking by decreasing Medicare spending growth without impacting quality. Findings showed that higher-cost areas have more referrals and use of hospitals, even though the care quality is similar to that in lower-cost areas. Further studies hypothesize that physicians' decision to refer a patient to a higher-cost provider is possibly influenced by the physicians' employer, given the rates of vertical integration across hospitals in the U.S. (Whaley 2021)

5.2. Data Collection

This section will overview the data we will use as input for our model and to enrich our analysis. Three datasets will integrate this part: the referrals dataset within the 30 days, the NPI attributes dataset, and public research data from the DAHC. For the latest, we are using Medicare Reimbursements (Appendix, Table 1) , and Primary Care Access and Quality Measures (Appendix, Table 2), each by HRR and for the year of study, 2015 (Dartmouth Atlas of Healthcare Data 2022). We used additional supplementary datasets for the crosswalk between the physician's zip code and HRR ("Dartmouth Atlas of Healthcare Data, "Supplemental Data" 2021). For the primary practice business location and zip code of each provider, we extracted all the NPIs from the subset of our study and matched them in the NPES Data File (CMS 2021, under "Data Dissemination"). From this data, we included only physicians that had a matching zip code to a Florida HRR. For the referrals' dataset already mentioned under Section 3, we included whether each referral was within the HRR as a binary column. This could provide insights to evaluate the relevancy of location for referrals.

Regarding the Medicare Reimbursements rates by HRR, we obtained the number of total Medicare enrollees for that year; and the spending for total, hospital % skilled nursing facility, physician, outpatient facility, home health agency, hospice, and durable medical equipment reimbursements, all per enrollee for 2015. We considered only price, age, sex & race-adjusted rates, given that prices of areas such as Miami may differ from other areas in the state. Because Medicare

spending varies across regions (Gottlieb et al 2010), using the latter would guarantee a more precise analysis. As referred by the DAHC, "Among the 306 hospital referral regions in the United States, price-adjusted Medicare reimbursements varied twofold in 2016, from about \$7,400 per enrollee in the lowest spending region to more than \$13,000 in the highest spending region." (Dartmouth Atlas of Health Care). This standardization factors in the price and wage variations so that any disparities we observe are not due to region differences. Moreover, it would help to discern if price differences are observed due to the volume of services requested. Further information on the standardization process is available in their technical report (Austin et al 2020).

We also collected Primary Care Access and Quality measures to address quality against the spending of a specific region and cluster of physicians. This would enrich our analysis when investigating the reason behind higher Medicare spending in certain areas. These metrics include the number of diabetic patients per region, average percentages of diabetes screening exams such as eye examination, hemoglobin tests, and blood lipid tests annually, leg amputations per 1,000 enrollees, and the average percent of female Medicare enrollees aged 67-69 having at least one mammogram over two years. These are considered measures to evaluate primary care quality per region, given that their frequency may help avoid more serious and/or chronic illnesses. Furthermore, for each HRR, we assigned the proportion of rural areas by total zip codes. We first extracted each RUCA code by zip code and matched it with the HRR. Because the RUCA codes are from 1 to 10, with 1, 2, and 3 being urban, and the rest rural, we assigned the rural proportion based on how many zip codes from the total per HRR were rural (Appendix, fig.1).

From the NPI attributes dataset, key financial attributes are total submitted charges, total Medicare allowed amount, and total Medicare paid amount. We will also use total beneficiaries and total services provided. For PC and SC providers, we calculated the average number of referrals made and received, respectively. We also added the number of unique physicians with whom providers had a referral relationship; and number of out-of-HRR referrals based on the providers' practice location.

5.3. Exploratory Data Analysis

All financial variables showed a right-skewed distribution, with many data points close to smaller values and a tail distribution with few values with high amounts (Appendix, Figure 2). More details can be found in Table 5.1. For example, the average submitted charges per patient, for each physician, can range from \$13 to \$45,804, where the median does not reach \$1,000.

Table 5-1. Summary Statistics of Financial Variables

CMS NPI Features	Descriptive Statistics						
	mean	std	min	25%	median	75%	max
<i>Total Medicare Payment Amount</i>	285,245	464,491	277	95,310	171,033	302,726	10,193,799
<i>Total Submitted Charges</i>	874,311	1,518,662	684	261,552	497,023	924,841	31,606,718
<i>Total Services</i>	10,373	40,712	12	1,419	2,890	6,130	1,572,970
<i>Total Beneficiaries</i>	886	1,024	11	348	585	999	20,414
Aggregated Features							
<i>Avg Submitted Charges per Patient</i>	1,213	1,841	13	509	797	1,258	45,804
<i>Avg Medicare Payment per Patient</i>	406	518	6	184	290	440	11,284
<i>Avg Submitted Charges per Service</i>	223	272	3	97	153	235	6426

Regarding within and out-of-HRR referrals, most physicians have within-HRR referrals as expected. Half of the primary care providers send their patients to a range from 1 to 7 unique specialists within 30 days. In contrast, half of the specialists receive their patients from 1 to 3 primary care providers (Table 5-2). This could indicate either missing primary care physicians in the area or some unknown bias in the referral system.

Table 5-2 Summary Statistics of Referral-Based Variables

CMS NPI Features	Descriptive Statistics							
	count	mean	std	min	25%	50%	75%	max
Unique # of specialists that PCPs send their patients to	5318	13	16	1	2	7	18	188
Unique # of PCPs that specialists receive their patients from	10121	7	9	1	1	3	8	81
Referrals within HRR	15439	7	11	0	1	3	9	182
Referrals out of HRR	15439	2	4	0	0	0	1	81

5.4. Research Question and Hypothesis

Our key concern is to investigate if there is any relationship between referrals and Medicare spending. Can we detect if the physician referral network explains part of the price variations across

hospital referral regions after accounting for age, sex, race, and price; with the aid of node embeddings? Are there any biases in the referral network that could lead to higher Medicare spending, avoidable otherwise? These questions could help answer trivial policymaking issues and save federal budget money with the aid of Artificial Intelligence. If the current manual referral decision is biased towards higher spending, lack of efficiency across physicians, or lesser healthcare quality, what impact could an AI-based automatized referral process have on the healthcare system efficiency and U.S. federal budget? While this question remains to be answered, it represents our motivation to pursue this investigation. Hence, we hypothesize that a relationship exists between the physicians' referral network patterns and Medicare spending. This study does not aim to exhibit what type of relationship, if any, that might be. Given the unsupervised nature of the algorithms we use, we will attempt to find patterns that could plausibly explain higher Medicare costs and evaluate whether physicians with certain referral patterns relate to specific levels of healthcare quality and spending.

5.5. Model Results

Regarding the GNN, we limited testing the walk length from 2 to 5 since a longer walk would probably start losing information about the local neighborhood and obtaining more from the overall network, which may result in more similar embeddings. We selected 2 as number of walks and 5 as our walk length. We used two layers with a size of 50 each, and a dropout rate of 0.1. The sigmoid activation function performed best for our model when compared to ReLu. As node attributes, we considered average submitted charges per service, total Medicare paid amount and average submitted charges per patient. Since we use an unsupervised algorithm, we cannot conclude whether a variable is dependent or independent. Hence, our goal is limited to understand what the embeddings communicate about the relationship between referral patterns and financial features in Florida. After hyperparameter tuning, we selected the parameters mentioned and obtained a GNN binary accuracy of 0.6242 and a binary cross-entropy loss of 0.6480 after training for 5 epochs.

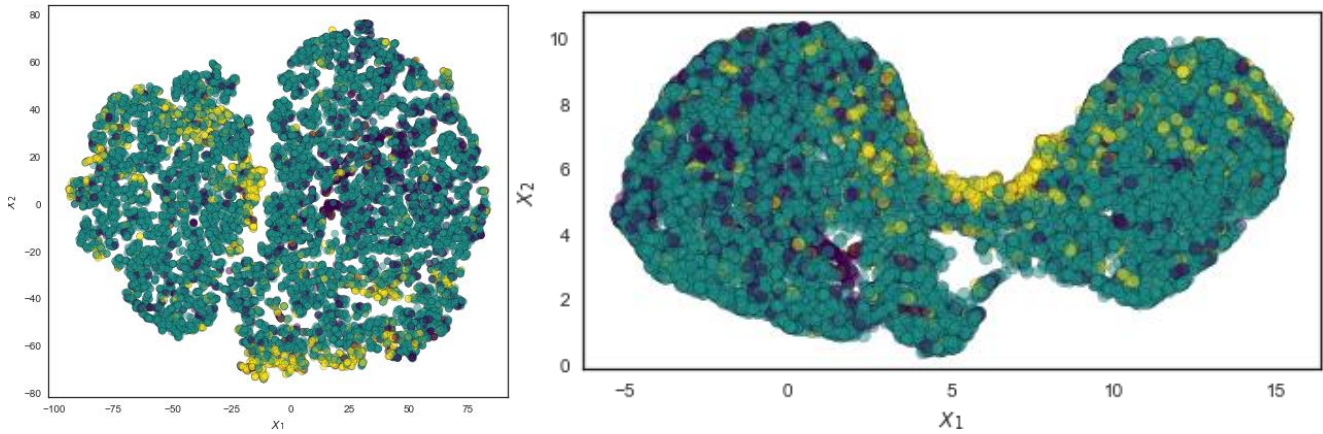


Figure 5-1. Node embeddings visualization using TSNE (left) and UMAP (right), where yellow represents the top 10% of physicians with highest average submitted charges per services.

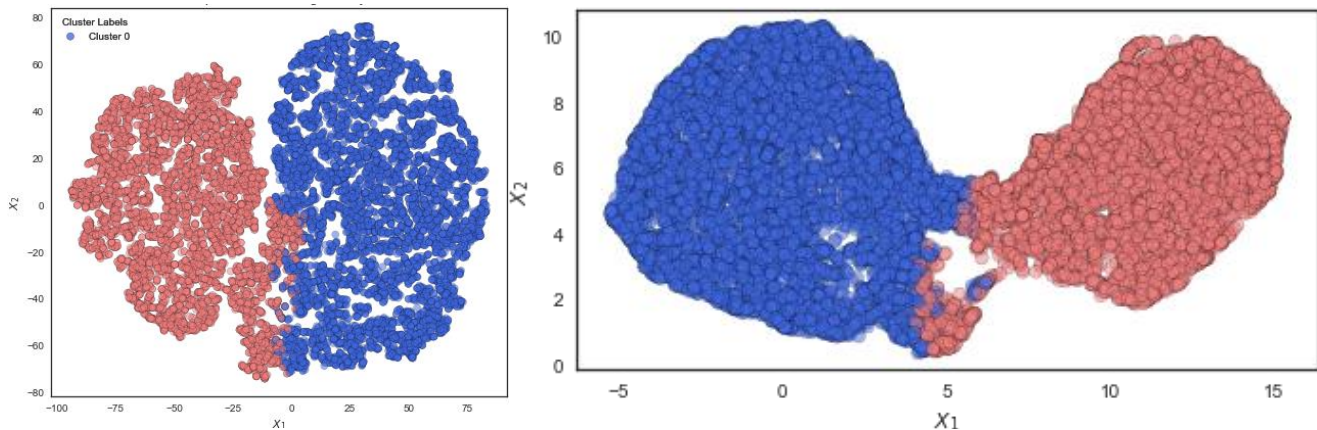


Figure 5-2. Node embeddings visualization using TSNE (left) and UMAP (right), where blue represents “Cluster 0” and coral represents “Cluster 1” as labeled by K-Medoids clustering.

To get a grasp of our embeddings, we visualized them by reducing dimensionality through T-SNE and UMAP. From a vector of size 50 we reduced to size 2, and plotted the new two-dimensional data on a scatterplot. Additionally, we added average submitted services as color, with yellow as top 10%, purple the lowest 10%, and green the remaining data points in between. (Figure 5-1)

Concerning the clustering results, we selected K of 2 for the K-Medoids clustering algorithm. Given the skewed distribution of the financial data, we decided that K-Medoids would guarantee better performance because of its robustness against outliers. To decide on the number of clusters, we calculated the inertias and silhouette scores for K ranging from 2 to 20 given that there are 18 HRRs in Florida is 18, and clusters formed based on these regions was possible. Yet, based on our business understanding, we suspect that a K of 2 or 3 may group physicians according to high, low and medium

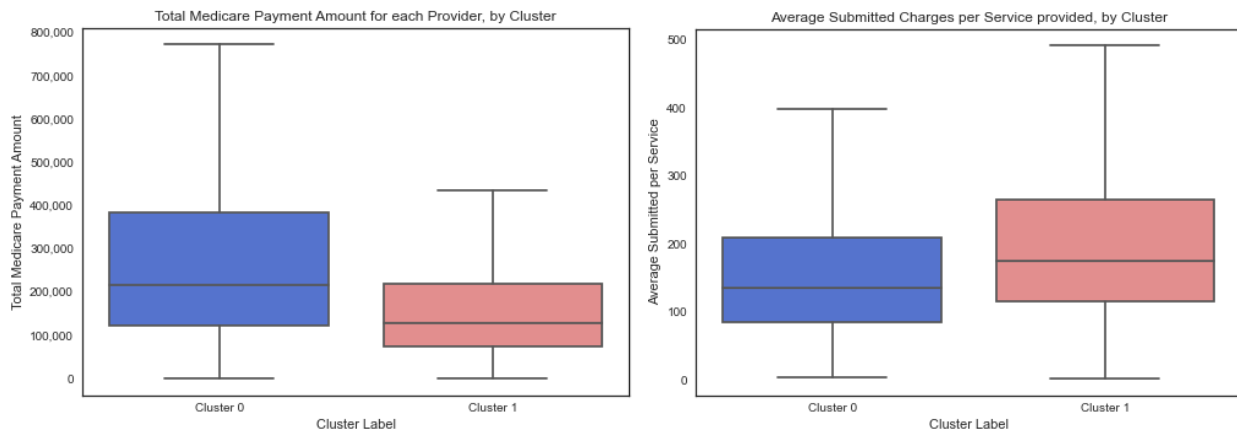


Figure 5-3. Boxplot distributions of Average Submitted Charges per Service and Total Medicare Payment Amount each Provider, by Cluster (outliers were excluded from the results)

charges and/or Medicare payment, with this being also suggested by the two-dimensional embeddings visualization (Figure 5.1) The average silhouette score of our results was 0.63 (Appendix, fig 3), with 56 % of physicians assigned to cluster 0 and 44 % to cluster 1. We suspected that 2 clusters may represent those with higher submitted charges but not necessarily higher Medicare payments, against those with lower submitted charges but higher Medicare coverage. This might result from physicians' specialties and services. We also contemplated a K of 3, yet the drop in the silhouette score was significant enough for us to choose K of 2 instead. We believe this way our analysis would be more precise since clusters are better defined. Two-dimensional node embeddings were plotted again, coloring by cluster label (Figure 5.2)

5.6. Results Analysis

We plotted distributions for the average submitted charges per service (fig. 5-3, right) and per patient by cluster (Appendix, fig. 5); we conclude they are lower in cluster 0. However, the total Medicare payment amount median and maximum values are higher for cluster 0 (fig. 5-3, left). It seems that physicians in cluster 0 submit higher charges than physicians in cluster 1 even though Medicare payment is lower. Furthermore, by observing the distributions plotted as histograms in log-scale (fig. 5-4; Appendix, fig 4) we can conclude that both averages submitted charges per patient and service from cluster 0 seem shifted to the left when compared to those distributions from cluster 1.

However, the total Medicare payment seems shifted to the right for cluster 0 when compared to cluster 1. Thus, there is a higher gap between submitted charges and Medicare coverage for those patients seeing physicians in cluster 0 than in cluster 1. These findings may occur if primary care physicians were clustered in cluster 0 and specialists in cluster 1. Nonetheless, in cluster 1 we have a total of 38.7 % of primary care providers; in cluster 0, 31 %. Fig. 5-5 provides a distribution count. The top specialties in both clusters seem to have a similar proportion, making it harder to state that one cluster represents "more expensive" or "less urgent" specialties than the other when trying to discover the reasoning behind the financial differences found. Table 3 from Appendix provides the % of specialties per cluster, as well as Gender proportions per cluster, which are also similar.

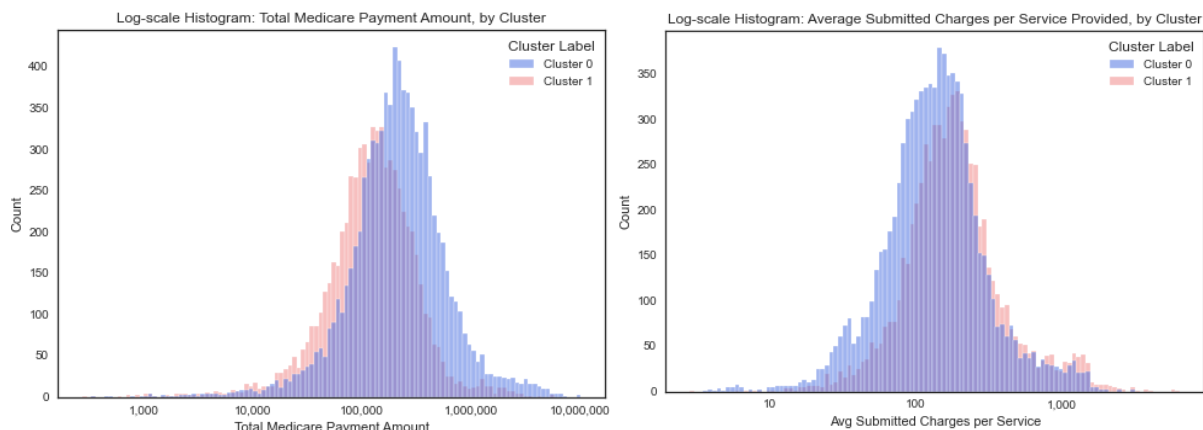


Figure 5-4. Log-scale histograms for Total Medicare Payment Amount and Average Submitted Charges per Service provided, by cluster.

Other main factor that may explain this discrepancy is rurality. Possibly, Medicare expenses are higher in urban areas. As expected, cluster 1 has 3% of physicians from rural areas, while cluster 0 only has 0.9% (Appendix, fig. 6, 7 and 8.) Since the majority of physicians are from urban areas (approximately 98%), it is challenging to state if this is a significant difference. Hence, we need to carefully interpret the results, given that other variables that are not being held equal may interfere. Additionally, we ought to look at patients and services in each cluster, to determine if the volume of these are different. Log-scale histogram shows cluster 0 shifted to the right, meaning it is the cluster with higher volume of services and patients (Appendix, fig. 9). This may explain why it is the cluster with higher total Medicare payment amount.

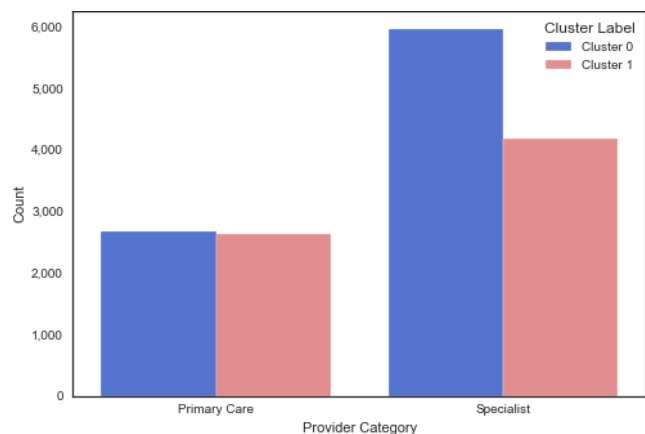


Figure 5-5. Count of Primary Care providers and specialists, by cluster

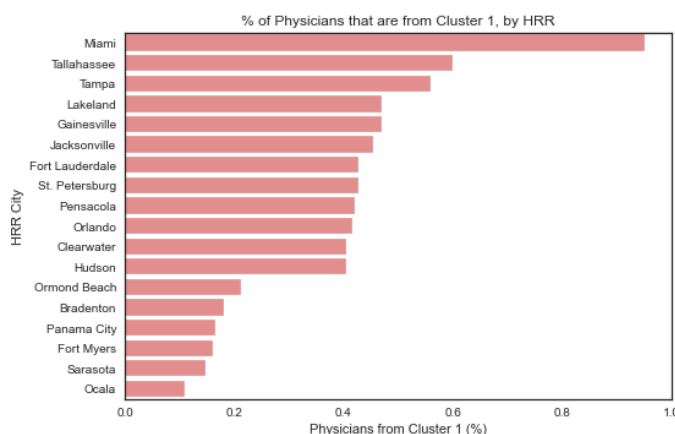


Figure 5-6. Percentage of Physicians for that region that were assigned to cluster 1

Given that the only available data at the geographic level was by HRR, we will analyze whether there is any relationship between each HRR Medicare reimbursement rate, primary care access, and quality measures, considering the clusters found and the HRR where these physicians work. After calculating the clusters' proportions within each HRR, the two most significant regions were Miami, with 95.2% of physicians found in cluster 1; and Ocala, with 89.2% of physicians found in cluster 0 (fig. 5-6). Concerning total Medicare reimbursements per enrollee, adjusted as explained under Section 5.2, Miami receives close to \$13,000 while Ocala receives around \$10,000. However, physician reimbursements per enrollee are not very different between these regions. (Table 5-3)

Table 5-3. Price, age, sex & race-adjusted Medicare reimbursement rates 2015 Miami & Ocala

Medicare Reimbursement Rates 2015 (Price, age, sex & race-adjusted)	HRR City	
	Miami	Ocala
Medicare enrollees	145,033	104,695
Total Medicare reimbursements per enrollee (Parts A and B)	13,109.2	10,236.6
Hospital & skilled nursing facility reimbursements per enrollee	5,191.65	3,993.02
Physician reimbursements per enrollee	3,701.58	3,915.66
Outpatient facility reimbursements per enrollee	1,556.16	1,118.49
Home health agency reimbursements per enrollee	1,593.65	615.65
Hospice reimbursements per enrollee	568.85	366.82
Durable medical equipment reimbursements per enrollee	211.54	208.39

Related to the quality measures, it appears that Ocala performs better overall, with higher percentages of diabetes-related tests, less percentage of leg amputations and higher percentage of older females having at least one mammogram every 2 years (Table 5-4). Hence, even though Miami has a higher total Medicare reimbursement per enrollee, this does not seem to translate into higher healthcare quality. Miami contains 95% of physicians from cluster 1. On average, per service and

patient, those physicians submit higher amounts of charges, but total Medicare payment is lower. Yet, the average Medicare payment per patient has the same median for these two clusters. If physicians are not reimbursed more in one region, Miami has a higher reimbursement rate, physicians that charge more, and apparent lower quality care, what explains either the lower quality or the higher submitted charges in Miami against Ocala? In contrast, Ocala contains a majority of physicians from cluster 0, those that submit fewer charges. Even though Ocala receives fewer Medicare reimbursements per enrollee than when compared to Miami, our analysis suggests that it performs better in terms of primary care access and quality. Even more intriguing, Miami has a total of 4.2 primary care providers per 1,000 enrollees; Ocala only has 1.6.

Table 5-4. Primary Care Access and Quality Measures for Medicare Enrollees, 2015 (Miami & Ocala)

Primary Care Access and Quality Measures	HRR name	
	Miami	Ocala
Average annual % of diabetic Medicare enrollees age 65-75 having hemoglobin A1c test	85.91	88.9
Average annual % of diabetic Medicare enrollees age 65-75 having eye examination	64.27	72.55
Average annual % of diabetic Medicare enrollees age 65-75 having blood lipids (LDL-C) test	86.04	89.37
Average % of female Medicare enrollees age 67-69 having at least one mammogram over a two-year period	56.75	75.22
Leg amputations per 1,000 Medicare enrollees	0.69	0.46
Total Mortality: ASR-adjusted % of deaths among Medicare enrollees	3.5	4.06
Primary care providers by 1,000 Medicare enrollees	4.25	1.64
Female Medicare enrollees % aged 67-69	7.7	8.9
Diabetic Medicare enrollees % aged 65-75	10.56	11.19
Rural zip codes %	7.36	6.06

Table 5-5. Referral Measures by Cluster

Referral-Based Metrics	Cluster Label							
	Cluster 0				Cluster 1			
	mean	50%	75%	max	mean	50%	75%	max
Unique # of specialists that PCPs send their patients to	15.4	9	22	188	9.94	5	14	114
Unique # of PCPs that specialists receive their patients from	7.53	4	10	81	5.43	2	6	66
Referrals within HRR	8.25	3	11	182	5.87	2	7	104
Referrals out of HRR	1.73	0	1	81	1.31	0	1	59

For referral-related patterns found in each cluster, the unique number of physicians that each cluster member sent/received their patients to/from within 30 days, on average, is higher in cluster 0. Within this cluster, PCPs send patients to 15.4 unique specialists, and specialists receive referred patients from 7.5 different PCPs, on average (Table 5-5). Thus, it appears that cluster 0 interacts in a

more widespread manner than cluster 1.

Table 5-6. Results and Statistical Significance

Variable	Cluster 0	Cluster 1
Average submitted charges per service	Lower ^a	Higher ^a
Average submitted charges per patient		
Total Medicare payment amount	Higher ^a	Lower ^a
Primary Care Providers	31 %	38.7 %
Physicians from rural areas	0.9%	3%
HRR with most physicians from this cluster	Ocala (89.2%)	Miami (95.2%)
Unique physicians interacted with	Higher ^a	Lower ^a
Patient referrals	Lower ^a	Higher ^a
Proportion of referrals out of HRR	Lower ^a	Higher ^a
Average Medicare payment amount per patient	Not different ^b	

^aThe results are significant at a significance level of 0.01 determined by Mann-Whitney U Test

^bEven though the Total Medicare Payment Amount was found to be significant, after accounting for total patients and considering the average, difference was not statistically significant (p-value = 0.08)

To conclude, the implementation of Graph Neural Networks and K-Medoids clustering led us to discern differences between healthcare providers and analyze their relationship with other demographic measures. Cluster 0 physicians showed lower submitted charges per patient/service, and fewer referrals overall and out of HRR. Yet, they interacted with more physicians; thus, their network was wider. In contrast, cluster 1 physicians showed higher submitted charges per patient/service, and more referrals overall and out-of-HRR, yet they interact with a smaller range of physicians; thus, their physician network was narrower. The most representative region of Cluster 0 was Ocala; and of Cluster 1, was Miami. Ocala's total reimbursements and most of the reimbursements' subcategories were lower than those in Miami; however, healthcare quality metrics considered for our study were overall lower for Miami – the HRR with 95 % of its physicians from cluster 1. Based on our analysis, we conclude that cluster 1 physicians seem less efficient than those from cluster 0, and one of the reasons could be that their network is smaller. Table 5-6 shows a summary of the results found and their statistical significance. Given the absence of normality in our data, we implemented the Mann-Whitney U test in each of the variables of the table to compare across clusters. Many variables may interact with these results that were not considered for this analysis, such as total population, economy, and education rates per region; physicians' employees; hospitals' management; among others. Hence, results should be interpreted cautiously.

6. Discussion

Given that referral is a human-based decision, our work is motivated by exploring if there are any biases affecting them. Physician referrals that are unnecessarily burdensome affect the efficiency and cost of the healthcare system. Our research analyzes these overlooked problems.

In this study we dissected the physician referral network for 2015 from Florida. In Section 5, we analyze different referral periods to uncover whether there are any significant differences when considering specialties. Our belief is that there might be differences in specialists collaboration in different timeframes, since there might be changes in the healthcare system that our data cannot control for. Yet, it is also possible that healthcare specialists have different patterns of collaboration based on their area of expertise explaining specialty-based findings. Additionally, chronic diseases may affect physician collaboration, since long-term medical conditions require constant check-ups. Thus, various healthcare specialists may require to cooperate in a way that is unique for patients with these type of diseases. For Section 6 we considered multiple financial variables from the physicians, such as submitted charges per patient and Medicare payments. Besides, we factored into this analysis the Medicare reimbursement rates, as well as PC access and quality measures for each HRR, considering that demographic factors are likely to intervene. The motivation was to uncover if any relationship exists between physician referrals and Medicare spending. Findings showed that physician submitted charges and their referral patterns are related to Medicare spending, PC access and quality measures of the region. However, given the unsupervised nature of our study, we are not able to conclude in what way variables affect one another. For Section 7, we explore the relationship between experience and physician referral patterns. Understanding this relationship is useful in identifying barriers to establish effective referrals and collaboration between physicians. Experienced physicians are responsible for passing information and practice experience to their younger counterparts, however, they might prefer to refer to an experienced

specialist because the latter may have more experienced or an established relationship that eases collaboration. Policymakers, educators, and healthcare organizations may find this information to be useful in developing plans and solutions to help doctors raise the standard of care and identify possible disparities in the healthcare system.

Concerning limitations, our work was based on the referral dataset from 2015 based on Medicare claims. As a result, physicians that were involved in the study were Medicare participants with claims from Medicare enrollees, and do not represent the overall population. Additionally, we acknowledge that each state may behave differently and results may not be applicable to a national level. Moreover, the number of referrals considered for the analysis were based on the assumption that if a patient visits two providers within our period of study, it would count as referral. This issue arises from the inability to disclaim patient records. Moreover, HRR-level rates may be too broad of a metric to understand primary care access and healthcare quality, with the possibility that more local metrics would add precision to our analysis. Similarly, factors such as education, race, population levels and income specific to the area of study have not been included.

Since we were limited regarding computing processing power, future work with appropriate equipment should explore the algorithms considered and reach a solution close to optimal. Further areas could be integrated into the study, especially states that are markedly different with one another. Additionally, access to patient claims data would provide more precision into the analysis. Regarding GNNs, a comparison of performance across other GNNs such as attri2vec, node2vec (etc.) would provide more insights and robustness to our findings.

Overall, the paper provides useful insights for healthcare organizations to improve collaboration between physicians and develop strategies to improve patient care. By better understanding referral practices of primary care specialists, organizations can improve coordination of care.

References

- “42 U.S. Code § 234 - Health Care Professionals Assisting during a Public Health Emergency.” Legal Information Institute. Legal Information Institute, June 24, 2019. https://www.law.cornell.edu/uscode/text/42/234#d_2.
- Akbari-Sari, Ali, Maryam Seyed-Nezhad, and Batoul Ahmadi. “Factors Affecting the Successful Implementation of the Referral System: A Scoping Review.” *Journal of Family Medicine and Primary Care* 10, no. 12 (December 10, 2021): 4364. https://doi.org/10.4103/jfmprc.jfmprc_514_21.
- An, Chuankai, A. James O'Malley, Daniel N. Rockmore, and Corey D. Stock. “Analysis of the U.S. Patient Referral Network.” *Statistics in Medicine* 37, no. 5 (2017): 847–66. <https://doi.org/10.1002/sim.7565>.
- Austin, Andrea M.; Gottlieb, Daniel J.; Carmichael, Don; Chakraborti, Gouri; Sutherland, Jason M.; Andrews, Kathryn Gilman; Raymond, Stephanie R.; Tomlin, Stephanie; Zhou, Weiping; Song, Yunjie; Skinner, Jonathan, 2020, "Technical Report: A Standardized Method for Adjusting Medicare Expenditures for Regional Differences in Prices", <https://doi.org/10.21989/D9/VJE2ZM>, Dartmouth Dataverse, V2
- “Avoidable Mortality (Preventable and Treatable).” Avoidable mortality (preventable and treatable) | Health at a Glance 2019 : OECD Indicators | OECD iLibrary. Accessed October 18, 2022. <https://www.oecd-ilibrary.org/sites/3b4fdbf2-en/index.html?itemId=%2Fcontent%2Fcomponent%2F3b4fdbf2-en>.
- Barnett, Michael L., Bruce E. Landon, A. James O'Malley, Nancy L. Keating, and Nicholas A. Christakis. “Mapping Physician Networks with Self-Reported and Administrative Data.” *Health Services Research* 46, no. 5 (2011): 1592–1609. <https://doi.org/10.1111/j.1475-6773.2011.01262.x>.
- Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler. “The Effect of Hospital/Physician Integration on Hospital Choice.” *Journal of Health Economics* 50 (2016): 1–8. <https://doi.org/10.1016/j.jhealeco.2016.08.006>.
- Barnett, Michael L, Nicholas A Christakis, James O'Malley, Jukka-Pekka Onnela, Nancy L Keating, and Bruce E Landon. “Physician Patient-Sharing Networks and the Cost and Intensity of Care in US Hospitals.” *Medical care*. U.S. National Library of Medicine, February 2012. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3260449/>.
- Broido, Anna D., and Aaron Clauset. “Scale-Free Networks Are Rare.” *Nature News*. Nature Publishing Group, March 4, 2019. <https://www.nature.com/articles/s41467-019-08746-5>.
- Carlin, Caroline S., Roger Feldman, and Bryan Dowd. “The Impact of Hospital Acquisition of Physician Practices on Referral Patterns.” *Health Economics* 25, no. 4 (2015): 439–54. <https://doi.org/10.1002/hec.3160>.
- Centers for Medicare & Medicaid Services (CMS). (2013). *National Provider Identifier Standard (NPI): Educational Resources: Guidance Portal*. National Provider

- Identifier Standard (NPI): Educational Resources | Guidance Portal. Retrieved December 4, 2022, from <https://www.hhs.gov/guidance/document/national-provider-identifier-standard-npi-educational-resources-0>
- Centers for Medicare & Medicaid Services Data. “Medicare Physician & Other Practitioners - by Provider.” Centers for Medicare & Medicaid Services Data, July 27, 2022. <https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners/medicare-physician-other-practitioners-by-provider>.
- “Clustering with K-Medoids and Common-Nearest-Neighbors¶.” Scikit. Accessed December 4, 2022. <https://scikit-learn-extra.readthedocs.io/en/stable/modules/cluster.html#k-medoids>.
- CMS. “Accountable Care Organizations (ACOs),” December 1, 2021. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO>.
- CMS. “Data Dissemination.” CMS. Accessed December 4, 2022. <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination>.
- CMS. “Referral Data Faqs.” CMS, December 1, 2021. <https://www.cms.gov/Regulations-and-Guidance/Legislation/FOIA/Referral-Data-FAQs>.
- Congressional Budget Office. “The Budget and Economic Outlook: 2022 to 2032,” May 25, 2022. <https://www.cbo.gov/publication/57950#:~:text=In%20CBO's%20projections%2C%20assuming%20that,by%203.1%20percent%20this%20year>.
- Dartmouth Atlas of Health Care. “Research Methods.” Dartmouth Atlas of Health Care, July 13, 2018. <https://www.dartmouthatlas.org/research-methods/>.
- Dartmouth Atlas of Health Care. “FAQ.” Dartmouth Atlas of Health Care, December 6, 2021. <https://www.dartmouthatlas.org/faq/#research-methods-faq>.
- Dartmouth Atlas of Healthcare Data. “Medicare Reimbursements.” Dartmouth Atlas DATA, February 3, 2022. <https://data.dartmouthatlas.org/medicare-reimbursements/>.
- Dartmouth Atlas of Healthcare Data. “Selected Primary Care Access and Quality Measures.” Dartmouth Atlas DATA, March 30, 2022. <https://data.dartmouthatlas.org/primary-care/>.
- Dartmouth Atlas of Healthcare Data. “Supplemental Data.” Dartmouth Atlas DATA, June 10, 2021. <https://data.dartmouthatlas.org/supplemental/#crosswalks>.
- Dijkstra, E. W. “A Note on Two Problems in Connexion with Graphs.” *Numerische Mathematik* 1, no. 1 (1959): 269–71. <https://doi.org/10.1007/bf01386390>.
- Disney, Andrew. “Social Network Analysis: Understanding Centrality Measures.” Cambridge Intelligence, October 27, 2022. <https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/>.

- Duarte, Regina, Qiwei Han, and Claudia Soares. “Referral Prediction in Healthcare Using Graph Neural Networks,” September 2021. <https://europe.naverlabs.com/wp-content/uploads/2021/09/DuarteEtAl2021.pdf>
- Choi et al. “Gram: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.” ACM Conferences, August 1, 2017. <https://dl.acm.org/doi/abs/10.1145/3097983.3098126>.
- Erdem, Kemal. “T-SNE Clearly Explained.” Medium. Towards Data Science, July 21, 2022. <https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a>.
- Fienberg, Stephen E. “A Brief History of Statistical Models for Network Analysis and Open Challenges.” Taylor & Francis, December 12, 2012. <https://www.tandfonline.com/doi/abs/10.1080/10618600.2012.738106>.
- Fisher, Elliott S., Mark B. McClellan, John Bertko, Steven M. Lieberman, Julie J. Lee, Julie L. Lewis, and Jonathan S. Skinner. “Fostering Accountable Health Care: Moving Forward in Medicare.” *Health Affairs* 28, no. Supplement 2 (2009). <https://doi.org/10.1377/hlthaff.28.2.w219>.
- Fuchs, Peter, Fridtjof W. Nussbeck, Nathalie Meuwly, and Guy Bodenmann. “Analyzing Dyadic Sequence Data-Research Questions and Implied Statistical Models.” *Frontiers*. Frontiers, March 7, 2017. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00429/full>.
- Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S. Skinner, and Jason M. Sutherland. “Prices Don’t Drive Regional Medicare Spending Variations.” *Health Affairs* 29, no. 3 (2010): 537–43. <https://doi.org/10.1377/hlthaff.2009.0609>.
- Hamilton, William, Rex Ying, and Jure Leskovec. “Inductive Representation Learning on Large Graphs.” arXiv.org, September 10, 2018. <https://arxiv.org/abs/1706.02216>.
- Han, Qiwei, Mengxin Ji, Inigo Martinez de Rituerto de Troya, Manas Gaur, and Leid Zejnilovic. “A Hybrid Recommender System for Patient-Doctor Matchmaking in Primary care.” *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018. <https://doi.org/10.1109/dsaa.2018.00062>.
- Herrin, Jeph, Pamela R. Soulos, Xiao Xu, Cary P. Gross, and Craig Evan Pollack. “An Empiric Approach to Identifying Physician Peer Groups from Claims Data: An Example from Breast Cancer Care.” *Health Services Research* 54, no. 1 (2018): 44–51. <https://doi.org/10.1111/1475-6773.13095>.
- Hoare, Jake. “How t-SNE Works and Dimensionality Reduction.” Displayr, June 9, 2021. <https://www.displayr.com/using-t-sne-to-visualize-data-before-prediction/>.
- Illinois, Aravind Sankar University of, Aravind Sankar, University of Illinois, Yanhong Wu Visa Research, Yanhong Wu, Visa Research, Liang Gou Visa Research, et al. “DySAT: Proceedings of the 13th International Conference on Web Search and Data Mining.” ACM Conferences, January 1, 2020. https://dl.acm.org/doi/abs/10.1145/3336191.3371845?casa_token=eMKa1nKJxFQAAA

[AA%3AdBdvMTwNl2Wz7OKnypM2NhFx5UcFhRieFYj1tqJyUxd6PBwb48ND8uGxtlQwxVDiIKp7G7VEqYbu.](https://doi.org/10.1007/s13042-020-01155-x)

- James D. Reschovsky, PhD. “Evolving Delivery System and Market Factors and Their Influence on Physician Networks and Patient Care.” *JAMA Internal Medicine*. JAMA Network, January 1, 2018. <https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2664064>.
- Jauhar, Sandeep. “Referral System Turns Patients into Commodities.” *The New York Times*. The New York Time, May 25, 2009. <https://www.nytimes.com/2009/05/26/health/26essa.html>.
- Jia, Peng, Fahui Wang, and Imam M Xierali. “Evaluating the Effectiveness of the Hospital Referral Region (HRR) Boundaries: A Pilot Study in Florida.” *Annals of GIS* 26, no. 3 (2020): 251–60. <https://doi.org/10.1080/19475683.2020.1798509>.
- Kaufman, Brystana G., David Klemish, Andrew Olson, Cordt T. Kassner, Jerome P. Reiter, Matthew Harker, Laura Sheble, Benjamin A. Goldstein, Donald H. Taylor, and Nrupen A. Bhavsar. “Use of Hospital Referral Regions in Evaluating End-of-Life Care.” *Journal of Palliative Medicine* 23, no. 1 (2020): 90–96. <https://doi.org/10.1089/jpm.2019.0056>.
- Kaufman, Leonard, and Peter Rousseeuw. “Partitioning around Medoids (Program Pam).” *Finding Groups in Data*, Wiley Series in Probability and Statistics, March 8, 1990, 68–125. <https://doi.org/10.1002/9780470316801.ch2>.
- Landon, Bruce E., Jukka-Pekka Onnela, Nancy L. Keating, Michael L. Barnett, Sudeshna Paul, Alistair J. O’Malley, Thomas Keegan, and Nicholas A. Christakis. “Using Administrative Data to Identify Naturally Occurring Networks of Physicians.” *Medical Care* 51, no. 8 (2013): 715–21. <https://doi.org/10.1097/mlr.0b013e3182977991>.
- Landon, Bruce E., Nancy L. Keating, Jukka-Pekka Onnela, Alan M. Zaslavsky, Nicholas A. Christakis, and A. James O’Malley. “Patient-Sharing Networks of Physicians and Health Care Utilization and Spending among Medicare Beneficiaries.” *JAMA Internal Medicine* 178, no. 1 (2018): 66. <https://doi.org/10.1001/jamainternmed.2017.5034>.
- Landon, Bruce E., Nancy L. Keating, Michael L. Barnett, Jukka-Pekka Onnela, Sudeshna Paul, A. James O’Malley, Thomas Keegan, and Nicholas A. Christakis. “Variation in Patient-Sharing Networks of Physicians across the United States.” *JAMA* 308, no. 3 (2012). <https://doi.org/10.1001/jama.2012.7615>.
- Li, Yang, Buyue Qian, Xianli Zhang, and Hui Liu. “Knowledge Guided Diagnosis Prediction via Graph Spatial-Temporal Network.” *Proceedings of the 2020 SIAM International Conference on Data Mining*, 2020, 19–27. <https://doi.org/10.1137/1.9781611976236.3>.
- Liu, Sicen, Tao Li, Haoyang Ding, Buzhou Tang, Xiaolong Wang, Qingcai Chen, Jun Yan, and Yi Zhou. “A Hybrid Method of Recurrent Neural Network and Graph Neural Network for next-Period Prescription Prediction.” *International Journal of Machine Learning and Cybernetics* 11, no. 12 (2020): 2849–56. <https://doi.org/10.1007/s13042-020-01155-x>.

- Liu, Zheng, Xiaohan Li, Hao Peng, Lifang He, and Philip S. Yu. "Heterogeneous Similarity Graph Neural Network on Electronic Health Records." *2020 IEEE International Conference on Big Data (Big Data)*, 2020.
<https://doi.org/10.1109/bigdata50022.2020.9377795>.
- Marshall, Kayley. "How Does Epoch Affect Accuracy?" Deepchecks, November 14, 2022.
<https://deepchecks.com/question/how-does-epoch-affect-accuracy/>.
- "Medicare Physician & Other Practitioners - by Provider Data Dictionary." Centers for Medicare & Medicaid Services Data. Accessed October 1, 2022.
<https://data.cms.gov/resources/medicare-physician-other-practitioners-by-provider-data-dictionary>.
- "Medicare Reimbursements." Dartmouth Atlas of Health Care, August 17, 2021.
<https://www.dartmouthatlas.org/interactive-apps/medicare-reimbursements/>.
- Meghanathan, N. "Small World Networks." Lecture, Jackson State University, Jackson, MS.
<https://www.jsu.edu/nmeghanathan/files/2015/08/CSC641-Fall2015-Module-6-Small-World-Networks-reduced.pdf>.
- "National Provider Identifier Standard (NPI)." CMS. Accessed October 8, 2022.
<https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand>.
- "Node Representation Learning with GraphSAGE and UnsupervisedSampler¶." Node representation learning with GraphSAGE and UnsupervisedSampler - StellarGraph 1.2.1 documentation. Accessed November 26, 2022.
<https://stellargraph.readthedocs.io/en/stable/demos/embeddings/graphsage-unsupervised-sampler-embeddings.html>.
- NPI Files. NPI files. (n.d.). Retrieved December 4, 2022, from
https://download.cms.gov/nppes/NPI_Files.html
- NPES NPI Registry Help. NPES NPI Registry. (n.d.). Retrieved December 4, 2022, from
<https://npiregistry.cms.hhs.gov/help/help-details>
- "Physician shared patient patterns methodology - centers for Medicare" (n.d.). Retrieved December 10, 2022, from
https://downloads.cms.gov/foia/physician_shared_patient_patterns_technical_requirements.pdf
- Perez, Charles, and Rony Germon. "Graph Creation and Analysis for Linking Actors: Application to Social Data." *Automating Open Source Intelligence*, 2016, 103–29.
<https://doi.org/10.1016/b978-0-12-802916-9.00007-5>.
- Peter G. Peterson Foundation. "Budget Basics: Medicare," July 5, 2022.
<https://www.pgpf.org/budget-basics/medicare>.
- "Quality of Care." World Health Organization. World Health Organization. Accessed October 16, 2022. <https://www.who.int/westernpacific/health-topics/quality-of-care>.

- Rijal, Avishek. "Clustering: Part 2, Putting the 'k' in K-Means." Xabit, November 22, 2022. <https://xabitanalytics.com/2022/11/22/clustering-part-2/>.
- Sharma, N. (2022, November 14). *K-means clustering explained*. neptune.ai. Retrieved December 11, 2022, from <https://neptune.ai/blog/k-means-clustering#:~:text=Elbow%20point%20is%20used%20as,longer%20worth%20the%20additional%20cost.>
- Skinner, Jonathan S., Elliott S. Fisher, and John Wennberg. "The Efficiency of Medicare." NBER, August 8, 2005. <http://www.nber.org/chapters/c10359>.
- Skinner, Lucy, Sandra Wong, and Carrie Colla. "Rethinking Rurality: Using Hospital Referral Regions to Investigate Rural-Urban Health Outcomes - BMC Health Services Research." SpringerLink. BioMed Central, November 3, 2022. <https://link.springer.com/article/10.1186/s12913-022-08649-0>.
- Traag, V. A., L. Waltman, and N. J. van Eck. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." Nature News. Nature Publishing Group, March 26, 2019. <https://www.nature.com/articles/s41598-019-41695-z?ref=https%3A%2F%2Fgithubhelp.com>.
- Tran, Minh-Hien. "Comparing Umap vs T-Sne in Single-Cell RNA-Seq Data Visualization, Simply Explained." BioTuring's Blog, January 31, 2022. <https://blog.bioturing.com/2022/01/14/umap-vs-t-sne-single-cell-rna-seq-data-visualization/>.
- "Understanding Physician Network Dynamics Could Help Providers Bring down Health Care Costs." Deloitte Insights. Accessed December 4, 2022. <https://www2.deloitte.com/us/en/insights/industry/health-care/physician-networking-for-performance-improvement-in-health-care.html>.
- USDA ERS. "Rural-Urban Commuting Area Codes." USDA ERS - Rural-Urban Commuting Area Codes. Accessed December 15, 2022. <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx#:~:text=The%20rural%20urban%20commuting%20area,census%20tracts%20that%20comprise%20them>.
- Wang, Changzhen, Fahui Wang, and Tracy Onega. "Network Optimization Approach to Delineating Health Care Service Areas: Spatially Constrained Louvain and Leiden Algorithms." *Transactions in GIS* 25, no. 2 (2020): 1065–81. <https://doi.org/10.1111/tgis.12722>.
- Wang, Changzhen, and Fahui Wang. "GIS-Automated Delineation of Hospital Service Areas in Florida: From Dartmouth Method to Network Community Detection Methods." *Annals of GIS* 28, no. 2 (2022): 93–109. <https://doi.org/10.1080/19475683.2022.2026470>.
- Wang, Yi-Tang. "What Are Small-World Network Models?" Medium. Towards Data Science, April 16, 2020. <https://towardsdatascience.com/what-are-small-world-network-models-87bbcfe0e038>.

- Whaley, Christopher M., Xiaoxi Zhao, Michael Richards, and Cheryl L. Damberg. “Higher Medicare Spending on Imaging and Lab Services after Primary Care Physician Group Vertical Integration.” *Health Affairs* 40, no. 5 (2021): 702–9. <https://doi.org/10.1377/hlthaff.2020.01006>.
- Yue, Xiang, Zhen Wang, Jingong Huang, Srinivasan Parthasarathy, Soheil Moosavinasab, Yungui Huang, Simon M Lin, Wen Zhang, Ping Zhang, and Huan Sun. “Graph Embedding on Biomedical Networks: Methods, Applications and Evaluations.” *Bioinformatics*, 2019. <https://doi.org/10.1093/bioinformatics/btz718>.
- Zeltzer, Dan. “Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap.” *American Economic Journal: Applied Economics* 12, no. 2 (2020): 169–97. <https://doi.org/10.1257/app.20180201>.

Appendix

Table 1. Selected Medicare Reimbursement Rates by Florida HRR. Data from Dartmouth Atlas of Healthcare, 2015, Price, age, sex and race adjusted (“Dartmouth Atlas of Healthcare Data 2022”)

HRR #	HRR name	Medicare enrollees (2015)	Total Medicare reimbursements per enrollee (Parts A and B)	Hospital & skilled nursing facility reimbursements per enrollee	Physician reimbursements per enrollee	Outpatient facility reimbursements per enrollee	Home health agency reimbursements per enrollee	Hospice reimbursements per enrollee	Durable medical equipment reimbursements per enrollee
115	Bradenton	39,678.00	10,544.49	4,212.99	3,902.24	1,172.71	805.97	271.84	174.85
116	Clearwater	52,103.00	11,331.01	4,575.79	3,868.90	1,337.17	976.96	342.69	217.50
118	Fort Lauderdale	270,366.00	11,515.69	4,276.00	4,269.25	1,486.57	912.36	385.76	191.12
119	Fort Myers	192,018.00	9,980.16	3,672.06	4,002.06	1,098.89	625.67	345.11	216.44
120	Gainesville	63,302.00	10,328.01	4,553.13	3,177.88	1,196.98	748.38	452.12	212.59
122	Hudson	45,070.00	11,494.79	5,079.43	3,999.53	999.25	932.71	263.64	209.79
123	Jacksonville	159,629.00	11,143.78	4,599.15	3,568.60	1,554.04	709.42	495.72	225.16
124	Lakeland	29,141.00	11,032.98	4,831.43	3,468.60	1,193.59	889.65	425.65	211.94
127	Miami	145,033.00	13,109.15	5,191.65	3,701.58	1,556.16	1,593.65	568.85	211.54
129	Ocala	104,695.00	10,236.63	3,993.02	3,915.66	1,118.49	615.65	366.82	208.39
130	Orlando	354,310.00	11,017.52	4,600.60	3,590.60	1,372.76	781.06	441.48	226.26
131	Ormond Beach	52,061.00	10,043.62	3,773.12	3,210.42	1,533.12	773.86	588.84	171.31
133	Panama City	27,363.00	11,393.02	5,122.49	3,539.09	1,343.43	612.58	497.88	287.24
134	Pensacola	93,551.00	10,448.59	4,313.57	2,902.79	1,809.99	716.57	485.91	231.76
137	Sarasota	90,707.00	9,645.26	3,489.95	3,847.34	1,217.29	685.47	234.91	172.09
139	St. Petersburg	37,395.00	11,662.96	5,171.11	3,797.55	1,278.71	934.37	275.82	199.04
140	Tallahassee	62,676.00	9,944.33	4,416.51	2,690.74	1,630.85	493.29	493.38	223.01
141	Tampa	90,046.00	11,766.31	5,037.77	3,691.05	1,289.53	1,107.03	404.33	218.04

Source: Dartmouth Atlas Data. “Medicare Reimbursements” By 2015, 100% Samples.
<https://data.dartmouthatlas.org/medicare-reimbursements/>

Table 2. Selected Primary Care Access and Quality Measures by Florida HRR. Data from Dartmouth Atlas of Healthcare, 2015, Price, age, sex and race adjusted (“Dartmouth Atlas of Healthcare Data 2022”)

HRR #	HRR Name	Number of diabetic Medicare enrollees age 65-75	Average annual % of diabetic Medicare enrollees age 65-75 having hemoglobin A1c test	Average annual % of diabetic Medicare enrollees age 65-75 having eye examination	Average annual % of diabetic Medicare enrollees age 65-75 having blood lipids (LDL-C) test	Number of female Medicare enrollees age 67-69	Average % of female Medicare enrollees age 67-69 having at least one mammogram over a two-year period	Number of Medicare beneficiaries (Part A eligible)	Leg amputations per 1,000 Medicare enrollees	Discharges for ambulatory care sensitive conditions per 1,000 Medicare enrollees
115	FL-Bradenton	3641	86.18511	71.51881	81.79072	3770	69.62865	32010.8	-0.624	46.778
116	FL-Clearwater	4335	87.797	69.68858	87.56632	4229	68.40861	42957.8	-0.583	52.688
118	FL-Fort Lauderdale	21653	87.11033	71.14026	86.09431	21357	69.37772	223113.8	0.294	49.475
119	FL-Fort Myers	17568	87.67646	71.67008	87.10155	17416	75.06316	153701.3	0.492	42.758
120	FL-Gainesville	7573	84.68242	65.53545	78.87231	5840	62.44863	51914.3	0.685	70.686
122	FL-Hudson	5051	87.13126	68.3429	87.48763	3764	69.20829	35858.3	-0.602	64.64
123	FL-Jacksonville	20066	84.55597	65.20482	81.40138	16386	65.50104	130293	0.574	57.867
124	FL-Lakeland	3517	87.26187	68.89394	86.80694	2468	66.97731	23868	-0.723	70.332
127	FL-Miami	15311	85.91209	64.26752	86.04271	11166	56.75264	125793	0.687	64.498
129	FL-Ocala	11713	88.90122	72.55187	89.37078	9315	75.22276	81975.8	0.464	45.505
130	FL-Orlando	40143	86.46588	67.17734	86.21927	31326	67.15508	288960.8	0.579	56.699
131	FL-Ormond Beach	5311	87.83657	70.68349	86.66918	4756	70.94197	41910.8	-0.35	43.056
133	FL-Panama City	3471	79.48718	66.69548	76.37568	2595	58.84393	21699	-0.726	62.098
134	FL-Pensacola	11417	81.06333	65.73531	77.60357	9207	62.32215	74656.5	0.608	59.604
137	FL-Sarasota	6567	86.91945	74.43277	84.11756	7664	74.43894	71592	0.337	32.573
139	FL-St. Petersburg	3258	86.00368	67.24985	83.45611	3156	63.75158	31374	-0.604	57.236
140	FL-Tallahassee	7972	84.20723	62.20522	80.55695	5919	62.83156	51876.8	0.633	58.739
141	FL-Tampa	10295	84.56532	64.5459	83.38028	8313	63.27439	75918.8	0.534	60.444

Source: Dartmouth Atlas Data. “Selected Primary Care Access and Quality Measures” for 2015, 100% Samples. <https://data.dartmouthatlas.org/primary-care/>

Figure 1. Rural Zip Codes % per HRR

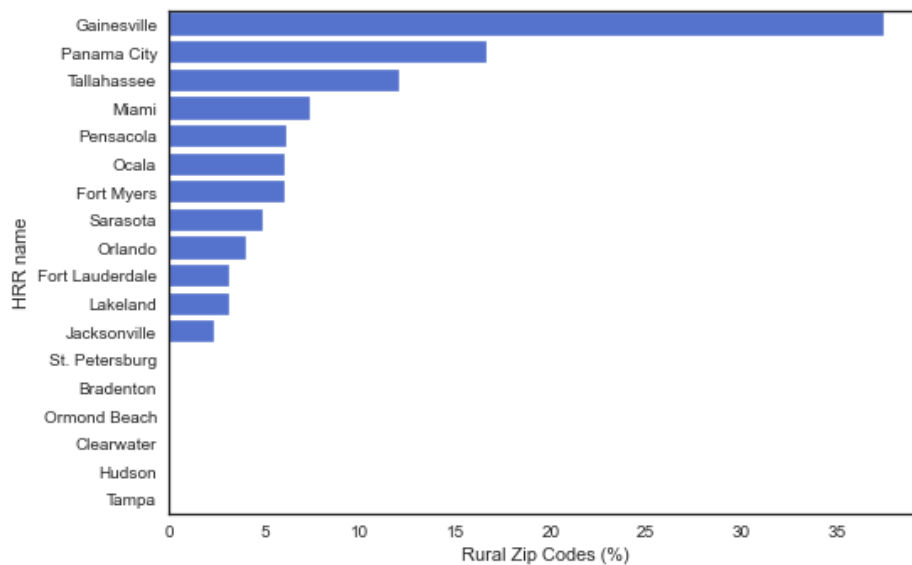
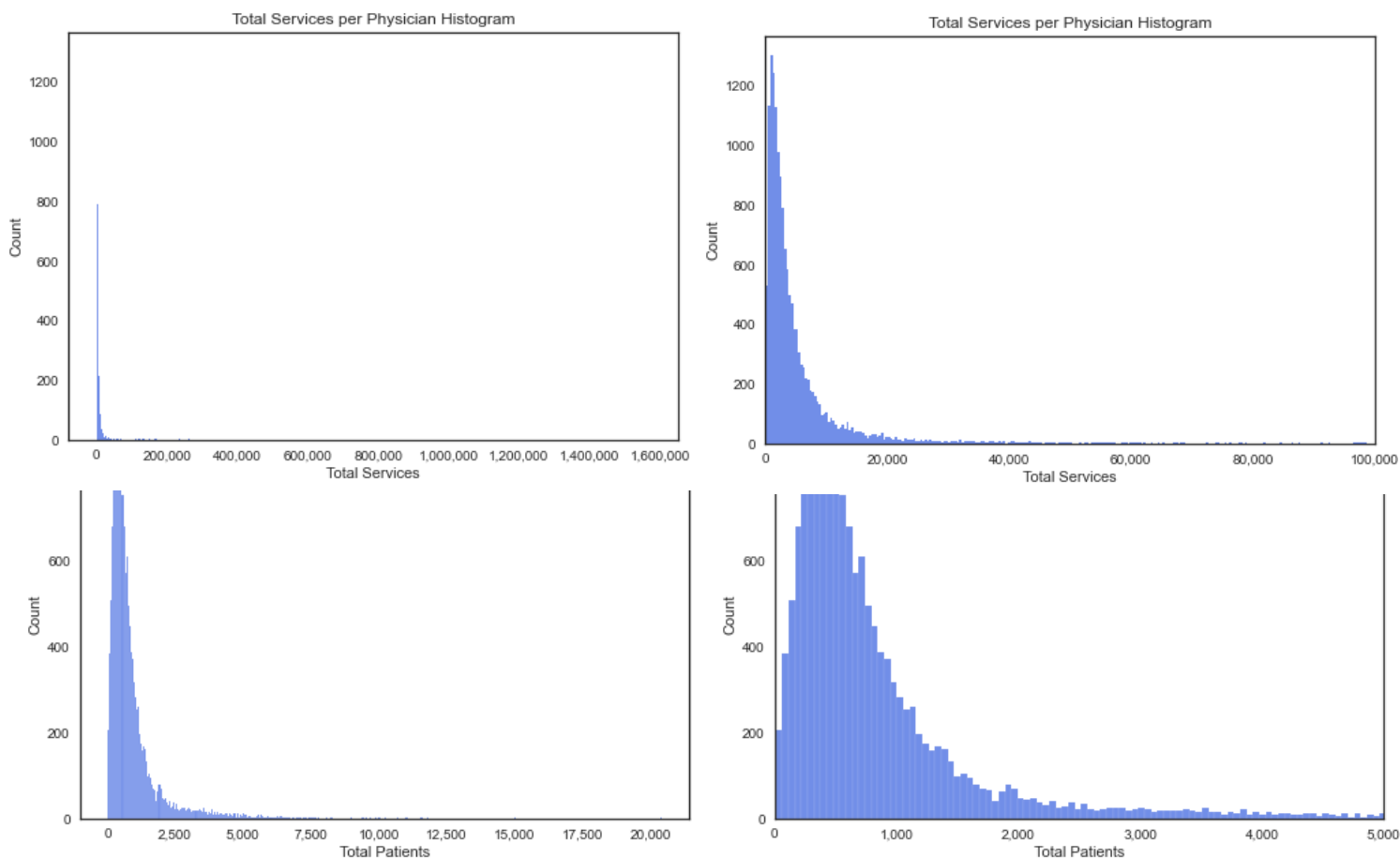
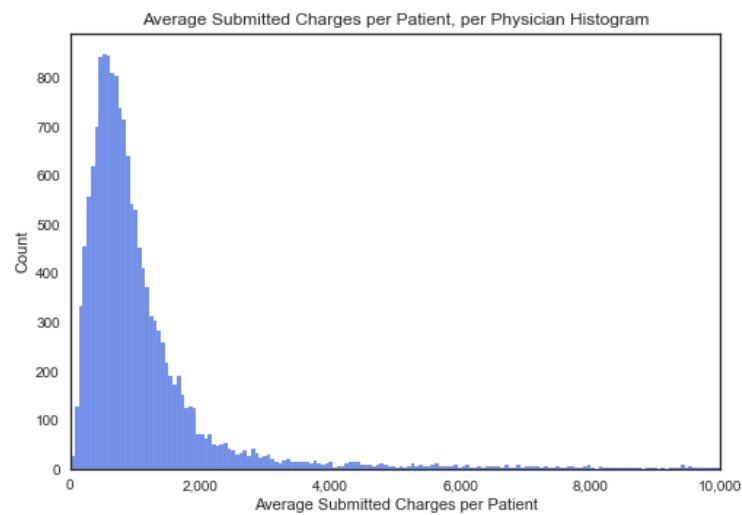
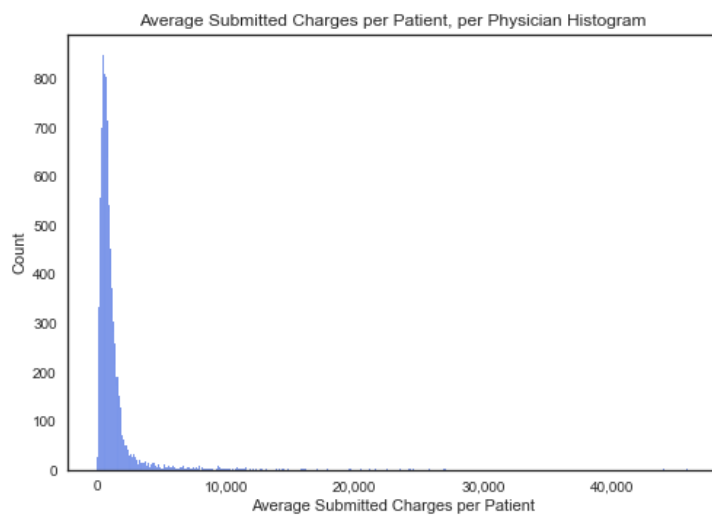
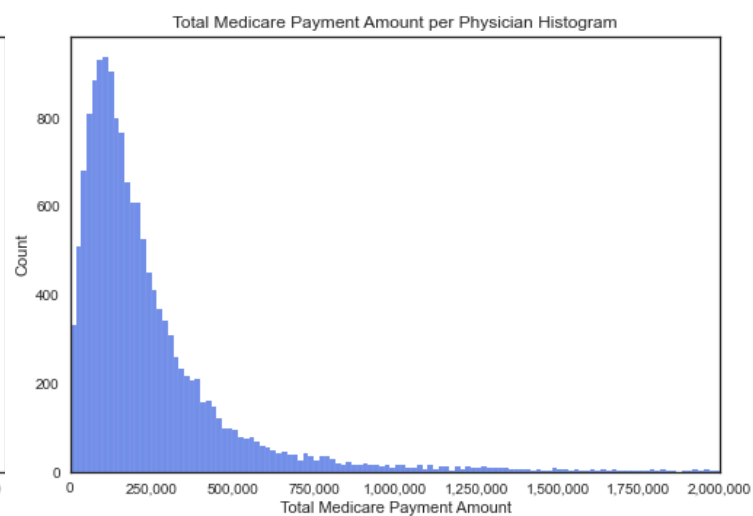
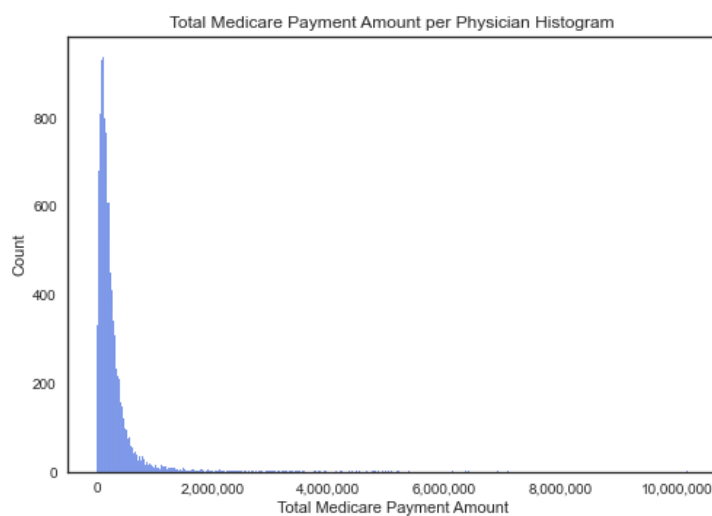
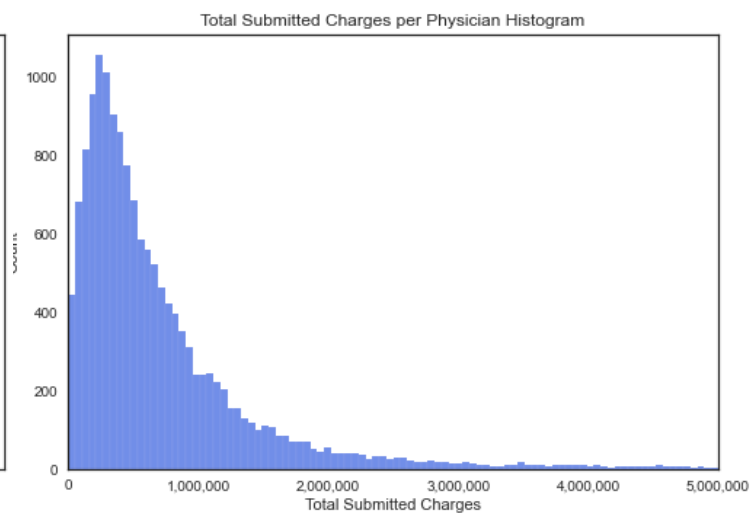
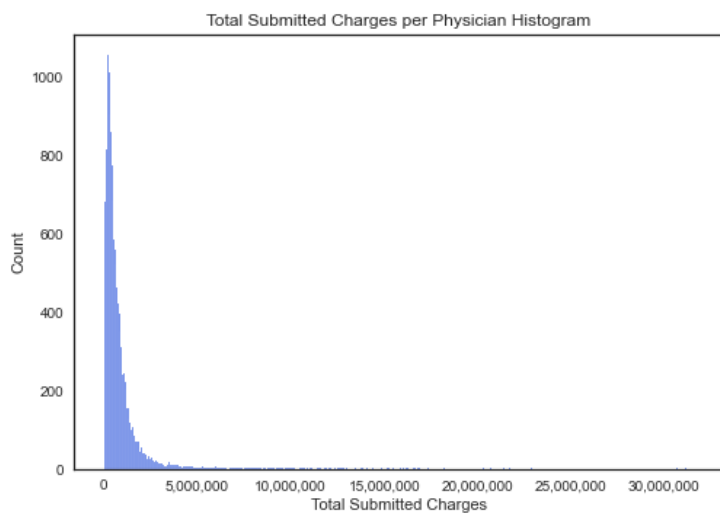


Figure 2. Exploratory Data Analysis for Section 6: Variables Distributions. All data (left), limited x-axis (right)





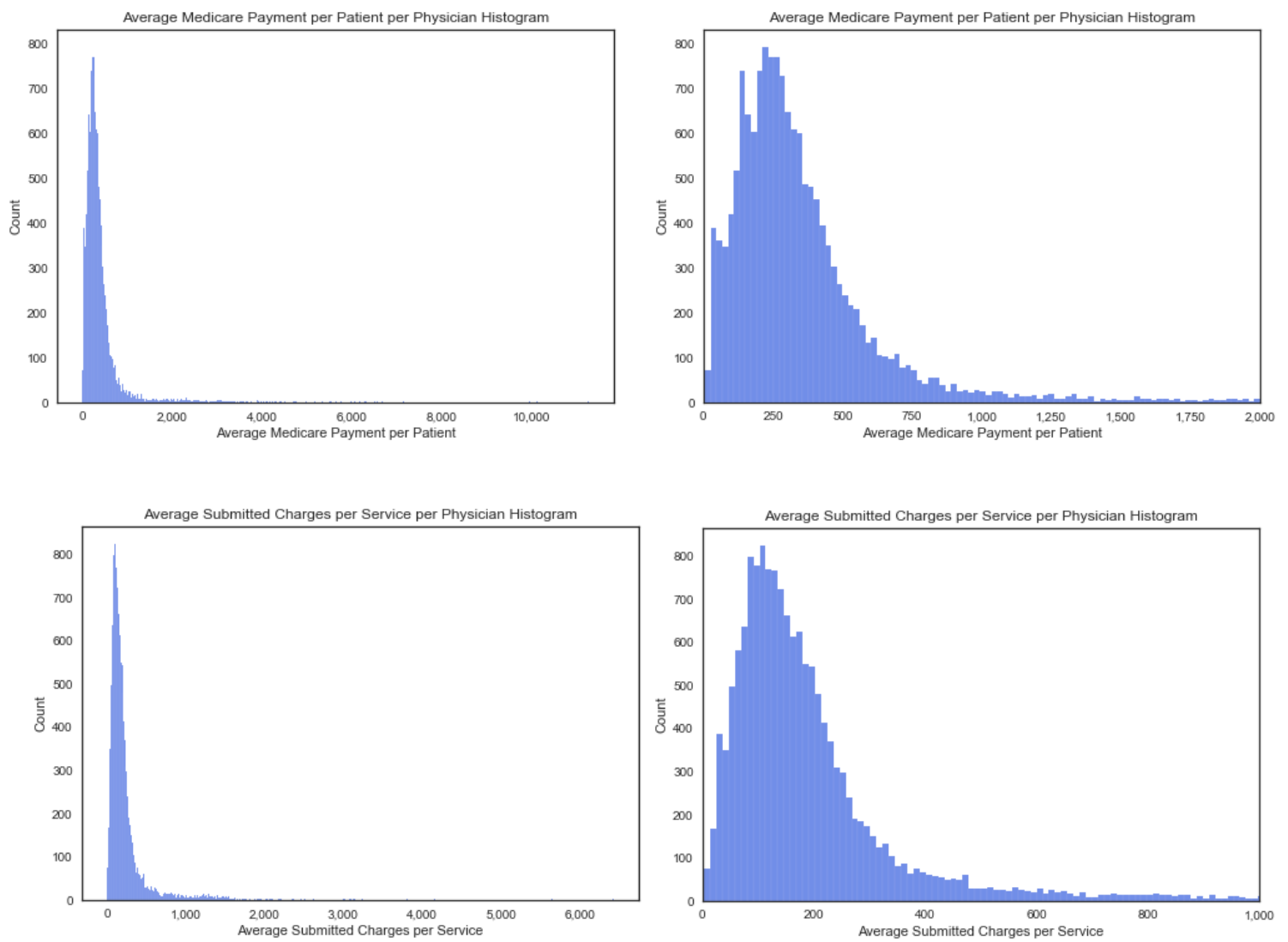


Figure 3. Average Silhouette Scores (left), Silhouette Plot for K-Medoids with 2 centers (right)

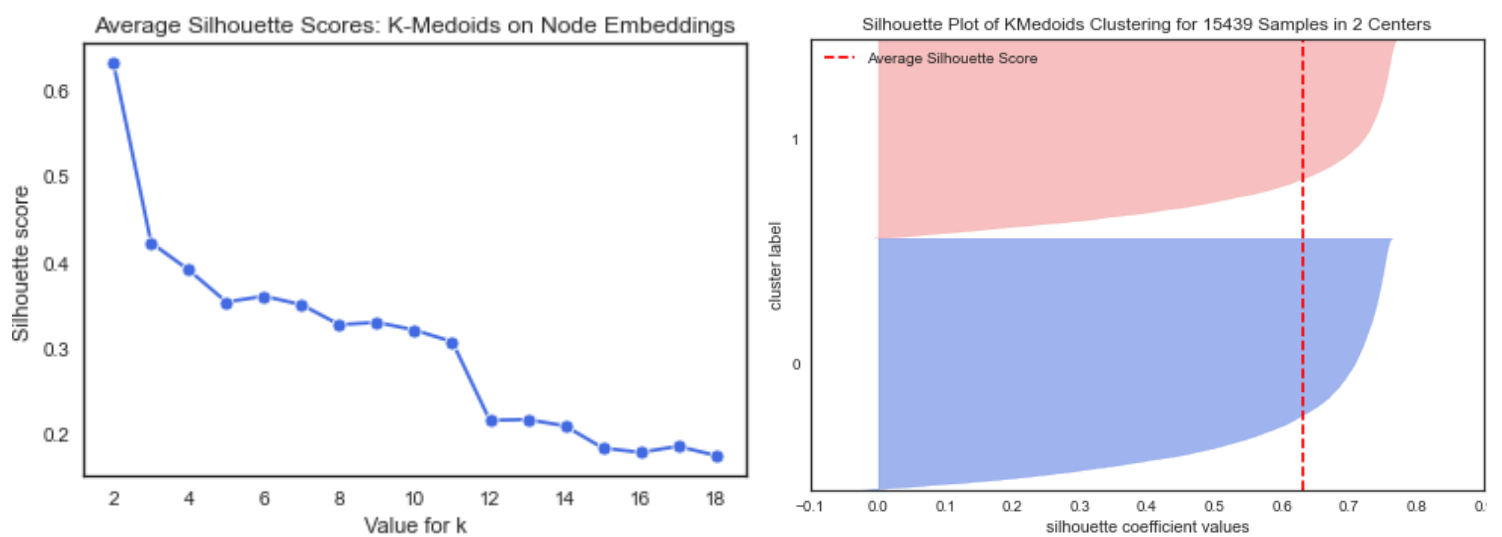


Table 3. Gender and Top 5 Specialties Distribution, by Cluster

Variable	Cluster Label	
	Cluster 0	Cluster 1
Gender		
<i>Male (%)</i>	76.64	72.95
<i>Female (%)</i>	23.36	27.05
Top 5 Specialties		
<i>Internal Medicine (%)</i>	19.95	25.53
<i>Family Practice (%)</i>	9.97	10.44
<i>Cardiology (%)</i>	7.87	6.55
<i>Diagnostic Radiology (%)</i>	6.93	5.44
<i>Nurse Practitioner (%)</i>	6.01	7.3

Figure 4. Log-scale histogram of Average Submitted Charges per Patient for each cluster

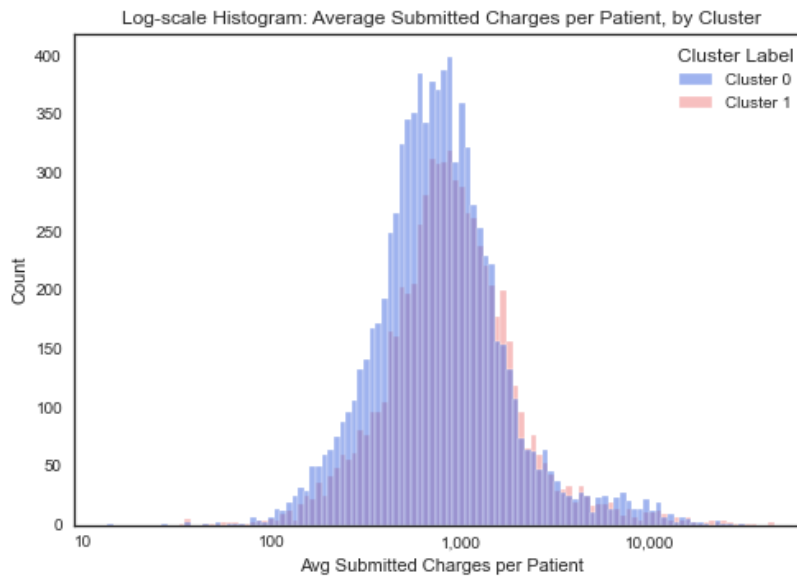
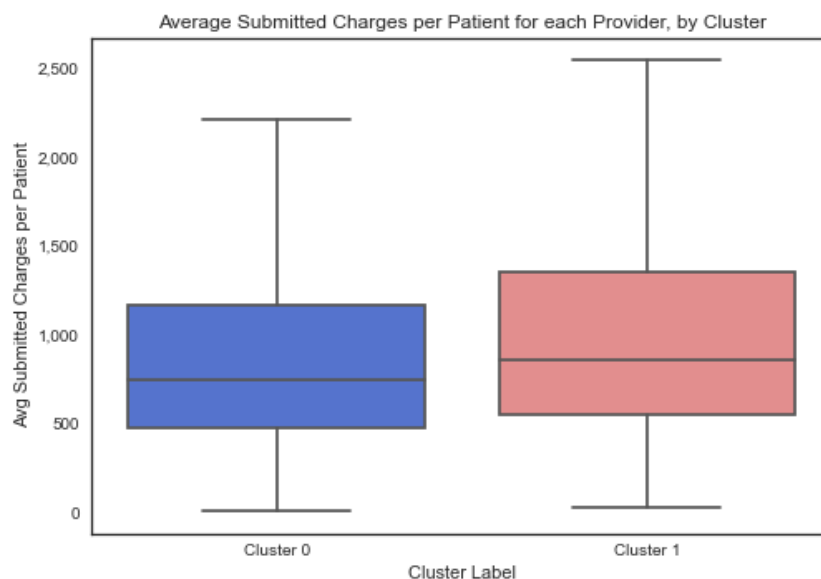


Figure 5. Average Submitted Charges per patient, Boxplot distribution for each cluster without outliers



Figures 6 and 7. Providers' Count for each RUCA code, by cluster (left), Providers' Count for each Rural RUCA code, by cluster (right)

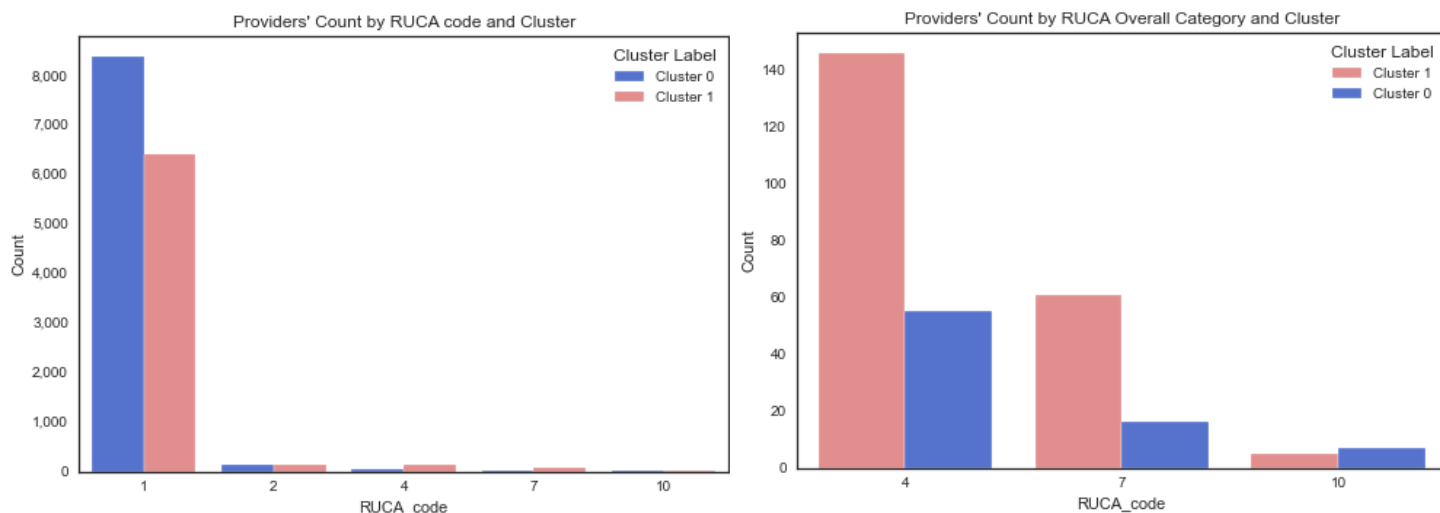


Figure 8. Providers' Count by RUCA rurality category

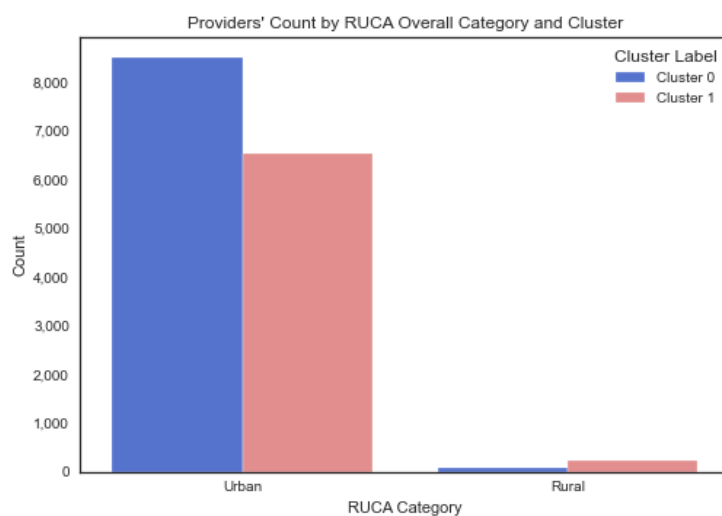


Figure 9. Log-scale histogram for Total Services per Physician by Cluster (left) and Total Patients per Physician by Cluster (right)

