NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Understanding the Referral Mechanism between General Practitioners and Specialists in Private Healthcare using Network Science

João Manuel Caixinha Casal da Veiga

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

UNDERSTAING THE REFERRAL MECHANISM BETWEEN GENERAL PRACTITIONERS AND SPECIALISTS USING NETWORK SCIENCE

by

João Manuel Caixinha Casal da Veiga

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management, with specialization in Knowledge Management and Business Intelligence

Co Advisor: Qiwei Han

Co Advisor*:* Miguel Castro Neto

11/2019

ACKNOWLEDGEMENTS

I would like to start by thanking Cláudia Soares (Instituto Superior Técnico) for her involvement in this research project.

Thank you to the partner who provided the data for this dissertation and made it possible.

My advisors Qiwei Han (Nova School of Business & Economics) and Miguel Castro Neto (Nova Infomartion Management School) for all the support and guidance that they gave during the development of this dissertation.

This dissertation was accepted in a prestigious international conference known as ComplexNetworks.

A special thank you to my grandmothers who are two of the most amazing women I ever met, Lucinda Rosa Gallego and Maria Manuela Marques de Amorim Casal da Veiga.

My parents and family who supported, motivated and encouraged me to finish this incredible challenge.

To my friends, my deepest thanks: Susana Maruta, Miguel Preto, Joana Frade, Inês Pereira Leonor Braz Teixeira, Carina Figueiredo, Margarida Teixeira Maia, Tomé Santos, Beatriz Passos, Joana Gomes, João Francisco Pereira, Miguel Garção, Pedro Barroso, Bernardo Sequeira, João Caetano, Joana Serralheiro, Inês Narciso, Eduardo Santana, Manuel Pereira, Maria Clara Costa, Catarina Mendes, Pedro Sousa, José David, Sérgio Silva, André Mestre, Carolina Amorim, Filipe Rosado, Catarina Sousa, Wilson Ramos, Luís Rodrigues, Bianca Garcia, Miguel Ferreira, Pedro Santos, Marc Brander, Miguel Saraiva, Inês Filipa Prazeres Rosário, Inês Rico, Maria Alves, Raquel Santos, Inês Alexandre, Patricia Tomás, Marta Gomes, Gabriela Madureira, Manuel Pereira, Joana Amaral, Bruna Vieira, Frederico Matos, Catarina Fonseca, Cristina Snimshchikova, Ana Isabel, Catarina Nunes, Maria Prata, António Barroso, Luís Belchior, Maria do Mar Belchior, Paulo Belchior, Luis Martinez, Marcel Reinhardt, Fernanda Santos, Manuela Serralheiro, Rodrigo Augusto, Diego Durâes, Ram Bahadur Gurung, Bal Kumar Chhantyal, Ram Bahadur Budhathoki, Nand Kishore, Sabrina Fucuda, Bir Bahadur Magar, Tika Ram Thapa, Gabriel Borges, Emanuelle Raysa, Sissy Antoniolli, Germany Belo, Dannilson Félix, Daniell Félix, Virgínia Ficagna, Gustavo Arrigoni, Christyan Ventura, Pedro Navarro, Ricardo Pereira, Hyago Moraes, Jéssica Sousa, Camila kristi, Tudor dani, Deepa Shrestha, Vishal Gurung, Lucas Lisboa, Luan Durães, Mansingh, Dhan Prasad Gurung, Tej Narayan Gurung, Santiago Punk, Gislaine Silva. Any of these persons had an impact in my life, and if it were not for them I would not be able to complete what I can say to have been one of the hardest challenges in my life so far.

To the non-lucrative organization SABI. In particular a especial thank you, to Diana Azevedo Jesus, Joana Vaz, Maria Inês Vasconcelos, Cátia Martinho, Francisco Salva.

To all other parties that might have been directly or indirectly involve in this dissertation and my double degree thank you for everything.

ABSTRACT

The following document represents a master thesis dissertation, that consists in the development of a research project in which its main objective is to understand the referral mechanism between primary care physicians and specialists using network science and other tools. The referral network of a healthcare provider denotes an important source of costs and revenue, as it can affect directly the management of its clients and employees, namely through the quality of the care being provided and the level of satisfaction being achieved. The data provided for the development of the study was given by a European industry leader in private healthcare. It is important to highlight that this research study attempts maps the relationships between general practitioners and specialists using a large dataset of doctor's appointments during 2012 and 2017. These relationships were mapped under the assumptions that two doctors had to share at least one patient, and the period between the two appointments could not be longer than 30 days. Moreover, the impact of the dynamics of the relationships between the two types of doctors in the primary-specialty referral mechanism is done by analyzing the referral patterns exhibited in network, and the performance of the physicians in terms of the centrality measures degree, betweenness and closeness. Additionally, two regression analysis are performed with the objective of identifying potential characteristics that might be affecting the referral rates of doctors. These characteristics include the social network metrics and the physician's backgrounds.

KEYWORDS

Social Network Analysis; Referral Mechanism; Informal Networks; Private Healthcare, Primary Care; Differentiated Care

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

WHO	World Healthcare Organization
OCDE	Organization for Economic Cooperation and Development
CAGR	Compounded Annual Growth Rate
NHS	National Healthcare Service
PCP	Primary Care Physician
SP	Specialist
GP	General Practitioner
HIT	Healthcare Information Technology
PCMH	Patient Centered Medical Homes
ACO	Accountable Care Organizations
ACSS	Administração Central do Sistema de Saúde
SNA	Social Network Analysis

1. INTRODUCTION

The relationship between a primary care physician[1] (pcp) and a specialist (sp) represents a relevant communication channel for the healthcare system, as it can deeply influence the outcome of a patient's treatment quality and the level of satisfaction regarding the care provided by a healthcare institution (Kinchen, Cooper, Levine, Wang, & Powe, 2004). Additionally, it can act as a significant driver of costs (Barnett, Landon, O'Malley, Keating, & Christakis, 2011).

Having knowledge of the relationships established in a referral mechanism, between pcp and sp, can represent an important source of information to manipulate the overall network for the benefit of healthcare providers, as doctors are influenced by their respective networks (B.E. et al., 2012). An efficient healthcare referral system is proven to improve the ability of both types of physicians to provide better quality treatment to their patients (Care, Referral, Means, & Problems, 2019).

This dissertation presents an approach in which the main objective is to better understand the relationship between doctors in primary care and specialists. Specifically, what might be the potential impact of its dynamics in the informal structure, that is the primary-specialty referral network mechanism, using network science and other tools. From an extensive dataset with 9,173,891 patient consultations records, more than 1.3 million unique patients and 2171 unique doctors it was generated a bipartite social network with 459 connected general practitioners, 14 isolated family doctors, 1487 connected specialist and 211 isolated doctors certified in differentiated care. Additionally, the 1946 connected doctors are linked between each other by no more than 59,065 edges. This network was further augmented with the information regarding doctors' backgrounds, their age, the number of years of medical practice, gender and others.

A large percentage of the existent literature focus on the use of descriptive methods to identify and describe patterns based on the informal relationships between primary care doctors and specialists, such as doctor specialties, the patient diagnosis, and disease severity (An, O'Malley, Rockmore, & Stock, 2018). In that sense, this dissertation act as a complement to other research studies, as it allows to validate, or not, patterns that have been recognized to affect the referral decision process between the two types of doctors. Additionally, it is investigated if certain network science metrics have an impact or not in the primary-specialty referral mechanism.

1.1. RESEARCH OBJECTIVES

The research purposes of this dissertation were to add to the current literature more information about the referral patterns of a primary-specialty referral system modelled as a bipartite network and explore its potential inefficiencies. Moreover, it was presented the opportunity to test the hypothesis that the different characteristics of a doctor's background and social network centrality measures have an impact or not in the referral rate of physician. From the perspective of the one making the referrals (pcp) and from the perspective of the one receiving them (sp).

[1] Primary care physician, primary care doctors, family doctors, pcp, and general practitioners are used interchangeably throughout this research project. In addition, the same succeeds with specialist, sp and differentiated care doctors.

2. LITERATURE REVIEW

In this section it is provided a contextualization of the healthcare industry, by making a helicopter approach at the global landscape (zoom out) and by focusing on the Portuguese healthcare system (zoom in). Furthermore, an overview of the referral mechanism between primary care doctors and specialists is made, along with its impact on the different type of individuals involved in the process: patients, doctors and respective healthcare providers (the institutions). Additionally, a synopsis of the patterns so far studied is provided. To conclude, an overview on the methodology of social network analysis is written for further understanding on how it can be applied to study the referral mechanism between general practitioners and specialist doctors, along with the description of some of its characteristics.

2.1. HEALTHCARE INDUSTRY

According to the World Health Care Organization (WHO), health can be seen as a state physical, mental and social well-being and not only the absence of sickness or a disease (Larson, 1996). In 2017 it was estimated that In terms of access to essential healthcare half of the worldwide population did not have access (WHO, 2017).

The amount spent in healthcare by households varies according to the country. In 2017 was estimated that the healthcare spending per capita in India to be 238 dollars while in the United States of America was approximately 34 times higher (8047 dollars). In the particular case of Portugal, the household spending in health is roughly 2000 dollars, 1925 to be more precisely (OCDE, 2019).

Between the years of 2013 and 2017 it was registered an increase in healthcare spending (2.9%) and is still expected to grow by even larger numbers, as until 2022 is expected for healthcare spending to have an annual growth of 5.4% (CAGR). This growth is associated with the growing needs of healthcare of the elderly population, the increase of the labour costs in the industry, the advances in treatments, the development of new healthcare technologies, and the expansion of healthcare coverage broaden up to new markets (Deloitte, 2019).

The healthcare stakeholders are investing in a new vision for the healthcare industry in which, instead of their goal being to treat their clients when they are sick already, is to create a system that promotes the well-being, prevention and early intervention on minor or serious diseases. The objective of this to develop of the concept of smart health in order to drive innovation, to reduce costs both to the service provider and for the patient while increasing the quality of the treatments (Deloitte, 2019).

Healthcare organizations are experiencing a state of uncertainty in the health economy as profits have been shrinking, new competitors are entering new markets, personalized medicine is becoming a certainty, disruptive technologies[2] and clinical advances are constantly entering in the market.

[2] The healthcare sector is expected to invest 280.25 billion dollars by 2021 in health technologies, related products and services (Deloitte, 2019).

Therefore, this reality is creating a sense of urgency in healthcare providers to decide now whether they want to invest to be a market leader, a follower or a niche player (Deloitte, 2019). The factors that dictate the competition within private operators can be summarized to reputation, clinical excellence, technology, price and client satisfaction (Deloitte, 2019).

2.2. HEALTHCARE INDUSTRY IN PORTUGAL

When Portugal was under the government of a dictatorship the healthcare system in Portugal was fragmented in, private healthcare and state hospitals which were usually located in important urban centers. In 1958, the Portuguese health ministry was created. Until 1970 Portugal was one of the worst countries with access to essential care in the European Union. After 1971, it was created a new system where each districted had two different infrastructures implemented, healthcare centers and hospitals. The current national healthcare (NHS) was then created in 1974 with the democratic purpose of making healthcare available to everyone (Baganha, Ribeiro, & Pires, 2002). The coexistence of public and private healthcare is due to the fact the national healthcare service is not able to cover the entire Portuguese population (Baganha et al., 2002).

Currently the Portuguese NHS administration is divided into five major health regions: Norte, Centro, Lisboa and Vale do Tejo, Alentejo and Algarve. Their headquarters are respectively, Porto, Coimbra, Lisboa, Évora, and Faro. Each region is then divided into sub-regions corresponding to districts. In each district is expected for the healthcare services to be ensure by healthcare centers and hospitals. Nowadays, healthcare centers are in charge of providing the primary healthcare to the Portuguese population, with respect to the national public health sector (Baganha et al., 2002).

When compared with other countries worldwide, the number of doctors per 1000 inhabitants in Portugal is estimated to be above the average by registering a number of 4,26 doctors against the average estimation of 3,2 of the OCDE and its 34 state members. In Portugal, is has been forecasted that approximately 600 000 Portuguese still do not have a family doctor (Notícias, 2019). In addition, the number of patients that that a General Practitioner has to see has increase from 1550 in 2011 to 1900 in 2018 (Notícias, 2019), and there are regions that register extreme numbers like 2500 (Teixeira, 2018).

2.3. PRIMARY CARE DOCTORS AND SPECIALISTS

Primary care denotes a starting point for individuals that requests hospital services in a world where healthcare is more specialized/fragmented than ever. Additionally, they are the ones that are more likely to develop long term relationships, as they regularly follow patients for long periods (Han et al., 2018a).

The training of a pcp is particularly focused on giving practitioners the necessary tools to ensure that they have the best first contact and provide quality continuing care of patients undiagnosed signs. Furthermore, the medical expertise’s that are comprised in this category of doctors include general medicine, internal medicine and pediatrics[3] (American Academy of Family Physicians, 2016).

[3] The term Primary Care does not necessarily characterize exactly the activities or practices a General Practitioner performs or has to perform.

Therefore, the remaining specialties such as cardiovascular surgery, ophthalmology along with more 45 areas of expertise available in Table 1 - Doctor Specialties[4] (D. da República, 2018b), will determine who are the doctors certified in differentiated care. Moreover, a glossary of all specialties is in Appendix 9.1 – Doctor Specialties Glossary.

Doctor Specialties	
Anesthesiology	Physical Medicine and Rehabilitation
Pathologic anatomy	General and Family Medicine
Angiology and Vascular Surgery	Intensive Care Medicine
Cardiology	Internal medicine
Pediatric Cardiology	legal Medicine
Cardiac surgery	Nuclear medicine
Cardiothoracic Surgery	Tropical Medicine
General surgery	Nephrology
Maxillofacial Surgery	Neurosurgery
Pediatric surgery	Neurology
Reconstructive and Aesthetic Plastic Surgery	Neuroradiology
Thoracic surgery	Ophthalmology
Dermato-Venereologia	Medical Oncology
Infectious diseases	Orthopedics
Endocrinology and Nutrition	Otolaryngology
Stomatology	Clinical pathology
Gastroenterology	Pediatrics
Medical Genetics	Pneumology
Gynecology / Obstetrics	Psychiatry
Immunoallergology	Psychiatry of Childhood and Adolescence
Immunohemotherapy	Radiology
Clinical Pharmacology	Radioncology
Clinical Hematology	Rheumatology
Sports Medicine	Public health
Work Medicine	Urology

Table 1- Doctor Specialties

2.4. The referral Mechanism Between Primary Care Doctors and Specialists

In a significant number of countries, it is required for a patient to see a specialist or have access to other medical resources such as laboratory tests, first to have an appointment with a general practitioner (Liddy, Arbab-Tafti, Moroz, & Keely, 2017). However, nowadays is possible to schedule an appointment directly with a specialist the private sector.

[4] The specialties Tropical Medicine and Cardiothoracic surgery after 2018 were no longer officially recognized as specialties. However, given that the Portuguese government issued an article recognizing those two as a specialty in 2015 the same will happen throughout the research project (D. A. República, 2015).

The practice of referring patients to specialists is considered to be a common event in the United States of America (Chen & Glover, 2016). Interestingly, two different studies found similar results regarding the percentage of visits to family doctors appointments that include a referral to a specialist in the United States of America, 5% (Forrest, Nutting, Von Schrader, Rohde, & Starfield, 2006) and 4.5% (Kinchen et al., 2004). Moreover, in the latter case, 45% of new patients a physician receives is assigned to him or her by referrals. In the particular case of Portugal, it registers a referral rate of 5.56% between primary care and secondary care (Ponte et al., 2006).

2.5. Patterns in the referral Mechanism between PCP and SP

It is in the interest of policy makers and healthcare organizations to understand the patterns behind the referral networks as they attempt to control healthcare costs by manipulating how both types of doctors establish their relationships. Another reason is to cultivate the level of awareness regarding the patterns of referral is to maintain the referrals within a specific institution or organization (Barnet, Song, & Landon, 2012).

The existent referral mechanism between a primary care doctor and specialists is affected by a variety of complex factors as you might see in table 2 - Factors and Reasons for referral between Doctors [5]. These factors can go from the previous experience a patient had with a certain doctor, as the probability of a referral can depend on the gender of the physician or patient.

Factors & Reasons for referral between Doctors	
Medical skill	Anxiety regarding outcomes
Perceived clinical expertise	Perceived pressure to control referral costs
Patients past experience with a physician	Restricted ability to obtain surgical referrals
Timely availability of a doctor for appointments	Hours of patient care per week
Sharing medical records	Visits per day
Being referred by another doctor	Physician income
Work in the same hospital	Physician income structure
Type of doctor (pcp or sp)	Ownership practice
Doctor specialty	Level generalization/expertise of a pcp
Patient diagnose	Convenience
Disease severity	Risk aversion
Future impact of a referral on a patient's health care	Race
Getting advice on diagnosis or treatment	Adequate patient time visit
Overall shortage of physicians	Nurse involvement in the process
Costs of tests and procedures	Managed care contracts
Age over 17	Use of HITS
Patient gender	Patient Has chronic conditions
Physician gender	Specialist refers the same pcp
Fear of lawsuit of not consulting an expert	Pcp refers the same Specialist
Patient health insurance	Years of practice
Anxiety due to clinical uncertainty scale	

Table 2 - Factors and Reasons for referral between Doctors

[5] These table contains factors from multiple sources.

A few papers have studied the referral relationship between general practitioners and specialists, the reasons of their choices and their patterns which can extend beyond professional motives (Barnett et al., 2011). For example, according to the study Trends in Physician Referrals in the United States 1999-2009 the most influencing factors affecting referrals were: clinical report, Physician gender, Years of practice, Specialty, Herfindahl Index[6](Barnett, Song, & Landon, 2012).

When identifying the professional network of each pcp in the overall network, 66% of the referrals made by them were to colleagues inside their network (Barnett, Keating, Christakis, O'Malley, & Landon, 2012). Family doctors give higher relevance to reasons related with a physician access to one another. The level of availability of a specialist for appointments, the sharing or not of medical records, working in the same hospital or not, if a patient has insurance or not, and what type of insurance (Kinchen et al., 2004). In addition, physicians with less capabilities of handling clinical uncertainty were more likely to refer a patient to other doctors (Barnett et al., 2011). Moreover, in case the care that a patient requires exceeds the expertise that a primary care physician can give, the probabilities of referral increase (Franks, Williams, Zwanziger, Mooney, & Sorbero, 2000). Additionally, physicians have less encouragement to increase their volume of referrals if payments are bundled[7] rather than discriminating for every service (Jauhar, 2019).

Other sources have found evidence that doctors when referring patients to other colleagues, consider the professionals with who they are familiar with. However , they first consider the ones working at the same hospital (An, O'Malley, & Rockmore, 2018).

It is common for studies to register variations according geographic regions. It is not abnormal when specialists are found to have a more central role in smaller networks than larger. Despite physicians being assigned to one or more hospitals, the closer proximity to different cluster suggests the existence of multiple ties across different hospitals. In addition, the individual network of a doctor is more likely to have more and stronger ties within the hospital which the professional is affiliated to, despite not always being the verified scenario. Male doctors are more likely to have ties with other male doctors. The same happens with the female gender. Doctors with connections between them reveal a closer proximity in age when compared with unconnected individuals. Furthermore, this study found that not only connected physicians are more likely to be in closer geographical proximity than unconnected doctors as well to have patients with similar medical complexity (An, O'Malley, Rockmore, et al., 2018). Another study supports that doctors professional networks vary across geographic locations (An, O'Malley, Rockmore, et al., 2018). Additionally, it is valid to expect that doctors share patients with colleagues that have personality similarities with them, therefore they demonstrate to have the homophily concept present in their networks. However, the rapid adoption medical electronic records can lead to different findings (B.E. et al., 2012).

Other important factors in the referral process include timely communication of information regarding doctors' appointments and the need for the referral. When a doctor does not receive

---

[6] In this particular case the Herfindahl Index attempts to define the level of expertise of a primary care physician. A doctor who scores one in this index means that it is a specialist in a specific area. However, if the result is zero it means that he or she a generalist (has a general knowledge about different areas).

[7] Payments that covers all physicians' services and hospital care for each patient.

those information's in proper time or did not receive them at all, the quality of his or her treatment and patient satisfaction can be compromised. In addition, if a situation like this is verified it might indicate a pathway of communication that is not being efficient (Care et al., 2019). The use of healthcare information technology (HIT), patient-centered medical homes (PCMH)[8], and accountable care organizations (ACO)[9] have been contributing to the reverse of such trend as it is a facilitator for communication. Furthermore, a poor level of communication between the intervenient parties in the referral process suggest that the quality of the healthcare organization is threatened. Including nurse care in the treatment process is associated with higher rates of receipt of information from referrals and consultation as it helps to coordinate the care of a patient and there is a greater likelihood of greater inter-speciality communication (O'Malley, Ann S., MD & Reschovsky, 2011).

A patient diagnostic has a significant amount of explanatory power over the variation of referral rates of a pcp as the impact that a referral can have less or significantly higher in the future of a patient health. Additionally, the patient expectations and demands for seeing a specialist also increase. Furthermore, patients presenting common pain problems or depression symptoms have lower likelihood of referral. The costs of procedures and surgical interventions (specialist care) represent a constraint towards the referral rates as it is common for policy makers on healthcare organizations to implement barriers against inappropriate referrals. Moreover, so far it has not been found an association between the rate of referrals and their quality (Forrest et al., 2006).

The ethnicity of a patient can also have some explanatory power as communities where the degree of concentration of black residents is high have a lower tendency to get treatment when compared with white individuals from the same population (Ghomrawi, Funk, Parks, Owen-Smith, & Hollingsworth, 2018). Additionally, black doctors report to have a perception of racial discrimination by white doctors in the referral process (Kinchen et al., 2004).

Sometimes a certain degree of financial transparency from healthcare providers can be demanded to their employees when seeing patients to build stronger relationships between the two parties. In addition, it is common for countries to have laws that prevent doctors to pay other doctors directly for referrals (Schroeder, 2016). Moreover, the availability to a pcp of healthcare information technology, nurse care manager, adequate visit time and quality reports regarding patients with chronic conditions would be expected to increase the referral rates by from 63.9% to 82.7%. The increased administrative burden general physicians face along with diminish of reimbursements are creating heat on those doctors to see more patients in less time. Having this kind of structure decreases the effectiveness of the communication in a referral network and the quality of the care provided. Moreover, systematic structures, tools and processes for information creation, sharing, receipt and recognition from both parties, the sending and receiving are necessary to support medical care practices (O'Malley, Ann S., MD & Reschovsky, 2011). An important consequence to bear in mind when increasing the number of referrals, is that, that implies also an additional increase in the number of ambulatory visits for the average person, either in the primary care services and

[8] Team based model coordinated by a personal doctor who continuously manages a patient care throughout his or her lifetime to maximize its health.

[9] Groups of health care providers that can represent teams of doctors to constituting an entire hospital, which volunteer to give high-quality treatment for Medicare patients.

specialist care (Barnett, Song, et al., 2012). The study Trends in Physician Referrals in the United States, 2012 hypothesizes that an increase in referrals might be due to the increasing complexity of care itself, for instance common diseases with general or viral symptoms did not register such increase.

2.6. Social Network Analysis

Social Network Analysis (SNA) in simple terms is a research technique that allows to map the existent informal structure of information-knowledge sharing behaviour within a specific network with application in diverse areas such as health, business and others (Clark, 2006). In other words it allows to map the existent relationships (edges) of the individuals (nodes) of an organization, which do not necessarily match with its formal structure(Cross, Parker, & Borgatti, 2000).

Social networks are groups of nodes and edges where nodes can correspond to social entities such as organizations or individuals and edges represent the existent relationships between those actors. Networks can be classified according to their direction (directed or undirected)[10].Based on its origin[11]. Based on their centrality they can be defined as ego-centric or socio-centric[12]. Additionally, they can be classified according to their mode[13]. Finally, a network can be classified as weighted or unweighted[14].

Depending on the context SNA allows individuals to answer a wide variety of questions, such as: Who are the most relevant actors in the network? To whom people turn in advice? Have smaller clusters emerged from within the overall network. How knowledge actually flows among its actors? (Cross et al., 2000). SNA can be used to study the structure of a social system and to understand how this structure influences the behaviour of its actors (Kenis & Oerlemans, 2009).

Furthermore, is a relevant tool that allows to identify the individuals you have a central position in the network, and therefore have a deeper impact positive or negative in its level of effectiveness. When performing a Social Network Analysis, it is possible to analyse the following situation: The existence of bottlenecks[15]. The number of existent links in a network is it sufficient to coordinate the network efficiently? What is the average distance between the nodes? If the distance between them is relatively short, it is likely that the information traveling in the network is received in a timely pace. However, if they are too long is likely that the information traveling in network, takes too long to reach its final destination and it is likely to be distort. In addition, it can also be relevant to analyse the actors that are isolated in the network, or in other cases the ones that are in the periphery of the

[10] A directed network consists of network where the relationships between two different nodes are not necessarily mutual (Song, B. Keller, & Zheng, 2017).

[11] Depending on their origin social networks can be classified as explicit or implicit, one is the reverse of the other. Implicit network do not exist by default they have to be intentionally built (Frey, Jégou, & Kermarrec, 2011).

[12] Ego-centered networks are represented by a set of nodes that are within a certain distance of focal node, the central actor. Socio-centered networks represents a set of entities that are connected between each other (Directions, n.d.).

[13] Networks can be classified accordingly to the level of granularity regarding the number of social entities. If there are two different social entities being represented it is a bipartite network, also known as two-mode network (Song et al., 2017).

[14] Different ties can have different weights, strengths. The weight of a tie between an individual and an acquaintance is weaker than a tie with his or her best friend (Directions, n.d.).

[15] Individuals who are not contributing for an efficient flow of information in the network (Cross et al., 2000).

network, as they can be underutilized resources. Last but not least, it is important to analyse possible cluster that can emerge from an informal network(Cross et al., 2000).

2.7. Social Network Applications and Mapping

Despite being a recent topic of research, social network applied to the referral mechanism between primary care doctors and specialist there is some consensus on how to map the possible existent informal relationships. According to the study Variation in patient-sharing networks of physicians across the United States. *JAMA - Journal of the American Medical Association*, the connection can be formed under the assumption that two different types of doctors have a connection when they share a patient. Doctors who shared at least 8 patients are 80% more likely to have a valid information sharing relationship (Barnett et al., 2011).

Other studies declare to exist a connection when besides sharing a patient, the time interval between each doctor’s appointment respects a maximum of 30 days (An, O’Malley, & Rockmore, 2018).

Moreover, the weights of the links between the types of doctors varies according to the number of patients they share, meaning the more patients two doctors have the stronger is the connection (B.E. et al., 2012). Moreover, the same study to eliminate potential existing false positive relationships suggests to define a threshold that retains a percentage of the strongest links (B.E. et al., 2012).

3. METHODOLOGY

The current section represents the process of the development of the dissertation since the data pre-processing until all the tools and analysis made in order to better understand the referral mechanism between general practitioners. Additionally, it is presented the reasoning, assumptions and limitations that arose during the construction of the primary-specialty referral network and other frameworks.

3.1. INTRODUCTION

As previously mentioned, the objective of this study is to better understand the referral mechanism between primary care doctors and specialists using network science, and other tools. To do so a referral network between the two types of doctors is estimated. Moreover, in this network the referral patterns and certain social network metrics such as degree centrality will be explored. Finally, a regression analysis is produced with the aim of understanding if the different types of information available, that mean respect to the background of a doctor, including their centrality measures, if they have any impact in the referral rate associated to each doctor.

The database used for the development of the research study was provided by a European healthcare provider in 2017. The originated from Portugal. The database management software used for the storage of the data was Dbeaver. The Programming Languages used for the cleaning, visualization of the data, networks construction and analysis were python[16] and postgresql. The development of the dataset necessary for the progress of this report can be divided into two main stages. The first phase consisted in extracting, cleaning and visualizing all the relevant information regarding a doctors' profile. The second stage (explained in the chapter "Schema Evolution") involved adding and validating new information to complement the initial dataset.

[16] Besides using the traditional libraries of python for data analysis like matplotlib, pandas, and other, two packages were used in particular NetworkX.

3.2. DATA PRE-PROCESSING

This stage represents the selection, integration, cleaning, description, and visualization of final dataset which is used for the development of the respective dissertation. The development of the final dataset departed from a table called initial_doctor_table. This framework initially had 3500 observations, each corresponding to a unique doctor, being described by 13 variables. From those variable 7 represent text type variables, 3 bigint type, 1 numeric type variable and 2 timestamps with time zone variable. In addition, each variable has a role. There is 1 ID variable, 6 variables defined as interval, 4 variables defined as Nominal, 1 variable defined as Binary, and 1 variable declared as Ordinal (see table 3 – Initial_Doctor_table).

Variable	Variable Type	Role	Description
Doctor_id	Text	ID	Unique identifier of a doctor
Birth_year	Bigint	Interval	Birth of the corresponding doctor
Gender	Text	Binary	Sex of each doctor
Specialty	Text	Nominal	Area of expertise of a doctor
Specialty_2	Text	Nominal	Second area of expertise of a doctor
Academic_Degree	Text	Nominal	Level of education and the training area of a doctor
Years_Of_Practice	Bigint	Interval	Number of years of how long ago a doctor received a school certificate
School_Certificate	Text	Ordinal	Certificate for the conclusion of a certain stage of medical education
Grade	Numeric	Interval	Final grade of doctor school certificate
Learning_Location	Text	Nominal	Higher education institution, location, or both at which an individual concluded a certain stage of his or her medical education
Number_Of_Workshops	Bigint	Interval	Number of workshops a doctor took until 2017
Initial_Date	Timestamp with time zone	Interval	Initiation date of the school certificate
End_Date	Timestamp with time zone	Interval	End date of the school certificate

Table 3 - Initial Doctor Table

3.2.1. Missing Values

The first step to treat the initial doctor table, was to handle the existence of missing values in it. A particular problem was identified related with this issue. Some values in some variables were in blank (not null or erroneous values). Those variables were: specialty, specialty_2, Academic_Degree, School_Certificate, and Learning_Location. This is relevant because the meaning of a blank value is assumed to be the same as a null value. Therefore, to correct this situation the blank values were updated to null values. The output of such action produced the following amount of missing values that is possible to observe in table 4 – Initial Doctor Table Missing Values.

Variable	Number of Null Values	Percentage of Null Values
doctor_id	0	0,0%
birth_year	0	0,0%
gender	0	0,0%
specialty	2096	59,9%
specialty_2	3488	99,7%
Academic_Degree	226	6,5%
years_of_practice	0	0,0%
school_certificate	28	0,8%
grade	3136	89,6%
learning_location	28	0,8%
number_of_workshops	0	0,0%
initial_date	28	0,8%
end_date	1559	44,5%

Table 4 – Initial Doctor Table Missing Values

Considering the amount of null values in some variables, some had to be dropped because they could not add any value or information to the dataset. Therefore, the variables specialty_2 and grade, were dropped due to the fact that they had 99.7% and 89.6% of missing values.

Additionally, the variables initial_date and end_date were removed also due to the significant amount of missing values in the latter case. In addition, Academic_Degree is considered to be redundant because the information that is possible to extract from it, is provided by other variables. The different values that the variable can take can be seen in table 5 – Academic Degree Values. Furthermore, the variable number_of_workshops do not carry relevant information for the development of the dissertation. Therefore, 6 different variables were excluded from the study.

Academic Degree
L. Med. Mestr. Medicina Legal
L. Enfermagem
L. Medicina
L. Med. PG Medicina Legal
L. Medicina Dentária
L. Fisioterapia
L. Med. Dout. Medicina
L. Med. PG Saúde Pública
L. Desconhecida / Não especificada

Table 5 - Academic Degree Values

3.2.2. Outliers

Concerning the process of identifying outliers, the distribution of the quantitative variables and their respective descriptive measures were assessed, namely: Birth_Year, Years_of_Practice. The distribution of the variable birth_year, presents to have a slightly skewed distribution to the left. The youngest individual in this sample was born in 1994, while the oldest person was born in 1900, making it to have 117 in 2017. Approximately 75% of the population was born before or in 1980 (37 years old in 2017) (figure 1 – Birth Year Distribution). The average age of the sample is set to be 48, which means that the average birth year was 1969. Additionally, every individual claiming to be born before 1935 is considered to be an extreme value. By applying such rule, we are excluding only 2 observations.

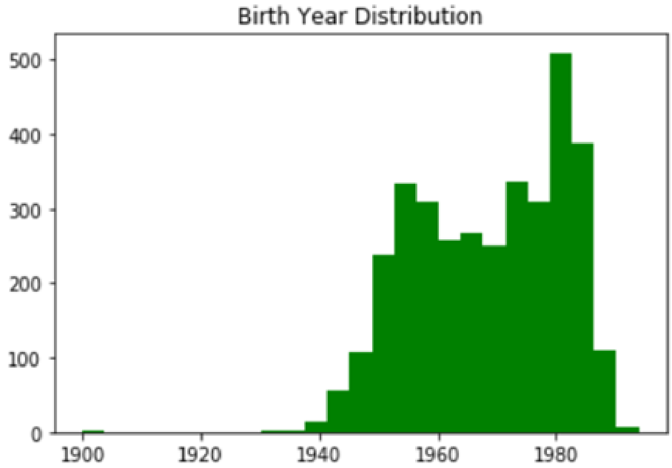


Figure 1 - Birth Year Distribution

The variable years_of_practice presents also a skewed distribution, but in this case is skewed to the right. Meaning the more years of practice, a doctor has the, the fewer are the number of doctors that are able to represent those years of practice. In this sample, on average, doctors finished their studies six years ago. Approximately 75% of the population finished their degrees no more than nine years ago. The range of this variables goes from zero to 39. The value zero can have two different meanings in our sample: It can either mean that a doctor has not finished his or her studies, or it has just finished them in 2017. In terms of outliers, it does not demonstrate to have extreme values (figure 2 – Years of Practice Distribution).

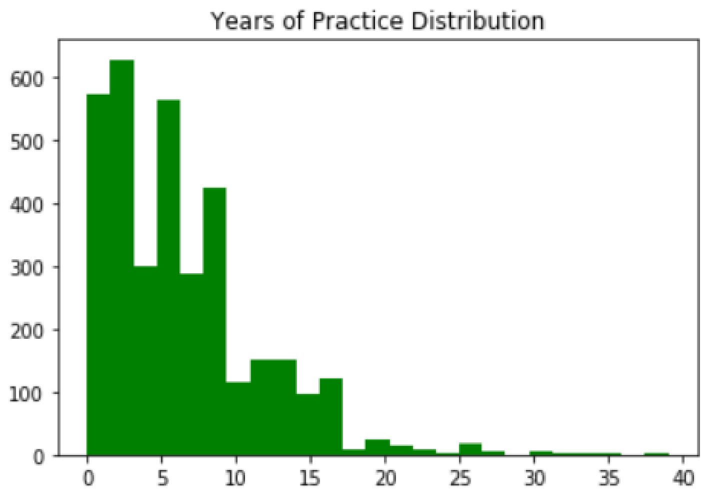


Figure 2 - Years of Practice Distribution

After applying the necessary rules to exclude the outliers from the sample only 1 observation is being removed, which represent 0.06% of the entire dataset. It is possible to observe it more detail in (table 6 – excluded outliers).

Variable	Excluded Observations	Excluded Observations %
Birth_year	2	0,06%
Total	2	0,06%

Table 6 - Excluded Outliers

3.2.3. Data Validation

To become a doctor is necessary to do at least six years of medical school (university), and a minimum of 5 to 7 years to conclude a medical post-graduation known as Internato Médico, that is divided in 1 or 2 years of Internato Geral and a minimum of 4 years to conclude the Internato Complementar, which can reach a maximum of 6 years depending on the specialty. Therefore, it is plausible to assume that the minimum age at which a doctor initiates its medical career in 2017 is at 29 years old. Before 2004, included, what is known today as Ano Comum of the Internato Médico, was in fact recognized as Internato Geral and it had a duration of two years. In the year of 2004 the Internato Geral started to be recognized as Ano Comum and its duration was modified to 1 year (Diário da República, 2004). However, it is important to know, that after concluding the general training of the Internato, a doctor is recognized as an autonomous individual, becoming able to start giving prescriptions to patients and scheduling appointments (D. da República, 2018a). In addition, since 2012 it is now possible for the private healthcare sector to start training medicine students, in order for them to get their post-graduation certificate. However a private healthcare provider has to be first certified by the institution Ordem dos Médicos[17] (Saúde, 2017).

Thus, it should not be appropriate to include individuals that have less than 26 years old in 2017. Because assuming that a person did not make any detour in its academic studies, a doctor has finished its high school degree with 18, finished its college degree with 24 and start its internato with 25 years old. Given that it's only possible to make referrals after concluding the first year of the internato it is not correct to have individuals in the sample of doctors that have 25 or less in 2017. Furthermore, as previously mentioned the objective of the study is to understand the referral mechanism between general practitioners and specialists, therefore, despite being valid to have individuals with 25 years old, they cannot make referrals, because they are in their first year of internato. They have to have at least initiated their second phase of the Internato Médico in 2017. Only then a person can schedule appointments and make prescriptions. By applying these rules only 1 observation is being excluded.

[17] The Healthcare Ministry is the official entity responsible for overseeing the practice of medicine in Portugal, ensuring its quality, security, and ethics (Governo da República Portuguesa, 2017). Additionally, the Administração Central do Sistema de Saúde (ACSS) has established protocols with other entities (Governo da República Portuguesa, 2019) such as Ordem dos Médicos and Conselho Nacional do Médico Interno that helps the Healthcare Ministry to manage the structure and oversee the post-graduation Internato Médico (Ordem Dos Médicos, 2018).

It is important to point out a limitation to the aforementioned rule which is the fact that it is not applicable for cases in which a doctor has more than 26 but it is in fact in its first year of internato. If considered the observations excluded plus the observations being left out, due to inconsistencies issues, the number of records not being considered represents approximately 0.08%[18] of the entire dataset.

3.2.4. Other Data Problems

The situation of the variable learning_location needs to be specifically addressed considering that it is a variable of extreme importance and it was also a variable seriously corrupted. Not only the information regarding where doctors did their internato and university was stored in the same variable, as well different individuals which attended the same institutions wrote different versions of their names, which end up creating 1064 unique institutions. The best solution found that for that problem was to manually check each value and make the data uniform. Moreover, given that some of doctors when writing the information regarding their institution or institutions, inserted the group of the hospital in which they did the internato the same was done to the other values. Meaning in a scenario in which we had for example the value "Hospital de Santa Maria", this would be converted to "Centro Hospitalar Universitário Lisboa Norte". Furthermore, in the potential case of having values that represented unique institutions that belonged to groups of hospitals that are currently closed they were converted to that same group anyways. Finally, the cases in which was not possible to recognize an official institution, the value was converted to "Desconhecido".

[18] The values include 2 observation from the exclusion of outliers and 1 observation which do not comply with the consistency rules.

3.2.5. Schema Evolution

3.2.5.1. Introduction

This dissertation originally departed from a table called Initial_Doctor_Table which composition can be seen in table 3.

To enforce the consistency rules and exclude the doctors determined to be outliers a new table was created called, doctor_table (figure 9 – Doctor Table) which is derived from the Initial_Doctor_Table (Table 3 - Initial_Doctor_Table). The decision process regarding the outliers is registered in the chapter "Outliers" section 3.2.2. Additionally, the reasoning for the development of the validation rules is under the chapter "Data Validation" section 3.2.3.

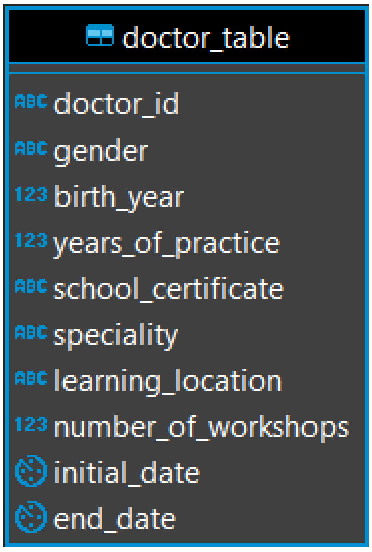


Figure 3 - Doctor_Table

3.2.5.2. Complementary Tables

Four different complementary tables were created, with the objective of solving specific issues with the data. Table institutions was a table created with the sole purpose of fixing the variable learning_location, because it required an extensive query to solve the respective problem. It was more reliable to develop such query on the side. Doctor_id was included in the table to facilitate the integration of the variable learning_institution later on. The variable learning_institution, has the corrected values originally from the table initial_doctor_table under the variable learning_location.

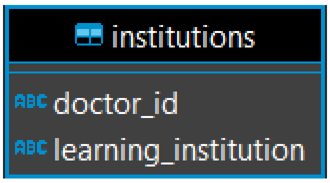


Figure 4 - Institutions

The table ic_specialities was created with the objective of validating the values regarding the doctors specialty inserted in the doctor_table. e1xpertise_id represents the unique identifier of the different areas of expertise. Expertise represents all the possible specialties a doctor can be trained for, that are recognized by the ACSS and Ordem dos Médicos. Expertise_category represents the respective category in terms of primary care or differentiated care. Expertise_category_id represents the unique identifier of each category.

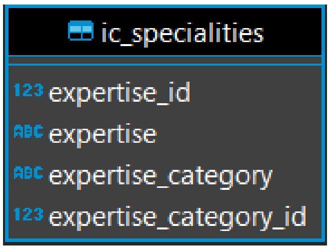


Figure 5 - IC_Specialties

Doctor_category is a table that represents an additional source of information regarding the variable specialty, existent in the doctor table, which was necessary due the fact that it has approximately 60% of its values missing. Without this variable is not possible to determine to which category a doctor belongs to. It was possible to reduce the amount of missing values to approximately 34%. Doctor_class represents the category associated with the type of specialty, it takes either the values “Cuidados Primários” (primary care) or “Cuidados Secundários” (differentiated care). Doctor_id is the

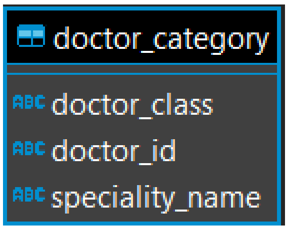


Figure 6 - Doctor_Category

unique identifier of each doctor. The variable specialty_name, as the name suggests represents the area of expertise of a doctor.

Working_hospital was created with the objective of assigning the hospital or hospitals in which a doctor is currently working in, to the respective doctor_id, which represents the unique identifier of a physician. Hospital_abbreviation consists in the shortening of the complete name of the hospital. The variable hospital has the complete name of the different hospitals. The hospital_code is the unique identifier of a hospital. Due to confidentiality reasons the names or abbreviations of the hospitals will not be disclosed.

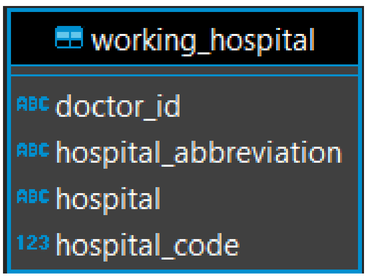


Figure 7 - Working_Hospital

3.2.5.3. Doctors Background Table

After creating each complementary table, a new table called final_doctor_table was designed with the purpose of integrating the different sources of information (figure 14 – Final_Doctor_Table).

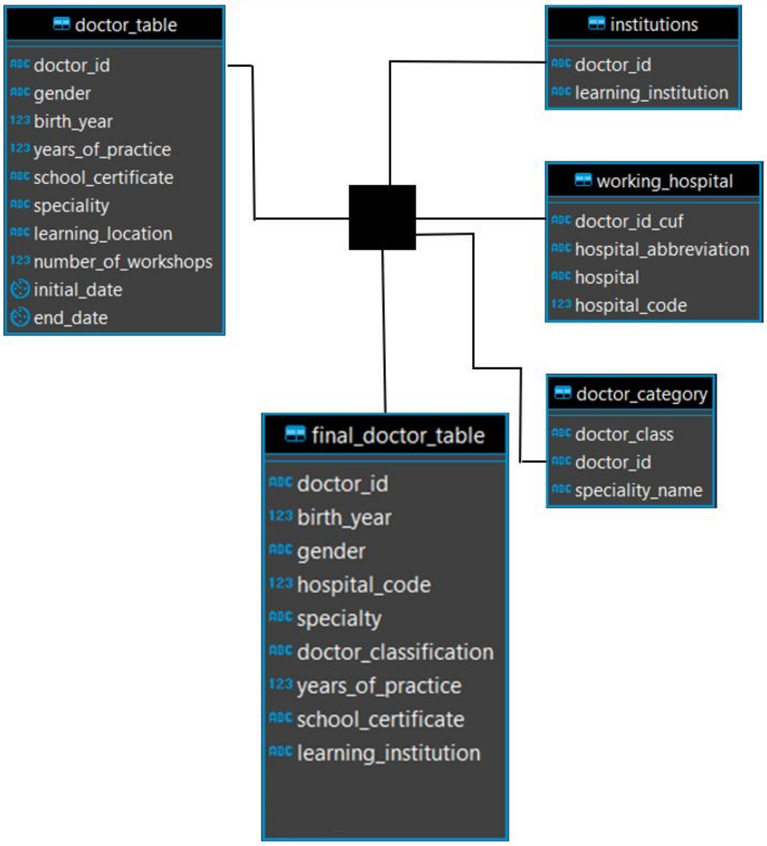


Figure 8 - Final_Doctor_Table

3.3. PRIMARY-SPECIALTY REFERRAL NETWORK

3.3.1. Primary-Specialty Mapping

The primary-specialty referral network was built under the purpose of mapping existent relationships between doctors that follow under the category of general practitioners, and physicians that practice specialties that belong to the differentiated care category. In order to do such, it was necessary to import 4 distinct variables: the date of each episode[19], the unique ids associated with each patient and doctor, regarding each episode, and classification the of each doctor.

The connections estimated to existed in this network, are based on four main assumptions: It cannot be longer than 30 days that between two appointments with different doctors. Only primary care doctors can make referrals. In addition, two distinct doctors have to share a patient. Additionally, in the case of a doctor having one specialty from each category, primary and differentiated care, only the former case was kept. However, if a doctor had two specialties from the same category only the one with the highest represent among the different specialties would remain. Finally, it is not possible to exiting referrals within the same group of doctors.

Initially, the data of the referral network had to be filtered according to the doctors in the final_doctor_table, which contains all the relevant information regarding the doctor's background. Moreover, the data was restricted to episodes that did not have urgency character, that were not cancelled and in which the date of the episode was available. The final output was a significant large sample with 9 133 477 observations. Afterwards, the dataset was filtered between the years of 2012 and 2017 to ensure a better quality of the data. By doing such, the main dataframe was reduced by 444 observations. Interestingly enough, 137 doctors where identified has not giving a single consultation between those years. Therefore, these doctors will not be taken into account in this research project, leaving us with 2171 doctors in total. However, it would be prudent to verify if such doctors are still active or not. Additionally, in the remaining 9 133 033, it was identified 1 305 361 unique patient ID's. Furthermore, it was registered 1 913 674 unique consultations on behalf of primary care doctors and 4 112 885 on behalf of specialists. Given that the objective is to study the referrals of primary care doctors to specialists, the sample of specialist should only represent doctors who are associated with patients present in the sample of general practitioners.

Unfortunately, the output of the relationships estimated is not valid, as there are multiple doctors which are associated with more referrals than consultations. This means they were estimated to give more appointments than the ones they actually gave. The next best solution found, was to only allow one referral for each unique primary care doctor consultation. The criteria to choose the link that would be assumed as the estimated relationship would the one with the shortest period. This represents a two important limitation to study. Not only, a general practitioner can refer a patient to more than one doctor of the same or different specialties, as well the link with the shortest period is not necessarily the referral that happened in reality. This resulted the estimation of 59 065 unique relationship between doctors.

[19] Episode represents the consultation of a patient. It's unique per visit.

The next toward the exploration of the relationship between the two types of doctors was built the actual network which was done using a library in python called Networkx. Given that it is only possible to have connections between doctors of distinct groups, the network was defined as a bipartite network (figure 9 – Bipartite Network).

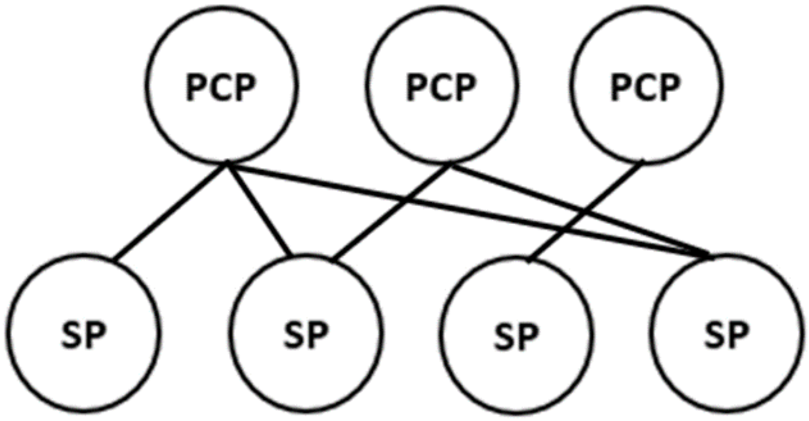


Figure 9 - Bipartite Network

Modelling a social network as a Unipartite network (figure 10 – Unipartite Network) would make it harder to understand the dynamics between the two types of doctors as the analysis and calculations of social network metrics are done are not specific to each type of doctor or do not have in consideration that condition.

Defining a complex network as two-mode imposes some restrictions and limitation both in the creation and analysis processes. Bipartite networks do not allow to have links within each of the groups in the network, only connections between doctors of different groups are permitted.

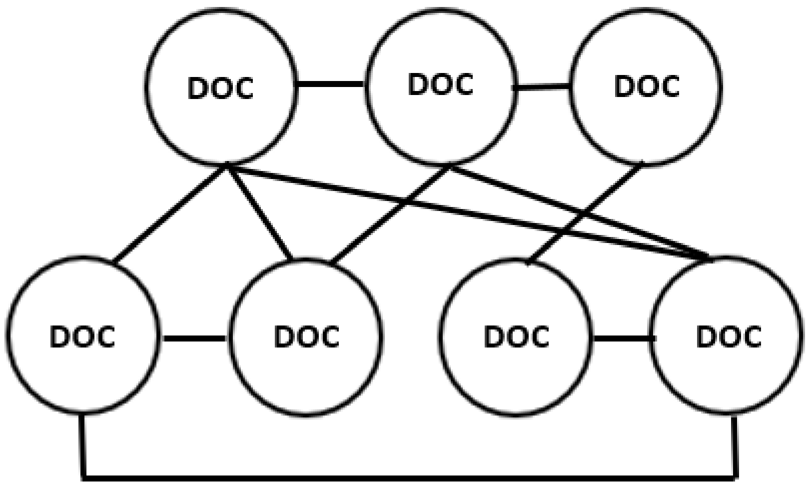


Figure 10 - Unipartite Network

Furthermore, the primary-specialty network was computed as a weighted network. In the particular case of this dissertation the weight assigned to an estimated link represents the number of patients each two doctors are expected to share. The nodes with whom there were no connections identified were included in the network grid as isolated nodes. Additionally, this network can be classified as socio-centered, implicit and undirected. To conclude this primary-specialty network does not necessarily corresponds to the formal structure of the private healthcare organization.

3.3.2. Edós-Renyi Random Network

A Érdos-Renyi Random network was necessary to build in order to compare the different characteristics of a random network against the macro and micro-level metrics found in the primary-specialty network. It was generated has a Bipartite random graph, where the size of the nodes is the same has the ones in the primary-specialty network. In addition, when creating a network of this nature it is necessary to assign a specific value to a specific parameter called p[20] (figure 11 – Erdós-Renyi random network edge probability), which represents the probability of an edge between two nodes being computed.

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

Figure 11 - Erdós-Renyi random network edge probability

[20] In the equation represented in figure 11 - Erdós-Renyi random network edge probability, μ embodies the weighted average degree of the primary-specialty referral network while symbolizes the number of nodes inserted in the network.

3.3.3. Centrality Measures and Others

3.3.3.1. Degree Centrality

The degree centrality measure in a bipartite network of a doctor is represented by the number of direct ties he or she has, divided by the total possibilities of the opposite partition (figure 12 – Bipartite Degree Centrality Equation). On the other hand, if the structure of the network was unipartite, the numerator would remain the same, but in this case the denominator always embodies the same value, which is the total number of nodes in the network (figure 13 – Unipartite Degree Centrality Equation).

$$\frac{\#\ Direct\ ties}{\#\ Nodes\ in\ the\ opposite\ partition}$$

Figure 13 - Bipartite Degree Centrality Equation

$$\frac{\#\ Direct\ ties}{\#\ Nodes\ in\ the\ network}$$

Figure 12 - Unipartite Degree Centrality Equation

Furthermore, to better understand the metric degree centrality a network visualization is provided with the objective of demonstrating which nodes have higher degree centrality. In figure, it is possible to observe that the nodes with higher degree centrality are general practitioners 2 and 4 and specialist 2 (Appendix 9.2 Degree Centrality Example Calculations)

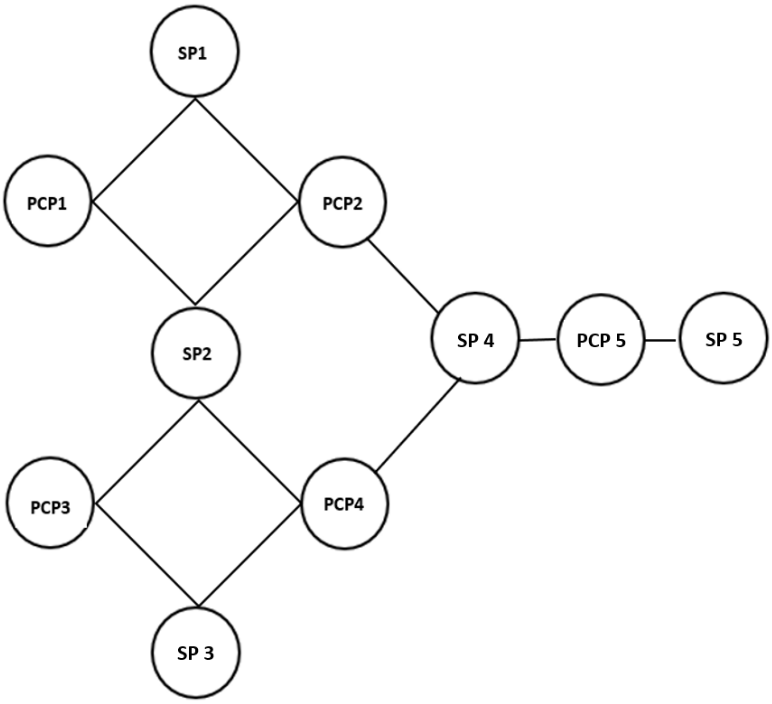


Figure 14 - Degree Centrality Example

3.3.3.2. Betweenness Centrality

Betweenness centrality is calculated similarly in both types of networks, bipartite and unipartite. For each node it is the number of occurrences on the shortest paths of each pair of nodes in the network (figure 14 – Betweenness Centrality Equation[21]). The main difference when computed in a two-mode network is that this metric is still normalized by the maximum possible betweenness centrality values that can be found for each node partition in a bipartite network. The equation for the calculation of such values can be seen in Appendix 9.3 Betweenness Centrality Normalization.

$$C_B(v) = \Sigma_{s,t \in v}\left(\frac{\sigma(s,t \mid v)}{\sigma(s,t)}\right)$$

Figure 15 - Betweenness Centrality Equation

Moreover, to have a clearer sense of what this metric is calculates it is possible to observe that the most important bridges in the example on figure 15 – Betweenness Centrality Example, are the primary care doctors 1 and 2 and the specialist 2. Specialist number 2 is on every shortest path between each node of the cluster on the left and right and vice-versa (Appendix 9.4 Betweenness Centrality Example Calculations).

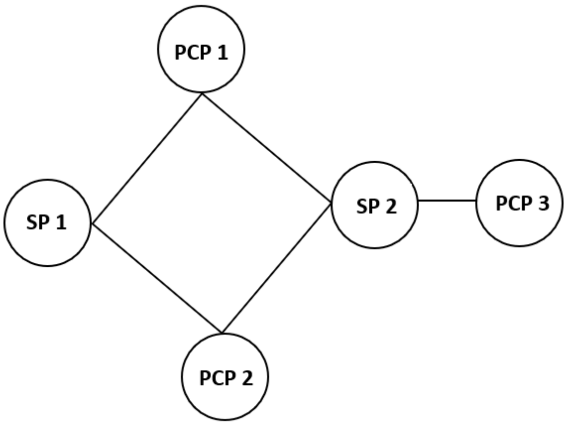


Figure 16 - Betweenness Centrality Example

[21] The parameters s and t represent each pair of nodes. The v parameter represents the node being scoped for the number of shortest paths in which an individual might be in.

3.3.3.3. Closeness Centrality

The last centrality measure calculated in this study was the metric Closeness. The purpose of such metric is to measure the distance of each node to the remaining ones in this network. In addition, given that this metric is being applied in a bipartite network it is normalized by the minimum distance possible, which in the case of a general physician to a specialist is 1 and in the case of doctors within the same partition is 2. The formulas of each group of doctors can be seen in figure 16 – Closeness Centrality Equations[22].

$$C_v = \frac{m\ + 2(n-1)}{d}, for\ v\ \in\ U,$$

$$C_v = \frac{n\ + 2(m-1)}{d}, for\ v\ \in\ V,$$

Figure 17 - Closeness Centrality Equations

In the figure below, figure 18 Closeness Centrality Example it is possible to observe that the nodes that are more central in the network are primary care physicians 1 and 2 and specialist 2 because when compared to the remaining individuals in the network, they are the ones that can reach everyone in the network faster given their position in the network being more central (Appendix 9.5 – Closeness Centrality Example Calculations).

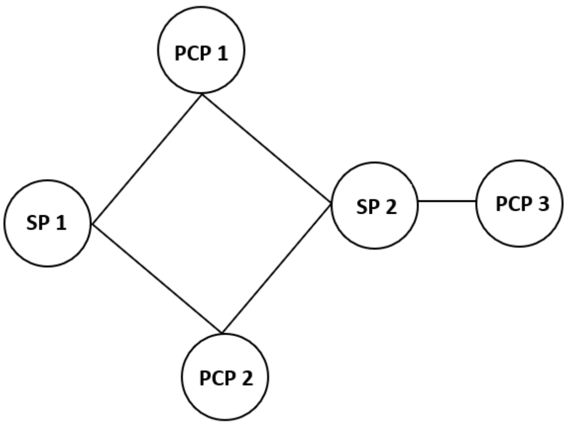


Figure 18 - Closeness Centrality Example

[22] The parameter m represents the number of nodes belonging to the group of doctors in partition V while n embodies the number of individuals in partition U. In both cases de letter d, is the aggregated sum of all distances of node v to the remaining nodes in the network. The normalization process is not reflected in this formula.

3.3.3.4. Global Clustering Coefficient

The global clustering coefficient is a measure that has the purpose of measure the probability of individuals to cluster together. In a unipartite network, what this metric does is to find all closed triangles in a network (figure 19 – Closed Triangle) over all triangles that are possible to be formed in

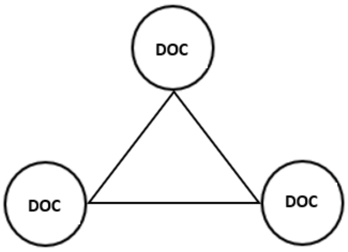


Figure 20 - Closed Triangle

a certain network. However, this is situation is not possible to be observed in a bipartite network as it is not possible to exist closed triangle (figure 20 – Bipartite Cycles). The alternative solution suggested is to find the total number of existing squares in the network over all possible squares that could exist in the network given its number of nodes per partition (figure 21 – Bipartite Global Clustering Coefficient Equation[23]).

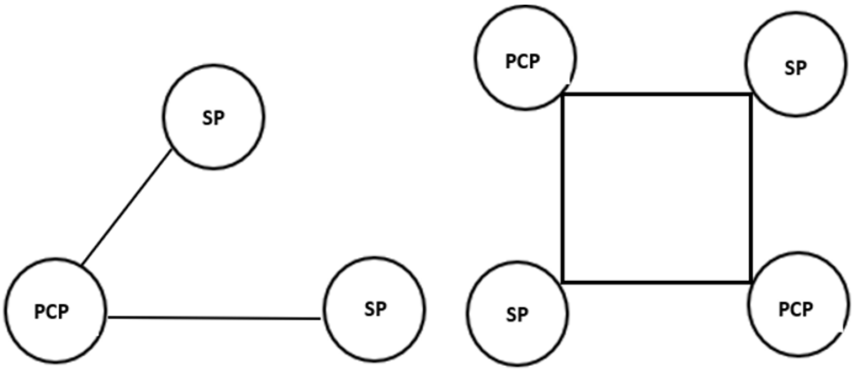


Figure 19 - Bipartite Cycles

$$C_{4,mn}(i) = \frac{q_{imn}}{(k_m - \eta_{imn}) + (k_n - \eta_{imn}) + q_{imn}}$$

Figure 21 - Bipartite Global Clustering Coefficient Equation

[23] In the equation represented in figure 21, m and n characterize a pair of nodes of i. The parameter q(i)mn is the number of existing squares in which i is an element. Kn and km are the number of nodes n and m repectively have. Finally, η(i)mn = 1 + q(i)mn + σ(i)mn. If m and n are connected σ(i)mn = 1, otherwise it's 0.

3.3.3.5. Community Detection

The algorithm used to detect possible existent communities in the network was the Louvain algorithm which is an optimization modularity-based algorithm. Communities in social network are no more than groups of nodes that are more densely connected when compared with other community. The Louvain algorithm measures the density between nodes through a metric called modularity. The algorithm starts by defining small communities through the optimization of aforementioned metric, locally on every node in the sample. Then the communities found are assembled into a single node respectively. The process repeats itself until the optimal number of communities is found, according the algorithm. Applying this algorithm represents a limitation on the current study as the algorithm is no appropriate for community detection in bipartite networks.

3.4. REGRESSION ANALYSIS

Two distinct regression analysis, representing both types of doctors were performed with the objective of testing the hypothesis that the variables available from the doctor's background and the centrality measures computed from the primary-specialty referral network had any statistical significance in the referral rate that each doctor makes or receives, depending on its specialty. Therefore, the dependent variable (referral rate) is defined as being the number of referrals a doctor has made or received, divided by the number of consultations he or she actually gave between 2012 and 2017. The isolated doctors present were included, with the dependent variable, and centrality measures equal to 0. The initial data can be observed in table 7 – Regression analysis variables. The model used was a multiple linear regression model.

Variable	Type	Attribute type
doctor_id	ID	ID
birth_year	Input	Numeric
gender	Input	Binary
specialty	Input	Nominal
number_of_hospitals	Input	Numeric
Doctor_classification	Input	Binary
number_of_specialties	Input	Binary
years_of_practice	Input	Numeric
school_certificate	Input	Binary
learning_location	Input	Nominal
degree	Input	Numeric
betweenness	Input	Numeric
closeness	Input	Numeric

Table 7 - Regression analysis variables

The variable hospital_code, which associates to each doctors the hospitals in which they work was converted into a new numeric variable called number_of_hospitals which as the name suggest represents the number of hospitals in which a doctor is currently working in. The variable number_of_specialties was created, with the objective of keeping the information which doctor had or did not have two specialties. The variable learning_location which mixes the university where a doctor pursued its medical degree and the institution at which a doctor made his post-graduation, was separated into 2 new variables, one called university and another one called hospital. The variable referrals were calculated has the sum of weights of each doctor in the network. The variable consultations as the name might suggest indicates the number of consultations each doctor gave between 2012 and 2017. In addition, the variable birth_year was converted into age. The output of the previous transformation can be seen in table 8 – Regression Analysis with New Variables.

Variable	Type	Attribute type
doctor_id	ID	ID
birth_year	Input	Numeric
gender	Input	Binary
specialty	Input	Nominal
number_of_hospitals	Input	Numeric
Doctor_classification	Input	Binary
number_of_specialties	Input	Binary
years_of_practice	Input	Numeric
school_certificate	Input	Binary
degree	Input	Numeric
betweenness	Input	Numeric
closeness	Input	Numeric
university	Input	Nominal
hospital	Input	Nominal
referrals	Input	Numeric
consultations	Input	Numeric
referral_rate	Target	Numeric

Table 8 - Regression Analysis with New Variables

Afterwards, all the variables that suffered a transformation were inspected. But before the dataset was divided in two. One representing general practitioners and another characterizing only specialists. From the 16 variables 5 of them were immediately ruled out. The variable doctor_id is composed only of unique values. The variable school_certificate was assumed not to carry any relevant information regarding the target variable, even if there was detected a linear relationship between the two variables. In addition, the variables concerning medical institutions were removed from the sample given the high number of missing values. In the primary care dataset, the variables university and hospital have a total amount of missing values of 53% and 38% respectively. Additionally, in the specialist's dataset the variable university and hospital have 79% and 53% of missing values respectively. Furthermore, the two samples were inspected for high values of correlation between each pair of the independent variables in order to avoid multicollinearity. The variables consultations and referrals not only reveal a high correlation between them, as well a strong correlation with the centrality measures, degree, betweenness, but not closeness (table 9 – PCP Regression Analysis Correlation, table 10 – SP Regression Analysis Correlation). This makes since the centrality measures were basically derived from those variables. Thus, the variables consultations and referrals will be excluded from the analysis on both datasets. Moreover, given that the centrality measures degree and betweenness share a high degree of correlation, betweenness will be removed.

	age	years_of_practice	closeness	betweenness	degree	referrals	consultations
age	1	0.405	0.120	0.195	0.226	0.209	0.211
years_of_practice	0.405	1	0.172	0.278	0.370	0.387	0.429
closeness	0.120	0.172	1	0.409	0.587	0.405	0.447
betweenness	0.195	0.278	0.409	1	0.821	0.864	0.765
degree	0.226	0.370	0.587	0.821	1	0.814	0.793
referrals	0.209	0.387	0.405	0.864	0.814	1	0.876
consultations	0.211	0.429	0.447	0.765	0.793	0.876	1

Table 9 – PCP Regression Analysis Correlations

´

	age	years_of_practice	closeness	betweenness	degree	referrals	consultations
age	1	0.440	0.042	0.086	0.101	0.096	0.165
years_of_practice	0.440	1	0.167	0.206	0.289	0.283	0.345
closeness	0.042	0.167	1	0.354	0.562	0.311	0.400
betweenness	0.086	0.206	0.354	1	0.744	0.693	0.706
degree	0.101	0.289	0.562	0.744	1	0.740	0.784
referrals	0.096	0.283	0.311	0.693	0.740	1	0.806
consultations	0.165	0.345	0.400	0.706	0.784	0.806	1

Table 10 -SP Regression Analysis Correlations

Furthermore, as multiple linear regression models are not able to work with non-numeric variables, all categorical variable had to be encoded (Appendix 9.7 – Primary Care Doctors Independent Variables, Appendix 9.7 – Specialists Independent Variables). However, once again to prevent the existence of multicollinearity in the model it was necessary to remove one column from each variable being encoded in order to avoid what is commonly known as the dummy variable trap. In the particular case of the family doctor's dataset, the variables gender_M, specialty_Pediatria, number_of_specialties_2.0 and number_of_hospitals_4.0 were excluded. Regarding the specialist's dataset, the variables gender_M, specialty_Medicina_Desportiva, number_of_specialties_2.0 and number_of_hospitals_6.0 were removed.

4. RESULTS AND DISCUSSION

The current section has the purpose of critically analyzing the empirical results found. Throughout the current chapter the aforementioned primary-specialty referral network, and regression analysis are used to better understand the existing dynamics between primary care physicans and specialists, and possibly how the relationship between the two of them might be affecting the healthcare provider. Firstly, the primary-specialty referral network is analyzed with the objective of identifying potential referral patterns in the overall network, and then in the communities identified within the network. Finally, the objective of performing a regression analysis was to check if variables like the age, social network centrality measures and other are affecting the number of referrals being made and received by doctors.

4.1. PRIVATE HEALTHCARE PROVIDER

4.1.1. Organizational Overview

The healthcare provider under analysis counted with 3500 doctors in 2017 according to their human resources department. However, due to impositions implemented during the pre-processing, our final sample contemplates only 2,171 doctors, which are associated to 7 of the hospitals under the care of the healthcare provider. From the 2,171 identified 473 represent primary care doctors and 1,698 represent specialists. In this particular sample, 37 doctors claim to have more than 1 area of expertise. From those doctors 15 of them are certified both as general practitioners and specialists. In addition, 18 doctors are certified in two specialties of differentiated care and 4 in two specialties of primary care.

The distribution of the doctors per hospital can be observed in figure 22 – Doctor Distribution per Hospital. Currently the private healthcare chain has 637 physicians working in more than one hospital. It is possible to highlight, that the majority of the doctors, are concentrated in the hospitals assigned with the codes 2,7 and 8, while the remaining hospitals register a lower number in their workforce. The distribution of the two types of doctors is more or less proportional to the size of the hospital int terms of doctors (table 11 – Doctor Classification Per Hospital).

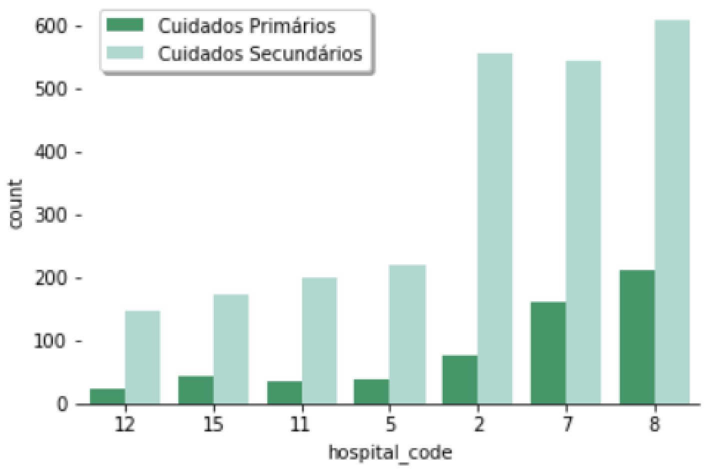


Figure 22 - Doctor classification per hospital

Hospital	Group	Doctors Count	Hospital Count	Distribution
2	PCP	76	631	12%
2	SP	555	631	88%
5	PCP	38	258	15%
5	SP	220	258	85%
7	PCP	160	703	23%
7	SP	543	703	77%
8	PCP	210	819	26%
8	SP	609	819	74%
11	PCP	36	237	15%
11	SP	201	237	85%
12	PCP	24	170	14%
12	SP	146	170	86%
15	PCP	44	216	20%
15	SP	172	216	80%

Table 11 - Doctor Classification Per Hospital

In terms of representation of doctor's specialties in the private healthcare workforce 44 distinct are identified. The areas of expertise that represent most employees (top 15) of the healthcare provider are identified in figure 23 – Top 15 Specialty Distribution. From the specialty Ginecologia/Obsterícia to Neurocirurgia they represent approximately 80% of the work force. If only considered the 5 most representative, they account for 42%. In addition, the three areas classified as primary care are represented in the top 15 specialties sample. In fact, pediatrics is the second specialty with the highest frequency (for more details Appendix 9.8 – Specialty Frequency). Furthermore, it is relevant to highlight that the specialties Farmacologia Clínica, Medicina Física e de Reabilitação, Medicina Intensiva, Medicina Legal, Medicina Tropical and Saúde Pública are not represented in dataset.

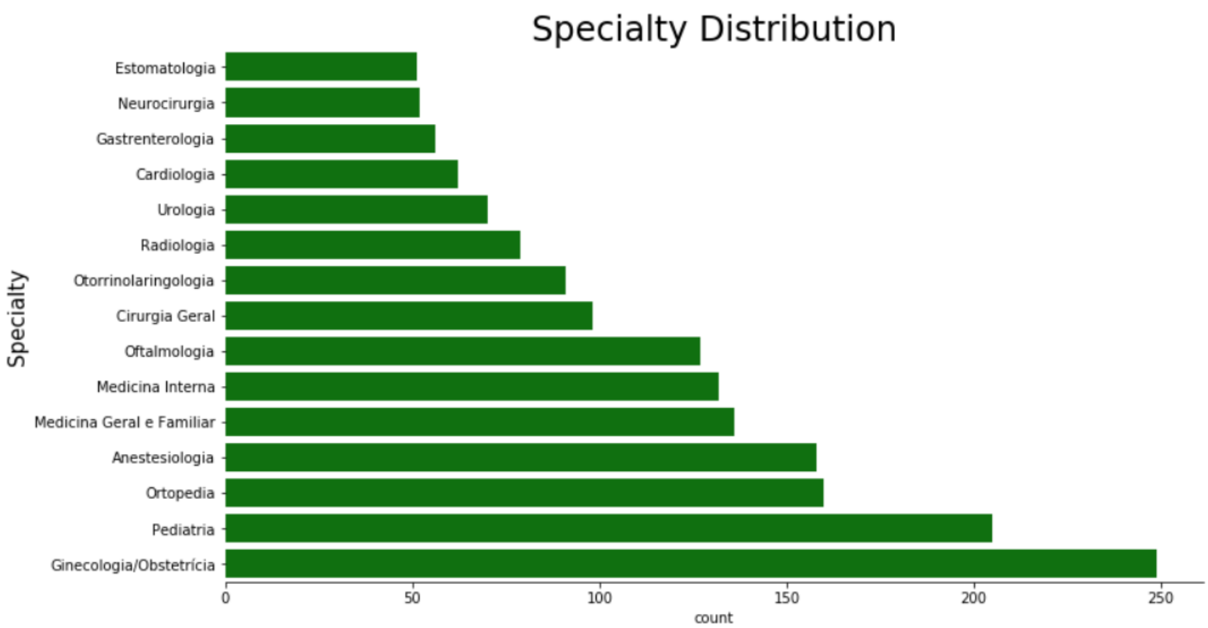


Figure 23 - Top 15 Specialty Distribution

4.1.2. Socio-Demographic Analysis

In what concerns the gender distribution, the sample of doctors is approximately evenly distributed, the male gender represents 51% while the female population constitutes the remaining 49% of the sample. By observing the female-male ratio per area of expertise, is interesting to verify that in general the areas related with childcare have a higher concentration of the female gender. While on the other hand, surgical related fields have a higher concentration of the male gender (figure 24 – Gender Distribution per Specialty). To observe the remaining specialties, observe Appendix 9.9 – Gender Distribution Per Specialty (Bottom 29). Moreover, the gender distribution per hospital is close to be evenly distributed (figure 25 – Gender Distribution Per Hospital). Therefore, when it comes to hire doctors there is no indication of exiting gender discrimination.

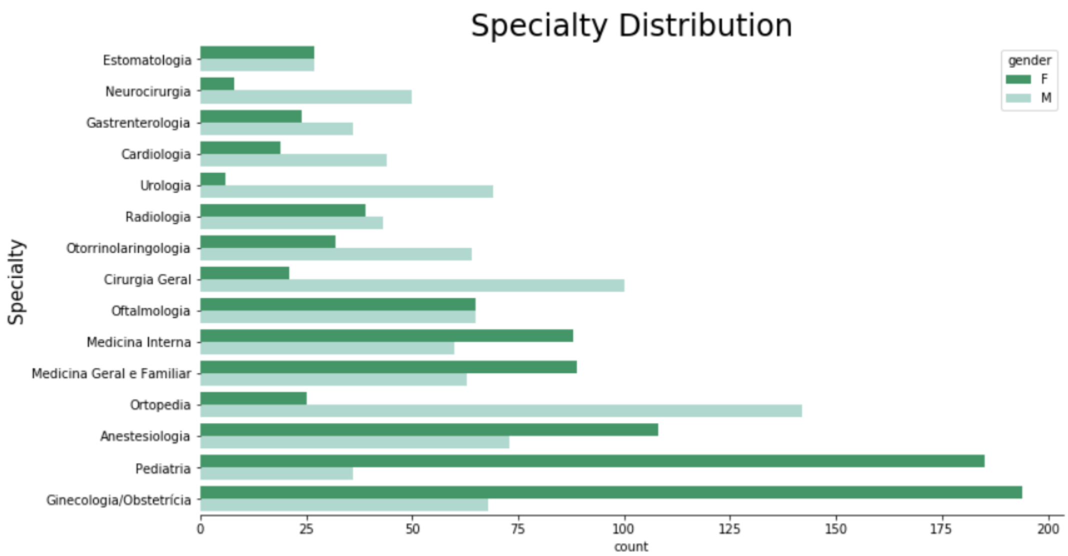


Figure 24 - Gender Distribution per Specialty

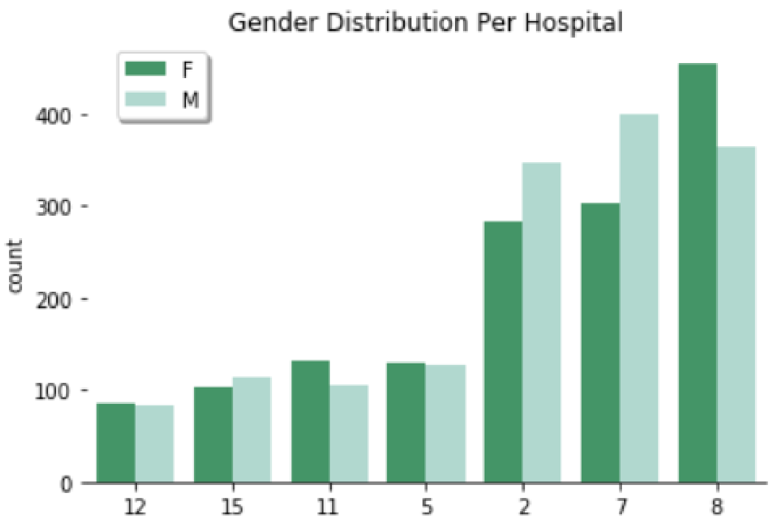


Figure 25 - Gender Distribution per Hospital

By applying the Empirical Cumulative Density Function to the distribution of the variable birth_year (figure 26 – Birth_Year Empirical Cumulative Density Function), is plausible to infer that the age structure of the sample is not old nor its young, it appears to be well balanced.

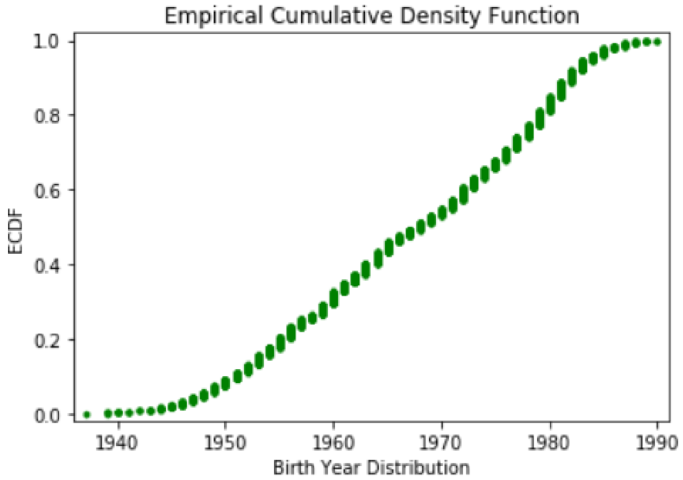


Figure 26 – Birth_Year Empirical Cumulative Density Function

In terms of years of practitice you might say that has a relatively young structure. In addition, when observing figure 27 – Years_of_practice Empirical Cumulative Density Function it is possible to conclude that approximately more than 85% of the sample has 15 years of experiencing or less by 2017.

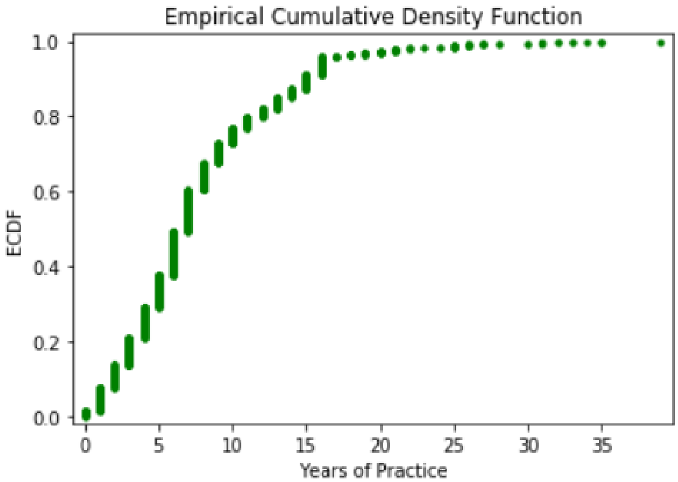


Figure 27 - Years_of_Practice Empirical Cumulative Density Function

4.2. PRIMARY-SPECIALTY

A large-scale patient consultation records was obtained from a private European healthcare provider between the years of 2012 and 2017, with 9,173,891 entries, 1.3 million unique patients, 2,171 unique doctors associated to 7 distinct hospitals.

As previously mentioned, a referral between two doctors is estimated to happen when a certain patient goes to see a family doctor and then no later than 30 days he or she meets a specialist for an appointment. The application of this methodology resulted in a weighted, undirected, explicit bipartite network in which 459 general physicians 1,487 specialists are connected. In total 1946 share at least a connection with another doctor. In addition, those doctors are linked by no more than 457,495 links which were converted into 59,065 unique edges.

The actual number of referrals made between each two doctors is measured through the weight associated to that same relationship, as that value represents the number of patients they share. In addition, given that only 1,946 of the 2,171 doctors are estimated to be connected this implies that 225 doctors[24], approximately 11% of the sample of physicians is expected to be isolated. Moreover, 89 doctors, 16 general practitioners, and 73 specialists are peripheral nodes in this primary-specialty referral network. This means that they only share a link with one doctor in the entire network. However, despite not being possible to ensure that this is actually the case, it still raises the possibility that those resources are not being optimized.

In addition, it was necessary to prove that the primary-specialty network does not exhibit characteristics of randomness, meaning that it does not have a similar structure of a random network. For that, it was tested if the weighted degree distribution of both groups of doctors followed or not a Poisson distribution. The results of both analyses returned a p-value of 0, allowing us to reject the hypothesis that both samples followed a Poisson distribution. Additionally, the average clustering coefficient was computed on both networks, on the Erdós-Renyi random network and on the primary-specialty. The average clustering coefficient computed on the doctor's network (0.0206364) revealed to be approximately 21 times higher than the one computed on the Erdós-Renyi network (0.0009752).

Furthermore, the human resources department of the healthcare provider was able to provide several important information's regarding doctor's background. It was possible to obtain their age, their level of medical expertise (number of years of practice), the specialty or specialties in which they are certified, their education, and the institution at which they did their post-graduation also known as "internato médico"

[24] In these 225 doctors that are estimated as not sharing connections, 14 of them belong to primary care and 211 to differentiated care.

4.2.1. Primary-Specialty Referral Patterns

Gender

In terms of gender distribution, the doctors belonging to the specialist group are more or less evenly distributed with a ratio of 44% compared to 56% male physicians. The scenario changes when considering the primary care doctors as approximately 69% are female and only 31% are male (figure 28 – Gender Distribution per Doctor Classification).

The distribution of the referrals per gender shows a preference for the male gender as 60% of the total of referrals was received by male specialists. From the total amount of referrals, the female gender has referred its own 26% of the times. But when referring the opposite gender this one represented 38% of the total referrals. In the scenario where the male gender is the one making referrals, women are referred 14% the times. The remaining 22% of the total referrals are made between men.

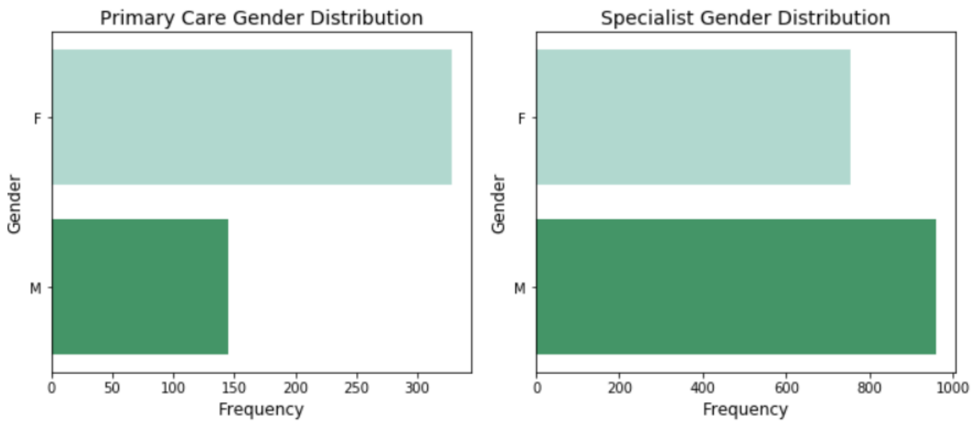


Figure 28 - Gender Distribution per Doctor Classification

In what concerns the 225 isolated nodes, on the general practitioner's side, the female-male ratio does not change, as it is 71%-29% respectively. Finally, the isolated specialists have a female-male ratio of 42%-58% respectively (figure 29 – Isolated Gender Distribution per Doctor Classification).

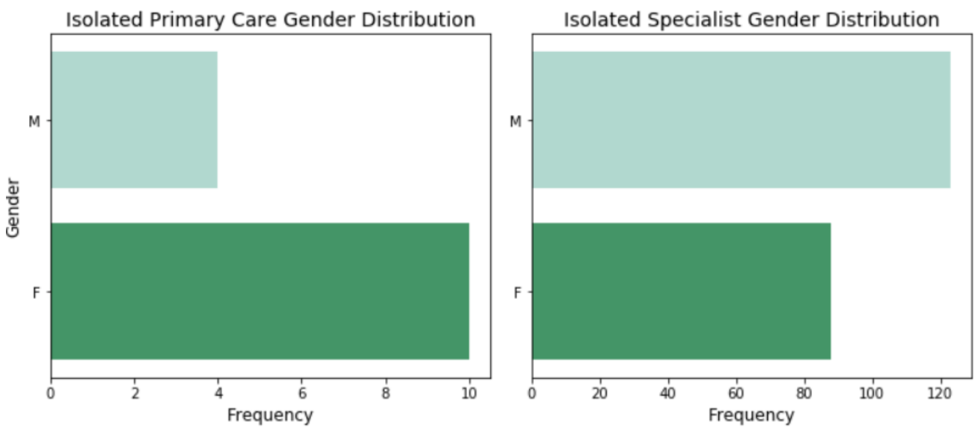


Figure 29 - Isolated Gender Distribution per Doctor Classification

Specialty

From the sample of doctors under analysis 44 distinct specialties are identified. As it turns out all of them are referred at least one time. Meaning they are all represented in the primary-specialty referral network. In the particular group of doctors representing the primary care, the specialty Medicina Geral e Familiar is the most active one (48%), while Medicina Interna (28%) and Pediatria (24%) are responsible for approximately the same number of referrals (figure 30 - Primary Care Referrals Distribution).

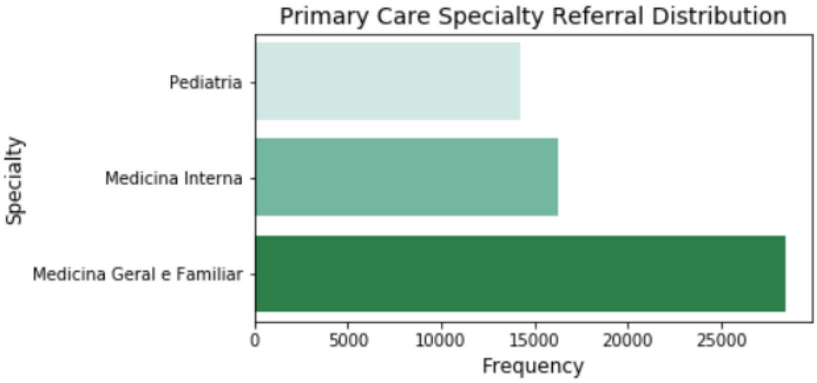


Figure 30 - Primary Care Referrals Distribution

According to the Portuguese national healthcare service the most common diseases that are currently affecting the Portuguese population are: Skin related diseases, depression, obesity related, and diseases such as low back and neck pain (*Retrato da Saúde 2018*, 2018). It is interesting to observe that most of the conditions aforementioned can be treated by at least one of the specialties in the of the most referred specialties (figure 31 – Differentiated Care Referrals Distribution). For example, the conditions of neck and low back can be treated by resorting to the specialties Ortopedia and Radiologia. Skin related conditions can be treated by the specialty Dermato-Venorologia. Morever, in terms of hearing and visual capabilities according to the portuguese national program “Para a Saúde da Visão ”, half of population in Portugal does not see properly (Público, 2005). Additionaly, according to a study called “Coping with noise ”, Portugal is the second country with worst hearing capabilities (Amplifon, 2017). These two conditions can be treated by the specialties Oftamologia and Otorrinolaringologia. Furthermore, the number of births in private healthcare as duplicated from 2000 to 2017, which is interesting given that it the fourthest most referred specialty, Ginecologia/Obstetrícia.

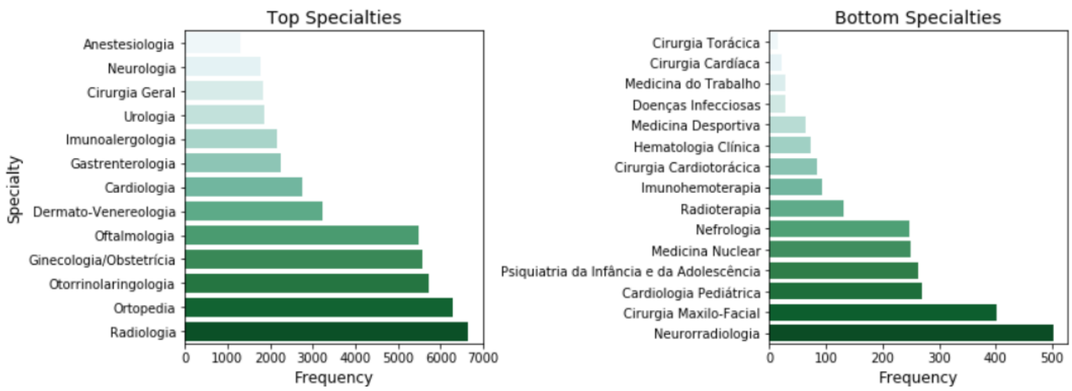


Figure 31 - Differentiated Care Referrals Distribution

In what concerns the isolated nodes the distribution of the specialties is interesting to see how the picture inverts (figure 32 – Isolated Primary Care Specialties Distribution). The specialty with the highest representation in the group of isolated general practitioners, Pediatria, is the one with the lowest number of referrals. Additionally, the specialty with the highest number of referrals made, has the lowest representation in this group of isolated doctors.

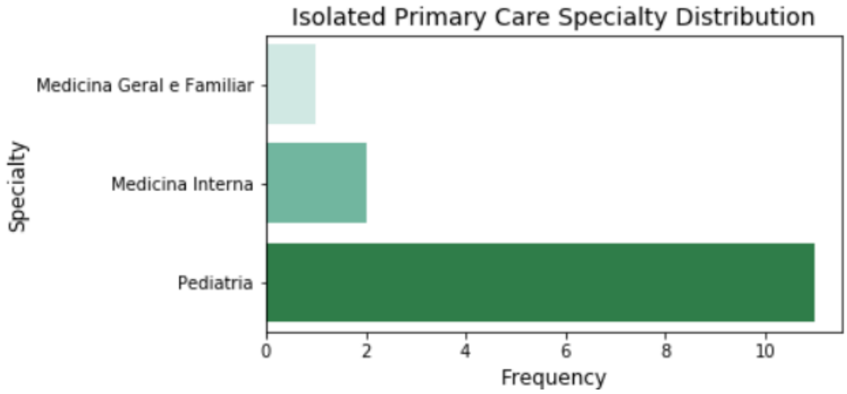


Figure 32 - Isolated Primary Care Specialties Distribution

The specialties represented in the group of doctors belonging to differentiated care (figure 33 – Isolated Differentiated Care Specialties Distribution), are all of them represent in the class "Top Specialties" of figure 31 – Differentiated Care Referrals Distribution, with exception of the specialties "Angiologia e Cirurgia Vascular", "Cirurgia Plástica Reconstrutiva e Estética", and "Neurocirurgia". Once more, this raises the question if the healthcare provider is being efficient in its decision making, it might not need that many professionals representing those specialties. However, it is not possible to ensure that this is the case in reality, therefore further analyses are required.

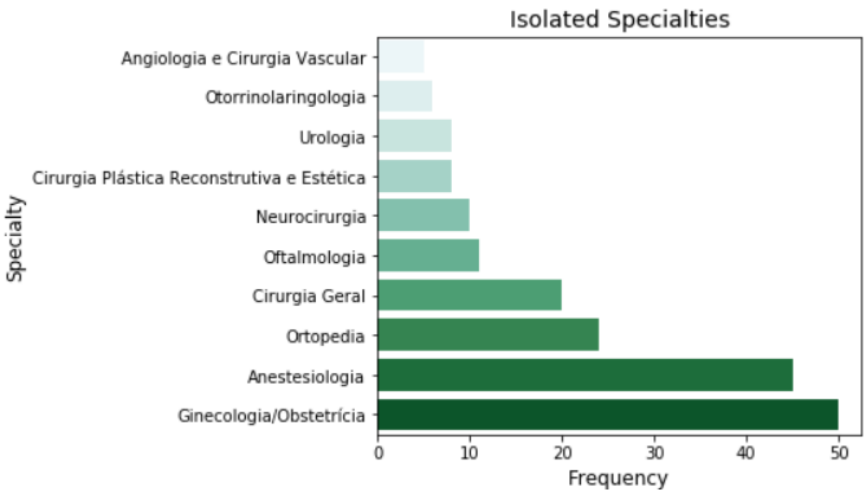


Figure 33 - Isolated Differentiated Care Specialties Distribution

Hospital

The distribution of doctors amongst the different hospitals can produce valuable insights as it is important to understand if the referrals between the two groups of professionals are being kept inside each healthcare institutions not. In the primary-specialty referral network approximately 76% of the referrals estimated only have one hospital in common. Therefore, it is plausible to assume that the vast majority of the referrals is kept inside each of the healthcare institutions. However, it is only possible to ensure that 37% of the distinct belong to the same hospital. Additionally, the data shows that in approximately 19% of the mapped relationships, physicians did not have hospitals in common (figure 34 - Number of Hospital In Common).

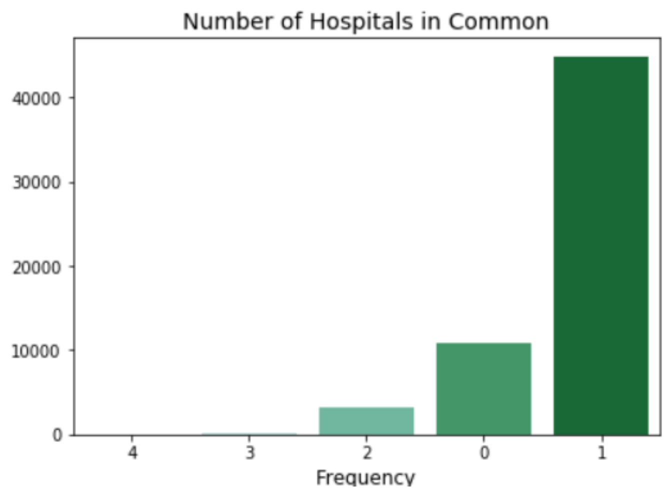


Figure 34 - Number of Hospitals In Common

Age

In terms of age difference, it is possible to conclude that close to 80% of the referrals found amongst the two type share an age difference inferior or equal to 20. Additionally, more than 40% of them do not have more than 10 years of difference. This might suggest a certain level of homophily. From figure 35 – Age Difference Distribution ECDF it appears that higher the age difference the less likely is that a specialist will be referred.

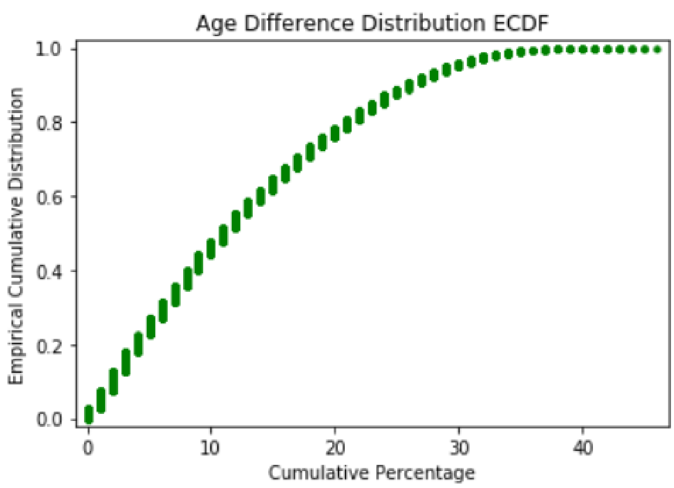


Figure 35 - Age Difference Distribution ECDF

Perceived Clinical Expertise

By observing the distribution of the specialists being referred in figure 36 – Specialists Perceived Clinical Expertise, the number of specialists in with more than 20 years of medical practice are practically none, in fact they represent less than 10% of the doctor's being referred. In addition, nearly 80% of them does not have more that 15-years practicing medicine. Given that more or less 85% of the entire sample of doctors does not have more than 15-years practising their profession (figure 27 - Years_of_practice Empirical Cumulative Density Function) it is hard to state a pattern in terms of referring specialists due to their level of perceived experience.

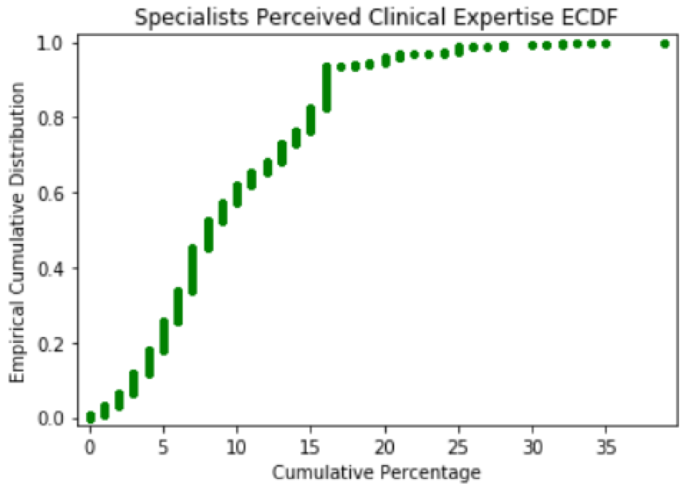


Figure 36 - Specialists Perceived Clinical Expertise

In addition, it was not associated any justification that would further help comprehend the situation of the isolated nodes in the sample (figure 37 – Isolated Age & Perceived Clinical Expertise ECDF). However, there are two individuals that distiante themselves from the overall sample of isolated doctors in terms of age, and one in terms of years of practice, which might be a contributive factor for being left out from the primary-specialty referral network.

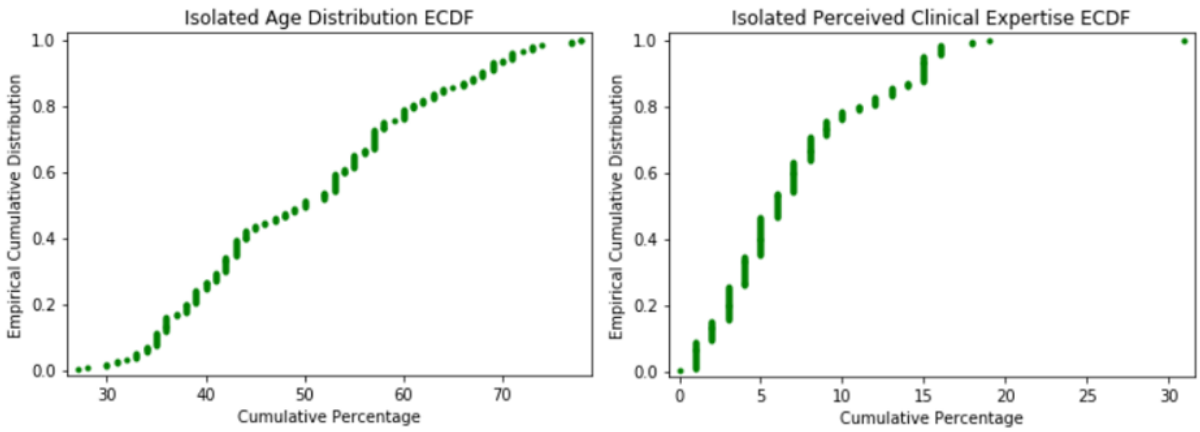


Figure 37 - Isolated Age & Perceived Clinical Expertise ECDF

4.2.2. Primary-Specialty Centrality Measures

Three distinct social network centrality measures were computed after instantiating the primary-specialty referral network. Those measures were: Degree centrality, Betweenness Centrality and Closeness Centrality. In terms of the centrality measure degree (figure 38 – Degree Centrality Distribution), it is possible to observe that both samples have a skewed distribution to the right. Furthermore, as the number of referrals increases the number of primary care doctors representing them decreases. A similar situation happens in the specialist distribution. However, the situation is not so much intense, meaning this distribution has a smaller tail. In addition, it was tested the possibility that the weighted degree distribution of both groups could be represented by power-law distribution (null-hypothesis). In the case of general practitioners this would mean that in a certain distribution only a few nodes are responsible for a large proportion of the referrals being made. For specialist doctors it would mean that only a few doctors are receiving a significant percentage of the referrals being made. It was only possible to prove that the primary care physicians group follows a power-law distribution, as the null-hypothesis cannot be rejected, given its p-value of 0.23.

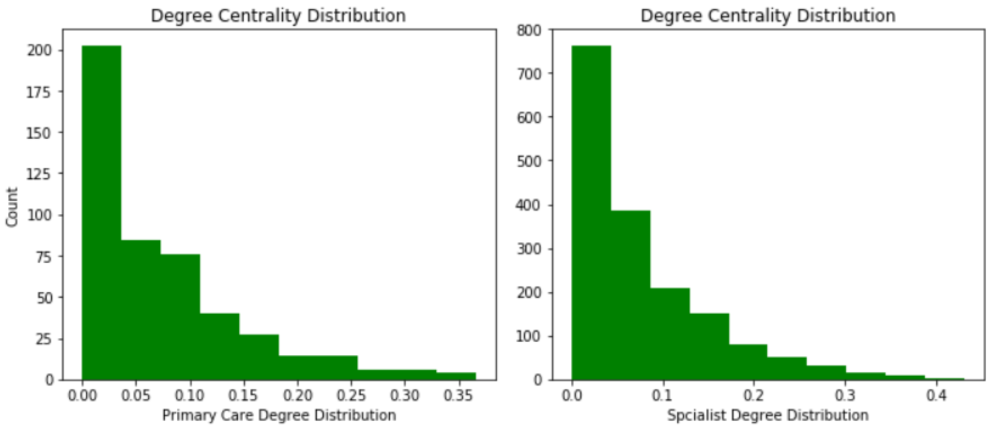


Figure 38 - Degree Centrality Distribution

Furthermore, in terms of the betweenness (figure 39 – Betweenness Centrality Distribution) there is a significant number of individuals on both distributions with values of zero, which means that they aren’t on the shortest path of any pair doctors. In addition, the values registered in this metric are extremely low. In fact, the highest values detected in the specialists and general practitioners are 0.007 and 0.035 respectively. It is also, interesting to notice that in general, family doctors have higher values in this metric when compared to specialist. For example, whereas 63% of the general practitioners have values superior to 0.0001, specialists only have 36% of their doctors represented.

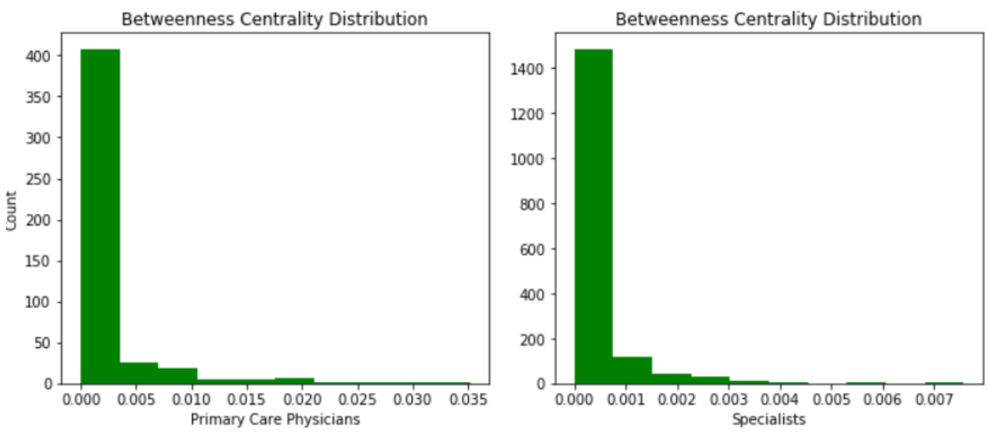


Figure 39 - Betweenness Centrality Distribution

In what concerns the metric closeness the distribution of both type of doctors is significantly different from the other centrality measures (figure 40 – Closeness Centrality Distribution[25]). The only doctors with closeness values equal to zero are the isolated nodes. The higher the values of closeness the more central nodes are in the network. It is interesting to see how the picture has inverted when comparing with the metric degree and betweenness centrality. According to the distributions of the metric closeness, specialists have a more central role in the network, meaning that in general they have more access to the rest of the doctors in the network. This might be happening, because only a few primary care doctors are responsible for most of the referrals being made, and thus, those doctors might be referring the same specialists, to some extension.

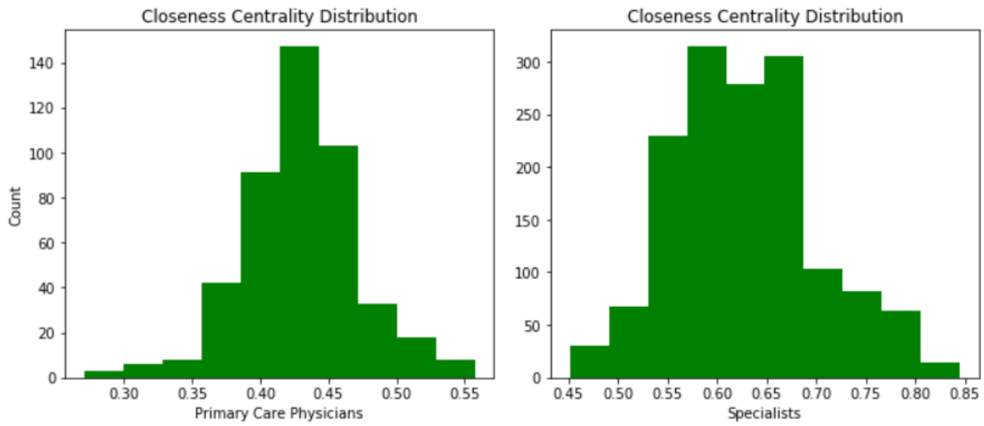


Figure 40 - Closeness Centrality Distribution

It has also been quantified the number of referrals made and received by the individuals that represents top and bottom 25% highest values for all the three centrality metrics (degree, betweenness, and closeness), regarding the two type of doctors. In case of the metric degree, the top 25% individuals are held responsible for 78.9% of the referrals, while the bottom 25% made only 0.33% of the referrals. The scenario, similar for the specialist doctors as the 25% individuals for having more direct ties received 77.8% of the referrals, while the bottom 25% received only 0.08% of the referrals. Moreover, the situation does not change for the remaining centrality measures. Betweenness centrality shows a 25% top-bottom ratio for primary care doctors of 77.8% - 0.34% and a top-bottom ratio for specialists of 73.3% - 0.1%. The metric closeness reveals a top-bottom 25% ratio for general practitioners of 79.2% - 1.4% and a top-bottom 25% ratio of 68.6% - 2.62%. Once again, the issue that the primary-specialty referral network is not being efficient is raised as the referral process appears to be ensured by only a few doctors, both on the making and receiving side of the referrals.

[25] This visualization does not account for the isolated nodes.

4.2.3. Primary-Specialty Referral Network Community Analysis

Community Structure

Communities in network science represent groups of nodes, which in this case are doctors who have a denser network amongst them. Additionally, the connections between groups are expected to be less condensed. They are important structures to be analysed, as they help to understand the dynamics of complex networks. The algorithm used for the detection of communities in this dissertation was the Louvain, a modularity-based optimization algorithm, which considered the weights of each relationship.

In total, the algorithm estimated the existence of 7 distinct communities. This is interesting as 7, corresponds to the number of unique hospitals that are in our sample (figure 41 – Primary-Specialty Communities[26]). In addition, it is possible to observe the communities individually in Appendix 9.10 – Individual Communities. It is also possible to observe the communities' structure in Appendix 9.11 – Communities Structure.

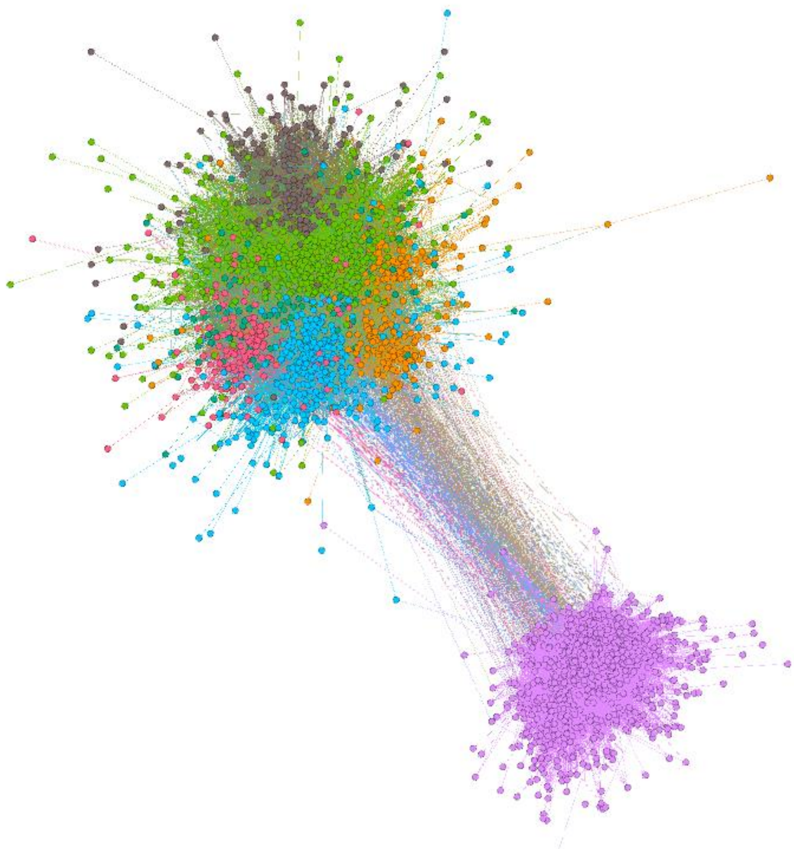


Figure 41 - Primary-Specialty Communities

From community 1 to community 7 the respective sizes of each of them is: 634, 80, 191, 198, 384, 322, 137. It is interesting to observe that in each community, there is one hospital that stands out in terms of representation. This might suggest that each community is representing a unique hospital of

[26] This visualization does not contain the isolated nodes.

the healthcare provider. In the case of community 3, the hospital with the highest representation is 8, but the concentration of individuals working in hospital 8 is significantly higher in community 5. Therefore, group 5 is associated with hospital 8. Thus community 3 will be associated with hospital 15 (Appendix 10 – Communities Structure). Although it is important to bear in mind that the particular situations of communities 3 and 4 are not so clear, when having to assign a hospital to them. It assumed that community 4 represents hospital 12.

Furthermore, the gender distribution amongst all communities is similar and more or less evenly distributed. Only community in community 3 it is verified a big disparity between the two genders as the female-male ratio for that group is 74%-26%. It also the only community where the number general practitioners and specialist nearly the same, which might be a contributing factor, as the female population is significantly higher in doctor certified in primary care.

In addition, the top 5 specialities with the highest representation in all communities produces a list of 13 unique specialties in which, Ginecologia/Obstetrícia, Ortopedia are represented in five communities and Oftamologia, Medicina Geral e Familiar and Anestesiologia are represented amongst 4 different communities.

In what concerns the distribution of the ages of doctors along the distinct communities this shows a very low variance for the metrics calculated (minimum age, average of age, and maximum age). Meaning that all groups estimated by the Louvain algorithm present a similar age structure. Moreover, in terms of perceived clinical expertise the situation is equivalent with exception of the maximum clinical expertise registered in the different communities. For example, in community 2 the doctor with most experience has 16 years on the job, while the most experienced doctor in community 6 has 39 years of medical practice. Even so, on average all communities reveal similar values.

Communities Referrals Distribution

In what concerns the distribution of the referrals amongst all the communities, the communities that were more active during the years of 2012 and 2017 were communities 4, 5, and 6 as they are the ones who registered the highest number of referrals per general practitioner. On the other hand, the ones with less activity were communities 1, 2, and 3. In addition, is also possible to observe in table 12 that the majority of the referrals is kept inside each community. In fact, only group 2 has the opposite situation. It is also interesting to observe that the only community that is distanced from all the other in geographical terms (Community 1) is the most isolated one. Meaning it is the one with the highest percentage of referrals made inside its community and the one with less referrals made to outside or received from other communities. This can be explained by the fact that it is the only hospital that operates in a different region of the country. Furthermore, from the 459 connected doctors approximately 20% refer exclusively doctors from their own communities. Moreover, in terms of specialists, nearly 31% are being exclusive to the communities in which they are inserted in.

	Community 1	Community 2	Community 3	Community 4	Community 5	Community 6	Community 7	Total
# of Referrals Made	101,423	9,355	47,954	49,891	131,035	94,827	23,010	457,495
% of Referrals Made	22.17%	2.04%	10.48%	10.91%	28.64%	20.73%	5.03%	100%
% of Referrals Inside	99.66%	41.62%	72.34%	73.07%	73.69%	69.78%	61.56%	-
% of Referrals Outside	0.34%	58.38%	27.66%	26.93%	26.31%	30.22%	38.44%	-
# of PCP	155	13	95	30	86	57	23	459
# of Referrals per PCP	654	720	505	1,663	1,524	1,664	1,000	997
# of PCP that refer inside	155	13	95	30	86	57	23	459
# of PCP that refer outside	79	13	91	29	82	52	23	369
# of Referrals Received	101,274	10,894	53,018	47,131	135,207	88,781	21,190	457,495
% of Referrals Received	22.14%	2.38%	11.59%	10.30%	29.55%	19.41%	4.63%	100%
% of Referrals Inside	0.19%	64.26%	34.57%	22.66%	28.59%	25.47%	33.15%	-
# of SP	479	67	96	168	298	265	114	1,487
# of Referrals per SP	211	163	552	281	454	335	186	308
# of SP that are referred inside	479	67	96	168	298	265	114	1,487
# of SP that are referred outside	118	63	90	155	264	238	108	1,036

Table 12 - Communities Referrals Distribution

Doctors with More Activity Outside their Community

In this sub-section it was identified the top 25% primary-care doctors of each community with more referrals made to outside their own. Regarding the group of specialists belonging to differentiated care it was identified the 25% doctors that received more referrals from outside their community. In terms of gender distribution per type of doctor classification, it is not strange that the number of female individuals higher in primary care as the female-male ratio in the sample of doctors is 69%-31% respectively. The same cannot be said for the doctor in differentiated care as the male gender in that group is significantly higher. Thus, there is a higher propensity towards the male gender to be referred by communities outside their own (figure 42 – Doctors with More Activity Outside their Community Gender Distribution).

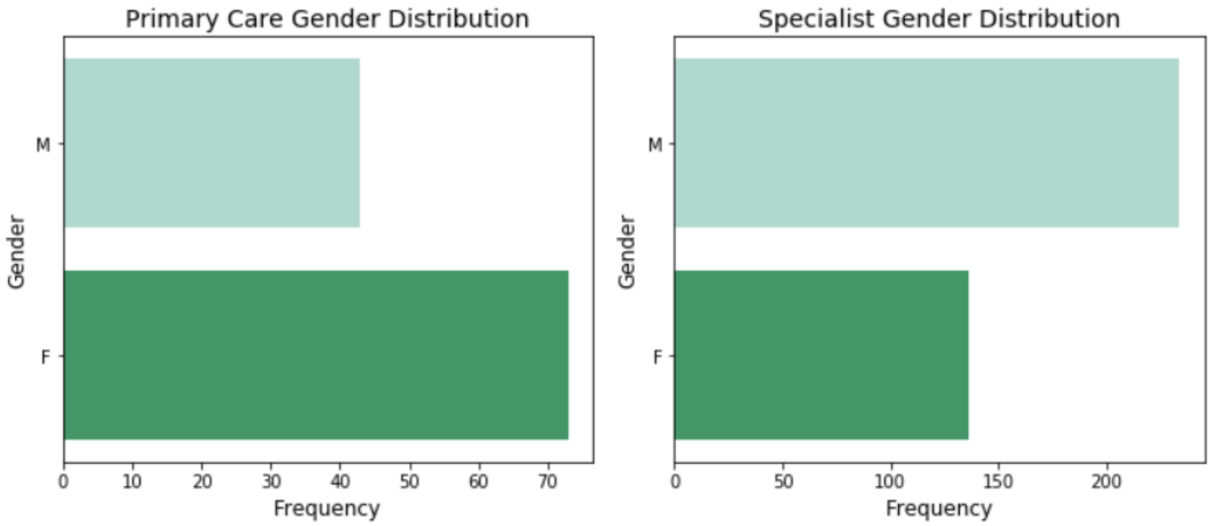


Figure 42 - Doctors with More Activity Outside their Community Gender Distribution

In terms of specialty activeness, in primary care (figure 43 – Doctors with More Activity Outside their Community Primary Care Distribution) Medicina Geral e Familiar is the most active one while the least active is Medicina Interna. Considering that the specialty that has given origin to more referrals was Medicina Geral e Familiar, it is not strange that also being the one making more referrals outside (figure 30 – Primary Care Referrals Distribution). However, Medicina Interna despite making more referrals than Pediatria, it ends up making less referrals to specialists outside their communities.

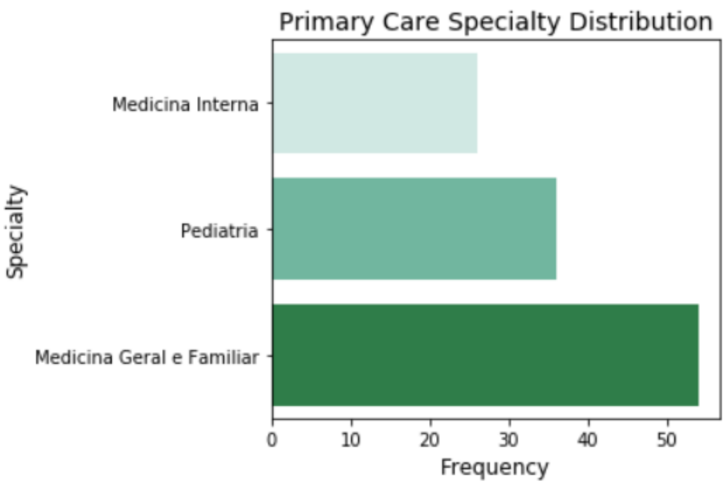


Figure 43 - Doctors with More Activity Outside their Community Primary Care Distribution

In what concerns the most active specialties, the framework is extremely similar to the one in figure 31 – Differentiated Care Referrals Distribution. In fact, the only differences are that “Pneumologia” is not in it, and the specialties “Oftamologia”, “Ginecologia/Obestetrícia”, “Urologia” and “Imunoalergologia” do not appear in the same order.

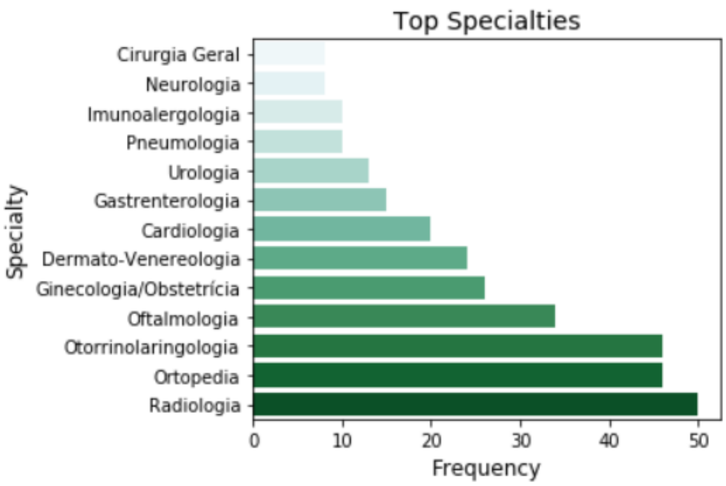


Figure 44- Doctors with More Activity Outside their Community Specialists Distribution

The age distribution of the doctors that are more likely to make ror receive referrals from outside their communities is not associated with any pattern in terms of age or perceived clinical expertise, as both distributions appear to be balanced. There are young doctors making referrals to outside their community as there are young physicians receiving referrals from outside and vice-versa (figure 45 – Doctors with More Activity Outside their Community Age Distribution).

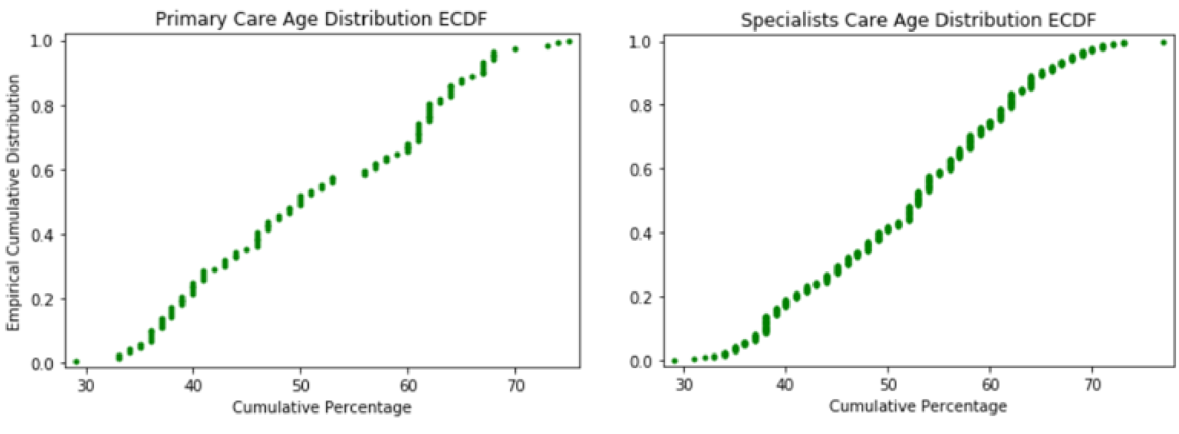


Figure 45 - Doctors with More Activity Outside their Community Age Distribution

In what concerns the perceived clinical expertise of doctors in general it shows a rather younger structure as more than 80% of the top 25% general practitioners that made more referrals to outside their respective communities as 15 or less years of medical practice. The situation is similar in the specialists case as only approximately 19% has more than 15 years of medical experience (figure 46 – Doctors with More Activity Outside their Community Perceived Clinical Expertise Distribution).

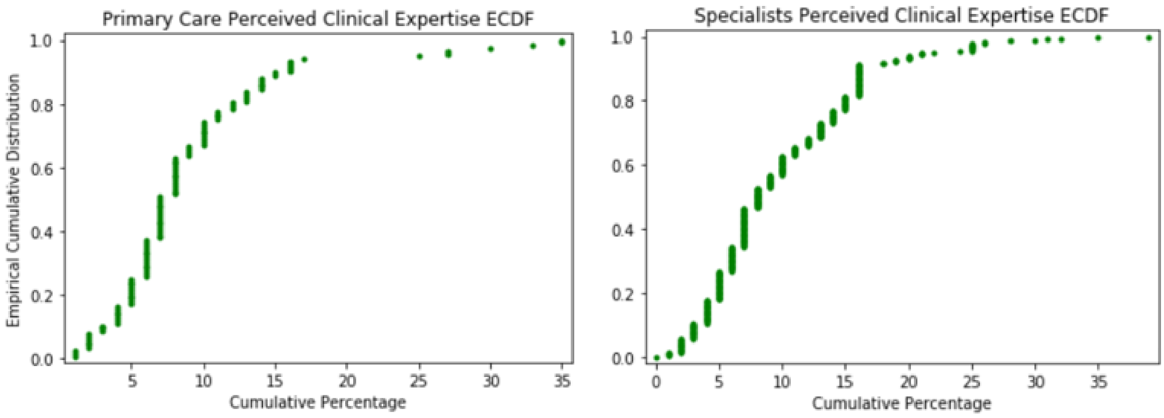


Figure 46 -Doctors with More Activity Outside their Community Perceived Clinical Expertise Distribution

When analysing the centrality measures is possible to observe that this doctors correspond not only to the ones who are more active outside their communities as they also belong to the group of more active doctors in general. For example by observing figure 47 – Overall and Most Active Doctors Outside their community Degree Centrality Distribution[27], it possible to see that some of the doctors hold the highest values for metric degree centrality.

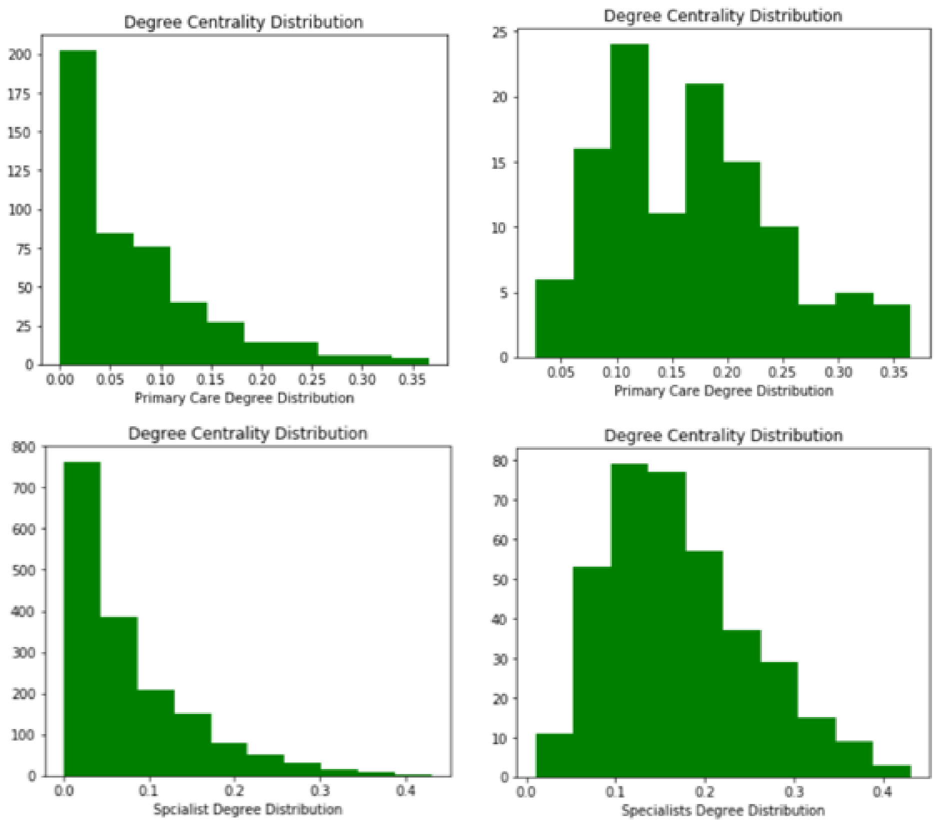


Figure 47 - Doctors with More Activity Outside their Community Degree Centrality Distribution

The same situation is verified for the centrality measures betweenness and closeness (Appendix 9.12 – Doctors with More Activity Outside their Community Betweenness Centrality Distribution and Appendix 9.13 – Doctors with More Activity Outside their Community Closeness Centrality Distribution).

27 The column of graphs on the left represents the overall distribution of the metric degree centrality. The column of graphs on the right represents

4.2.4. Primary-Specialty Regression Analysis

For the development of this chapter two multiple linear regressions were built, one for each type of doctors, general practitioners and specialists. In both regressions it was included all possible variables regarding the doctor's background and social network centrality measures. In addition, the variable referral_rate (dependent variable) was computed with objective of understanding which variables had an impact in it or not.

In the case of the primary care doctor's regression (figure 48 – Primary Care Regression Analysis and Appendix 9.6 – Primary Care Doctors Independent Variables), the first variable listed that has having an impact in the referral rate of doctors is the variable "age". Oddly enough, its impact in the dependent variable, regardless of how small it is, its negative. Meaning the older you, the less likely a doctor is to be referred. In what concerns the centrality measures of network, the only one that revealed has having an impact in the referral rate was the metric closeness. In this case this means that, the more central a general practitioner is in the network the higher is the likelihood of him or her to make referrals. Furthermore, the female gender is expected to have a weight in the referrals of the family doctors. This is not a surprising result given the unbalanced gender distribution in primary care doctors favouring women. Moreover, all the primary care physicians have an impact in the referral rate. In fact, the specialty "Medicina Interna" is expected to be the variable with the highest impact (coefficient wise). This is not so much intuitive as Medicina Geral e Familiar is more active in terms of making referrals. A unit increase in the age of a doctor is expected decrease the referral rate by 0.0013. If a doctor registers the maximum value in terms of closeness centrality this will translate in an increment in the referral rate of 0.2691. Additionally, having only one specialty is expected to increase the referral rate of a physician of 0.0930.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          referral_rate   R-squared:                       0.824
Model:                            OLS   Adj. R-squared:                  0.820
Method:                 Least Squares   F-statistic:                     196.7
Date:                Sat, 02 Nov 2019   Prob (F-statistic):         1.70e-166
Time:                        11:42:39   Log-Likelihood:                 334.03
No. Observations:                 473   AIC:                            -646.1
Df Residuals:                     462   BIC:                            -600.3
Df Model:                          11
Covariance Type:            nonrobust
===================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------
age                                -0.0013      0.000     -2.603      0.010      -0.002      -0.000
years_of_practice                   0.0014      0.001      1.050      0.294      -0.001       0.004
closeness                           0.2691      0.076      3.525      0.000       0.119       0.419
degree                             -0.0226      0.109     -0.207      0.836      -0.237       0.191
gender_F                            0.0394      0.013      3.100      0.002       0.014       0.064
number_of_hospitals_1.0            -0.0815      0.045     -1.798      0.073      -0.170       0.008
number_of_hospitals_2.0            -0.0988      0.048     -2.069      0.039      -0.193      -0.005
number_of_hospitals_3.0            -0.0674      0.054     -1.258      0.209      -0.173       0.038
number_specialties_1.0              0.0930      0.029      3.156      0.002       0.035       0.151
specialty_Medicina Geral e Familiar 0.1951      0.016     12.231      0.000       0.164       0.226
specialty_Medicina Interna          0.2824      0.015     19.388      0.000       0.254       0.311
==============================================================================
Omnibus:                      299.043   Durbin-Watson:                   2.070
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3797.419
Skew:                           2.540   Prob(JB):                         0.00
Kurtosis:                      15.918   Cond. No.                     1.13e+03
==============================================================================
```

Figure 48 - Primary Care Regression Analysis

In what concerns the regression of the specialists the only three variables that were concluded as having an impact were the characteristics age, closeness centrality, and the specialty Pediátrica. In this regression the variable age also has a negative impact in the referral rate of a doctor. (figure 49 – Specialists Regression Analysis and Appendix 9.7 – Specialists Independent Variables). A unit increase in the age of a doctor decreases the referral rate by 0.0019. Additionally, if specialist has a closeness centrality value equivalent to 1, it will represent an increase in the referral rate of 0.1264. It is interesting to notice that this variable has a bigger impact in the referral rate of a general practitioner. Moreover, being a specialist in Cirurgia Pediátrica contributes to an increase in the dependent variable of 0.2481. It is the variable with more impact in the referral rate (Appendix 9.14 – Specialist Regression Analysis).

5. CONCLUSIONS

The development of this dissertation is expected to add more insights and new methodologies to the existent literature concern about the referral mechanism between general practitioners and specialists. From a large-scale patient consultation dataset filtered between the years of 2012 and 2017, provided, by a private European healthcare institution, it was obtained a sample with more than 9 million entries, 1,305,361 unique patients, 2171 unique physicians associated to 7 different hospitals. From the overall sample of physicians in the sample 22% are general practitioners and 78% are specialists

The output of the estimated primary-specialty referral network accounted for all the 2171 doctors but only 1946 are expected to share at least one connection with another doctor from these, 459 are primary care doctors and 1487 belong to differentiated care. The remaining 225 of the sample represent isolated doctors, 14 are family doctors and the other 211 doctors are specialists. For these doctors it was not found a doctor with whom they might share a connection. Additionally, within the referral network there are 89 doctors who are identified as having unique relationships with whom they only have one patient in common. Therefore, without making any conclusions that group of doctors may represent underutilized resources for the healthcare provider. Thus, it would be in its best interest to make further analysis to settle if such statements are factual or not. In total the primary-specialty network is comprised of 59,065 distinct relationships that embody 457,495 referrals. The weight of each unique link characterizes the actual number of patients that two doctors share.

In addition, besides the doctor appointments records, it was provided the information regarding doctors' background by the Human Resources department. Such information as doctors' age, years of practice, hospital in which they are currently working in, and the higher learning institution where they got their university degree were given.

The gender composition for primary care doctors is 69% of female and 31% male, and 44% of female and 56% of male for specialists. Women show to be more likely in pursuing a career in the primary care than men. Additionally, it is also interesting to notice that in general the percentage of women in specialties related with childcare is higher. Whereas men, show higher concentration in areas related with surgical fields. In fact, in all the 6 surgical related fields recognized as a specialty in Portugal, including pediatric surgery women are outnumbered. In terms of the 4 specialties including only childcare only, women have a stronger presence in 3.

From the total referrals estimated, approximately 60% of them were to the male. In addition, the distribution of the references per gender shows higher propensity for women to refer men and not the other way around. From the total of referrals that are originated from women (63.9%), 38.3% of the total referrals made are to men. However, from the total of referrals made by men (36.1%) only 14% of them were to women. Additionally, it is detected a certain level of homophily[28] in the referrals established. Approximately 48% of the total referrals to the same gender. Moreover, close

[28] Homophily is a phenomenon that states that individuals have higher tendency to be connected with individuals with similar characteristics.

80% of the references made are made between doctors with an age difference smaller or equal to 20 and more than 40% were between individuals with an age difference smaller or equal to 10. More interestingly, the majority of the referrals made appear to be kept inside each of the 7 hospitals, as 76% of the total referrals were made amongst doctors that only have 1 hospital in common. Additionally, approximately 18% of the referrals do not have a single hospital in common. However, only close to 37% of the referrals made between two doctors can be ensured that they represent a unique hospital.

Furthermore, from all the referrals made to specialists more than 60% were made to doctors that claim to have no more than 10 years of medical experienced. However, given that approximately only 4.37% of the doctors who are connected specialists have more than 10 years of experience it might be plausible to infer but not conclude that experience of a doctor is relevant for the referral decision process as 40% of the total referrals made were to doctors with more 10 years of experience.

In what concerns the specific metrics of network science three metrics were computed based on the primary-specialty referral network: Degree Centrality, Betweenness Centrality and Closeness Centrality. In the case of general practitioners, it was possible to prove that only a handful of the primary care doctors is responsible for a significant percentage of the referrals mapped throughout the network. In fact, when selecting the 25% individuals with highest values of degree centrality they are responsible for approximately 79% of all the estimated referrals. Moreover, when quantifying the number of referrals made by the 25% with the highest values in terms of betweenness and closeness centrality they are responsible for 78% and 79% of all the referrals respectively. Therefore, a high percentage of doctors is represented on top 25% of the three centrality measures. In terms of the specialists, despite not having been proven that only small group of the doctors belonging to differentiated care, by selecting the top 25% of each network centrality measure, degree, betweenness and closeness centrality, those doctors represent 78%, 73% and 69% of the total referrals respectively. Therefore, according to these results the primary-specialty referrals network, once more, appears to not be efficient. This suggest that only a few primary care doctors and specialist are being responsible for the referral system.

In terms of communities, through the application of the Louvain algorithm, 7 distinct communities were identified. In fact, 7 is the number of hospitals present in the sample. In what concerns the structure of the communities they show to be similar in age, perceived clinical expertise, specialty distribution, gender and doctor classification. However, community 3 represents the exception as it verified an unbalanced gender distribution with a female-male ratio of 74%-26%. Moreover, in case the different communities are actually representing each unique hospital in the sample, it is possible to support once more that the majority of the referrals is kept inside each community. Considering that on average approximately 30% of the referrals made by a community is to outside, community 2 (58%) and 7 (38%) stand out. Moreover, it is interesting to notice that the only hospital that operates in a different region from the other six hospitals is the most isolated one as only 0.34% of the referral with origin in that community were not kept inside it. Even so, from the 101274 referrals they receive 22.14% were made from other communities. This might indicate that geographic distance between hospitals plays an important role in the referral decision process. Moreover, is common for doctors to guarantee that the referrals sources remain happy with them specially major cities and in particular metropolitan areas (Jauhar, 2019).

It is also interesting to noticed that in the top 25% of doctors that make more referrals outside their community and the doctors that receive more referrals from outside their community, is partially represented by the doctors that have the most relevant positions in terms of degree, betweenness and closeness centrality. This might suggest that the number of direct ties a doctor has in the network, the number of times it appears on the shortest paths of other nodes and the position it holds in the network might have an impact in the referral rate of a doctor. Moreover, the statement that the male gender in the specialists has represents a higher percentage of the referrals is evidenced once more.

Finally, by performing a regression analysis on both types of doctors' classification it was possible to conclude within our sample if the background and network metrics have or do not have an impact in the referrals rates associated with each doctor. In the case of primary care from the 11 variables included in that regression 7 were determined as having a statistical impact in the referral rate of general practitioner. The results that the age, the metric closeness, working in 1 or 2 hospitals and being certified in primary care influences the referral rate of doctor. It is interesting the fact one of the metrics that can the biggest impact in a referral rate of a doctor is the centrality metric closeness which depending on its values, it signals the position of the doctor in the network, how much access it has to other doctors. However, the one with the most significant impact is being certified in "Medicina Interna".

In the case of the specialties the results were not so promising. However, the concluded within our sample that the age and closeness centrality have impact in the referral rate of doctor belonging to differentiated care. In addition, from the 40 specialties taken into consideration in the specialists' group, only pediatric surgery revealed to have an impact (positive) in the referral rate of a doctor.

In both regression it was not found evidence that the number of connections a doctor has, or the number of times it appears on the shortest paths of other two doctors, to have any influence in the referral rate of a doctor.

6. RECOMMENDATIONS AND POTENTIAL APPLICATIONS

According to Deloitte: Global health care outlook Shaping the future, 2019, the anticipated challenges that the healthcare industry is expected to face are: ***“Creating financial sustainability in an uncertain health economy, adapting to changing consumer needs, demands, and expectations, using new care delivery models to improve access and affordability , maintaining regulatory compliance and cybersecurity, investing in innovation and transformation, and last but not least recruiting, developing and retaining top talent”.***

The referral mechanism of a healthcare provider can have an impact on all. Inefficient communication within a referral network can have a negative impact for the referral network, as it can result in ineffective care coordination by the healthcare provider (An, O’Malley, Rockmore, et al., 2018). The level of effectiveness of communication between members of a network is affected by how information and resources flow around the network (An, O’Malley, & Rockmore, 2018). In addition, given that good patient experience is related with higher hospital profitability (Deloitte, 2019), healthcare providers are once more affected by the referral system between the two types of doctors. The recommendations and application mentioned in this chapter focus only their impact on the primary-specialty referral network and consequently on the healthcare provider.

6.1. GENDER BIAS

In terms of the appropriateness, a referral should only happen if justified and if it is in the best interest of a patient and not of the doctor (Schroeder, 2016). Referrals of patients having a condition that has a low degree of severity by Primary Care Doctors to Specialists, is considered by physicians themselves a waste of “expert” consultations, that ends up adding unnecessarily costs, both to the healthcare provider as to the patient(Jauhar, 2019).

Therefore, considering that network as shown signs of inefficiency it is important to tackle those same problems. Thus, to start it would be prudent to develop campaigns that would empower women in differentiated care and create awareness for the trend that has been created where the male gender has been predominant in the referrals of specialist. The objective of such measure is creating a more balance distribution of referrals per gender.

6.2. PERCEIVED CLINICAL EXPERTISE DISTRIBUTION

Moreover, knowing that the structure of the healthcare provider in terms of its work force perceived clinical expertise is relatively young, it would be wise to develop policies that would increase the retention rate of doctors, in order to contribute to more even distribution. However, it should be important to keep the correct balance. In the same way that is important to have people with experience and knowledge it is also important to have young minds that can contribute to the innovation and transformation of current processes and tools being used daily by a doctor, that can help them perform a better job, by providing better quality care and attend emerging needs that their patients may have. Consequently, it is important to create programs and policies that ensure that that dynamics of innovation and transformation endures.

6.3. PRIMARY-SPECIALTY UNDERUTILIZED RESOURCES

Furthermore, having knowledge that only a hand of primary care doctors and specialists are being responsible for the referral's mechanism of the healthcare provider. This might mean that the doctors represented in this group are facing a significative overload of work which can lead to poorer quality care and customers satisfied with the healthcare provider. Thus, if the administration has doctors that are being underutilized, it can be helpful to distribute the amount of work by those resources, specifically amongst general practitioners. To solve the particular situation of the specialist it would be exciting to develop a recommender system. The function of such program would be to optimize the allocation of patients to general practitioners and the referrals being made to the different specialists in the network. Hopefully, this tool would contribute, for doctors to have on average a more central role in the primary-specialty referral network, making their closeness centrality values to go up. For the particular case of the allocation of patients to family doctors, a collaborative filtering recommender can be suggested in the study "Collaborative filtering recommender system in primary care: towards a trusting patient-doctor relationship" (Han et al., 2018b). This system can be further extended to allow primary care physicians to make referrals to specialists.

In addition, the healthcare provider would be advised against hiring primary care doctors to work in more than one hospital at the same time, as in the regression analysis of that group of doctors it was determined that working in two hospitals affects negatively the referral rate of physicians.

6.4. ISOLATED NODES

In what concerns the sample of isolated nodes in the primary-specialty referral network it would be prudent to analyze those cases in particular. In the group of isolated primary care physicians, all the 14 doctors have 15 or less consultations between the years of 2012 and 2017. Whereas in the sample of connected general practitioners only approximately 4% gave no more than 15 consultations. Additionally, the scenario is also preoccupying in the isolated specialists' group as more than 70% of them have less than 16 consultations. But in the case of connected specialists, they represent only 3% of that sample.

7. FUTURE WORK

7.1. VALIDATION OF THE CONCLUSIONS

Before implementing any initiatives to solve the identified problems of the primary-specialty referral network it would be in the best interest of the healthcare provider and any entity that produces a similar study to verify if their conclusions are factual. Meaning, the network was built under certain assumptions, which produced the final output of a bipartite network between specialists and family doctors and in this network it was discovered a bias towards the female gender. But if the network was built under other assumptions it might had produce other results. Therefore, is important to conduct further studies to verify if the conclusions taken from this research project are accurate. This could be done through surveys, or through a more innovative approach yet more costly by creating an application or program specifically designed to keep track of the referrals inside a healthcare provider.

7.2. POLICIES AND INITIATIVES

Initiatives such as the creation of campaigns to empower the female gender and the creation of networking event were suggested in the expectation of contributing to a more effective primary-specialty referral network. However, it is still necessary to develop a plan of action for those ideas. In addition, it would be exciting to conduct a study which would determine if healthcare provider and its patient would benefit from a higher level of cooperation between hospitals from the perspective of the referral mechanism.

7.3. STUDY THE PRIMARY-SPECIALTY REFERRAL THROUGH TIME

It would be interesting to study the evolution of the network throughout time would allow us to have insights about how the communities in the network are evolving over time and how their characteristics change. Additionally, it would uncover what characteristics of the network are constant which are not. Moreover, it could help understand what impact different policies implemented in different periods may had in this informal structure of a healthcare provider.

8. BIBLIOGRAPHY

American Academy of Family Physicians. (2016). Primary Care -- AAFP Policies, 1. Retrieved from http://www.aafp.org/about/policies/all/primary-care.html#1

Amplifon. (2017). PORTUGAL É O 2.º PAÍS NO MUNDO COM PIOR AUDIÇÃO. Retrieved from https://www.amplifon.com/web/pt/novidades-blog/-/portugal-e-o-2-pais-no-mundo-com-pior-audicao

An, C., O'Malley, A. J., & Rockmore, D. N. (2018). Referral paths in the U.S. physician network. *Applied Network Science*, *3*(1). https://doi.org/10.1007/s41109-018-0081-4

An, C., O'Malley, A. J., Rockmore, D. N., & Stock, C. D. (2018). Analysis of the U.S. patient referral network. *Statistics in Medicine*, *37*(5), 847–866. https://doi.org/10.1002/sim.7565

B.E., L., N.L., K., M.L., B., J.-P., O., S., P., A.J., O., … N.A., C. (2012). Variation in patient-sharing networks of physicians across the United States. *JAMA - Journal of the American Medical Association*, *308*(3), 265–273. https://doi.org/10.1001/jama.2012.7615

Baganha, M. I., Ribeiro, J. S., & Pires, S. (2002). O Sector da Saúde em Portugal: Funcionamento do Sistema e Caracterização Sócio-Profissional. Retrieved from https://ces.uc.pt/publicacoes/oficina/ficheiros/182.pdf

Barnett, M. L., Keating, N. L., Christakis, N. A., O'Malley, A. J., & Landon, B. E. (2012). Reasons for choice of referral physician among primary care and specialist physicians. *Journal of General Internal Medicine*, *27*(5), 506–512. https://doi.org/10.1007/s11606-011-1861-z

Barnett, M. L., Landon, B. E., O'Malley, A. J., Keating, N. L., & Christakis, N. A. (2011). Mapping physician networks with self-reported and administrative data. *Health Services Research*, *46*(5), 1592–1609. https://doi.org/10.1111/j.1475-6773.2011.01262.x

Barnett, M. L., Song, Z., & Landon, B. E. (2012). Trends in Physician Referrals in the US, 1999–2009. *Archives of Internal Medicine*, *172*(2), 163–170. https://doi.org/10.1001/archinternmed.2011.722

Care, P., Referral, P., Means, G., & Problems, M. (2019). Primary Care Physicians and Specialists Referral Gap Means Marketing Problems, 1–6.

Chen, J. E., & Glover, G. H. (2016). HHS Public Access, *25*(3), 289–313. https://doi.org/10.1007/s11065-015-9294-9.Functional

Clark, L. (2006). *Network Mapping as a Diagnostic Tool Manual*.

Cross, R., Parker, A., & Borgatti, S. (2000). A bird's-eye view: Using social network analysis to improve knowledge creation and sharing. *Knowledge Directions*, *2*, 48–61. https://doi.org/10.2307/1315064

Deloitte. (2019). *2019 Global Health Care outlook: Shaping the Future*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-2016-health-care-outlook.pdf%5Cnhttp://www.deloitte.com/healthcareoutlook.

Diário da República. Portaria n.o 1498/2004, 2002 § (2004). Retrieved from http://data.dre.pt/eli/port/1499/2004/12/28/p/dre/pt/html

Directions, F. (n.d.). SOCIAL NETWORK ANALYSIS. https://doi.org/10.3366/saj.2004.26.1-2.vi

Forrest, C. B., Nutting, P. A., Von Schrader, S., Rohde, C., & Starfield, B. (2006). Primary care physician specialty referral decision making: Patient, physician, and health care system determinants. *Medical Decision Making*, *26*(1), 76–85. https://doi.org/10.1177/0272989X05284110

Franks, P., Williams, G. C., Zwanziger, J., Mooney, C., & Sorbero, M. (2000). Why do physicians vary so widely in their referral rates? *Journal of General Internal Medicine*, *15*(3), 163–168. https://doi.org/10.1046/j.1525-1497.2000.04079.x

Frey, D., Jégou, A., & Kermarrec, A. M. (2011). Social market: Combining explicit and implicit social networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6976 LNCS*, 193–207. https://doi.org/10.1007/978-3-642-24550-3_16

Ghomrawi, H. M. K., Funk, R. J., Parks, M. L., Owen-Smith, J., & Hollingsworth, J. M. (2018). Physician referral patterns and racial disparities in total hip replacement: A network analysis approach. *PLoS ONE*, *13*(2), 1–10. https://doi.org/10.1371/journal.pone.0193014

Governo da República Portuguesa, M. da S. (2017). Ministério da Saúde – SNS. Retrieved from https://www.sns.gov.pt/institucional/ministerio-da-saude/

Governo da República Portuguesa, M. da S. (2019). Internato Médico - ACSS. Retrieved from http://www.acss.min-saude.pt/2016/09/12/internato-medico/

Han, Q., Ji, M., Gaur, M., Zejnilovic, L., Lisboa, N. De, & Economics, R. (2018a). A Hybrid Recommender System for Patient-Doctor Matchmaking in Primary Care. *IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 481–490.

Han, Q., Ji, M., Gaur, M., Zejnilovic, L., Lisboa, N. De, & Economics, R. (2018b). Collaborative filtering recommender system in primary care: towards a trusting patient-doctor relationship. *6th IEEE International Conference on Healthcare Informatics, ICHI 2018*, 377–379.

Jauhar, S. (2019). Referral System Turns Patients Into Commodities, 3–5.

Kenis, P., & Oerlemans, L. (2009). The Social Network Perspective: Understanding the Structure of Cooperation. *The Oxford Handbook of Inter-Organizational Relations*, (April 2015). https://doi.org/10.1093/oxfordhb/9780199282944.003.0011

Kinchen, K. S., Cooper, L. A., Levine, D., Wang, N. Y., & Powe, N. R. (2004). Referral of patients to specialists: factors affecting choice of specialist...: EBSCOhost. *Annals of Family Medicine*, *2*(3), 245. https://doi.org/10.1370/afm.68

Larson, J. S. (1996). The World Health Organization’s Definition of Health: Social versus Spiritual Health. *Springer*, *38*, 181–192. Retrieved from https://www.jstor.org/stable/27522925?seq=1#page_scan_tab_contents

Liddy, C., Arbab-Tafti, S., Moroz, I., & Keely, E. (2017). Primary care physician referral patterns in Ontario, Canada: A descriptive analysis of self-reported referral data. *BMC Family Practice*, *18*(1), 1–8. https://doi.org/10.1186/s12875-017-0654-9

Notícias, D. de. (2019). Ministra assume que deve falhar meta de dar médico de família a todos. *Diário de Notícias*. Retrieved from https://www.dn.pt/pais/interior/ministra-assume-que-deve-falhar-meta-de-dar-medico-de-familia-a-todos-os-portugueses-10730984.html

O'Malley, Ann S., MD, M., & Reschovsky, J. D. (2011). Referral and Consultation Communication Between Primary Care and Specialist Physicians. *Archives of Internal Medicine*, *171*(1), 56–65.

Ordem Dos Médicos. (2018). Ordem dos Médicos – Portal Oficial. Retrieved from https://ordemdosmedicos.pt/

Ponte, C., Moura, B., Cerejo, A., Braga, R., Marques, I., Teixeira, A., … Trancoso, J. (2006). Referenciação aos cuidados de saúde secundários. *Revista Portuguesa de Clínica Geral*, *22*(5), 555–568. https://doi.org/10.32385/RPMGF.V22I5.10282

Público. (2005). Metade dos portugueses tem problemas de visão. Retrieved from https://www.publico.pt/2005/03/30/sociedade/noticia/metade-dos-portugueses-tem-problemas-de-visao-1219451

República, D. A. Lei n.º 117/2015 (2015).

República, D. da. Decreto-Lei n.º 13/2018 (2018). Governo de Portugal. Retrieved from https://dre.pt/web/guest/home/-/dre/114766032/details/maximized?serie=I&print_preview=print-preview&day=2018-02-09&date=2018-02-01

República, D. da. Portaria n.º 78/2018 (2018). Retrieved from https://dre.pt/application/file/a/114880161

Retrato da Saúde 2018. (2018). Retrieved from https://www.sns.gov.pt/wp-content/uploads/2018/04/RETRATO-DA-SAUDE_2018_compressed.pdf

Saúde, J. de M. (2017). CUF forma os primeiros médicos internos do setor privado _ CUF. Retrieved from https://www.saudecuf.pt/cuf/noticias/cuf-forma-os-primeiros-medicos-internos-do-setor-privado

Schroeder, M. (2016). What's Behind That Medical Referral?, 1–8. Retrieved from http://health.usnews.com/health-news/patient-advice/articles/2016-04-07/whats-behind-that-medical-referral

Song, Y., B. Keller, F., & Zheng, L. (2017). *Social Network Analysis Methods and Examples*. *Sage* (Vol. 0249). Retrieved from https://us.sagepub.com/sites/default/files/upm-binaries/78651_Chapter_1.pdf

Teixeira, F. (2018). Médicos de família fazem em média 21 consultas por dia. *Público*. Retrieved from https://www.publico.pt/2018/07/13/sociedade/noticia/medicos-de-familia-fazem-21-consultas-por-dia-em-media-1837427

WHO. (2017). World Bank and WHO: Half the world lacks access to essential health services, 100 million still pushed into extreme poverty because of health expenses. Retrieved from https://www.malariaworldcongress.org/in-the-news/2018/1/11/world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses%0Ahttp://www.who.int/news-room/detail/13-12-201

9. APPENDIX

9.1. DOCTOR SPECIALTIES GLOSSARY

Doctor Specialties Glossary	
English	**Portuguese**
Anesthesiology	Anestsiologia
Pathologic anatomy	Anatomia Patológica
Angiology and Vascular Surgery	Angiologia e Cirurgia Vascular
Cardiology	Cardiogia
Pediatric Cardiology	Cardiogia Pediátrica
Cardiac surgery	Cirurgia Cardíaca
Cardiothoracic Surgery	Cirurgia Cardiotorácica
General surgery	Cirurgia Geral
Maxillo-facial Surgery	Cirurgia Maxilo-Facial
Pediatric surgery	Cirurgia Pediátrica
Reconstructive and Aesthetic Plastic Surgery	Cirurgia Plástica Reconstrutiva e Estética
Thoracic surgery	Cirurgia Torácica
Dermato-Venereologia	Dermato-Venereologia
Infectious diseases	Doenças Infeciosas
Endocrinology and Nutrition	Endocronologia e Nutrição
Stomatology	Estomatologia
Gastroenterology	Gastrenterologia
Medical Genetics	Genética Médica
Gynecology / Obstetrics	Ginecologia/Obstetrícia
Immunoallergology	Imunoalergologia
Immunohemotherapy	Imunohemoterapia
Clinical Pharmacology	Farmacologia Clinica
Clinical Hematology	Hematologia Clinica
Sports Medicine	Medicina Desportiva
Work Medicine	Medicina do Trabalho
Physical Medicine and Rehabilitation	Medicina Fisica e Reabilitação
General and Family Medicine	Medicina Geral e Familiar
Intensive Care Medicine	Medicina de Cuidados Intensivos
Internal medicine	Medicina Interna
legal Medicine	Medicina Legal
Nuclear medicine	Medicina Nuclear
Tropical Medicine	Medicina Tropical
Nephrology	Nefrologia
Neurosurgery	Neurocirurgia
Neurology	Neurologia
Neuroradiology	Neuroradiologia
Ophthalmology	Oftamologia
Medical Oncology	Oncologia Médica
Orthopedics	Ortopedia
Otolaryngology	Otorrinolaringologia
Clinical pathology	Patologia Clinica
Pediatrics	Pediatria
Pneumology	Pneumologia
Psychiatry	Psiquiatria
Psychiatry of Childhood and Adolescence	Psiquiatria da Infância e da Adolescência
Radiology	Radiologia
Radioncology	Radioncologia
Rheumatology	Reumatologia
Public health	Saúde Pública
Urology	Urulogia

9.2. DEGREE CENTRALITY EXAMPLE CALCULATIONS

Degree Centrality Example Calculations			
Node	**Direct Ties**	**Possible Ties**	**Degree Centrality**
PCP 1	2	5	0.4
PCP 2	3	5	0.6
PCP 3	2	5	0.4
PCP 4	3	5	0.6
PCP 5	2	5	0.4
SP 1	2	5	0.4
SP 2	4	5	0.8
SP 3	2	5	0.4
SP 4	3	5	0.6
SP 5	1	5	0.2

9.3. BETWEENNESS CENTRALITY NORMALIZATION

Nodes belonging to partition U:

$$\frac{1}{2}\left(\mathrm{m}^2(s+1)^2 + m(s+1)(2t-s-1) - t(2s-t+3)\right)$$

$$\mathrm{s} = \frac{(\mathrm{n}\text{-}1)}{\mathrm{m}}, \mathrm{t} = (\mathrm{n}\text{-}1) \bmod \mathrm{m},$$

Nodes belonging to partition V:

$$\frac{1}{2}\left(\mathrm{n}^2(\mathrm{p}+1)^2 + \mathrm{n}(\mathrm{p}+1)(2\mathrm{r}-\mathrm{p}-1) - \mathrm{r}(2\mathrm{p}-\mathrm{r}+3)\right)$$

$$\mathrm{p} = \frac{(\mathrm{m}\text{-}1)}{\mathrm{n}}, \mathrm{r} = (\mathrm{m}\text{-}1) \bmod \mathrm{n},$$

In both equations n represents the number of nodes in partition U, and m is the number of nodes in partition V.

9.4. BETWEENNESS CENTRALITY EXAMPLE CALCULATIONS

Betweenness Centrality Example Calculations	
PCP 1	0.25
PCP 2	0.25
PCP 3	0
SP 1	0.1
SP 2	0.7

9.5. CLOSENESS CENTRALITY EXAMPLE CALCULATIONS

Closeness Centrality Example Calculations	
PCP 1	1
PCP 2	1
PCP 3	0.75
SP 1	0.71
SP 2	1

9.6. PRIMARY CARE DOCTORS INDEPENDENT VARIABLES

Primary Care Doctors Idependent Variables	
age	number_of_hospitals_3.0
years_of_practice	number_of_hospitals_4.0
closeness	number_specialties_1.0
degree	number_specialties_2.0
gender_F	specialty_Medicina Geral e Familiar
gender_M	specialty_Pediatria
number_of_hospitals_1.0	specialty_Medicina Interna
number_of_hospitals_2.0	

9.7. SPECIALISTS INDEPENDENT VARIABLES

Specialists Idependent Variables	
age	specialty_Cirurgia Geral
years_of_practice	specialty_Cirurgia Plástica Reconstrutiva e Estética
closeness	specialty_Dermato-Venereologia
degree	specialty_Endocrinologia e Nutrição
gender_F	specialty_Psiquiatria
gender_M	specialty_Angiologia e Cirurgia Vascular
number_of_hospitals_2.0	specialty_Anestesiologia
number_of_hospitals_1.0	specialty_Anatomia Patológica
number_of_hospitals_3.0	specialty_Neurocirurgia
number_of_hospitals_4.0	specialty_Neurologia
number_of_hospitals_5.0	specialty_Cirurgia Cardíaca
number_of_hospitals_6.0	specialty_Hematologia Clínica
number_specialties_1.0	specialty_Neurorradiologia
number_specialties_2.0	specialty_Imunoalergologia
specialty_Oftalmologia	specialty_Nefrologia
specialty_Estomatologia	specialty_Cirurgia Cardiotorácica
specialty_Ortopedia	specialty_Cirurgia Pediátrica
specialty_Medicina Nuclear	specialty_Cardiologia Pediátrica
specialty_Ginecologia/Obstetrícia	specialty_Cirurgia Torácica
specialty_Cirurgia Maxilo-Facial	specialty_Radioterapia
specialty_Otorrinolaringologia	specialty_Medicina do Trabalho
specialty_Urologia	specialty_Patologia Clínica
specialty_Radiologia	specialty_Imunohemoterapia
specialty_Pneumologia	specialty_Doenças Infecciosas
specialty_Cardiologia	specialty_Genética Médica
specialty_Gastrenterologia	specialty_Psiquiatria da Infância e da Adolescência
specialty_Oncologia Médica	specialty_Medicina Desportiva
specialty_Reumatologia	

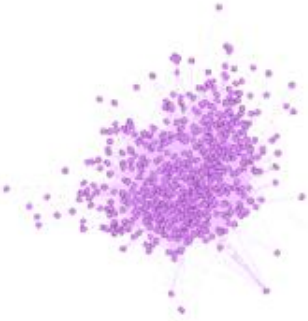
9.8. SPECIALTY FREQUENCY

Specialty Frequency	
Ginecologia/Obstetrícia	249
Pediatria	205
Anestesiologia	158
Ortopedia	160
Medicina Gerla e Familiar	136
Medicina Interna	132
Oftalmologia	127
Cirurgia Geral	98
Otorrinolaringologia	91
Radiologia	79
Urologia	70
Cardiologia	62
Gastrenterologia	56
Neurocirurgia	52
Estomatologia	51
Cirurgia Plástica Reconstrutiva e Estética	45
Neurologia	48
Dermato-Venereologia	45
Angiologia e Cirurgia Vascular	41
Pneumologia	31
Imunoalergologia	31
Psiquiatria	32
Oncologia Médica	23
Endocronologia e Nutrição	23
Reumatologia	21
Cirurgia Pediátrica	17
Cirurgia Maxilo-Facial	15
Anatomia Patológica	13
Others[29]	60

[29] Every Category with a frequency lower or equal to 9 was inserted in the category others. The specialties included in this segment are: Radioncologia; Psiquiatria da Infância e da Adolescência; Cirurgia Cardiotorácica; Medicina Nuclear; Cardiologia Pediátrica; Neurorradiologia; Medicina do Trabalho; Cirurgia Cardíaca; Medicina Desportiva; Cirurgia Torácica; Hematologia Clínica; Genética Médica; Patologia Clínica; Doenças Infecciosas and Imunohemoterapia.

9.9. Gender Distribution Per Specialty (Bottom 29 Specialties)

9.10. Individual Communities



Community 1

Community 2

Community 3

Community 4

Community 5

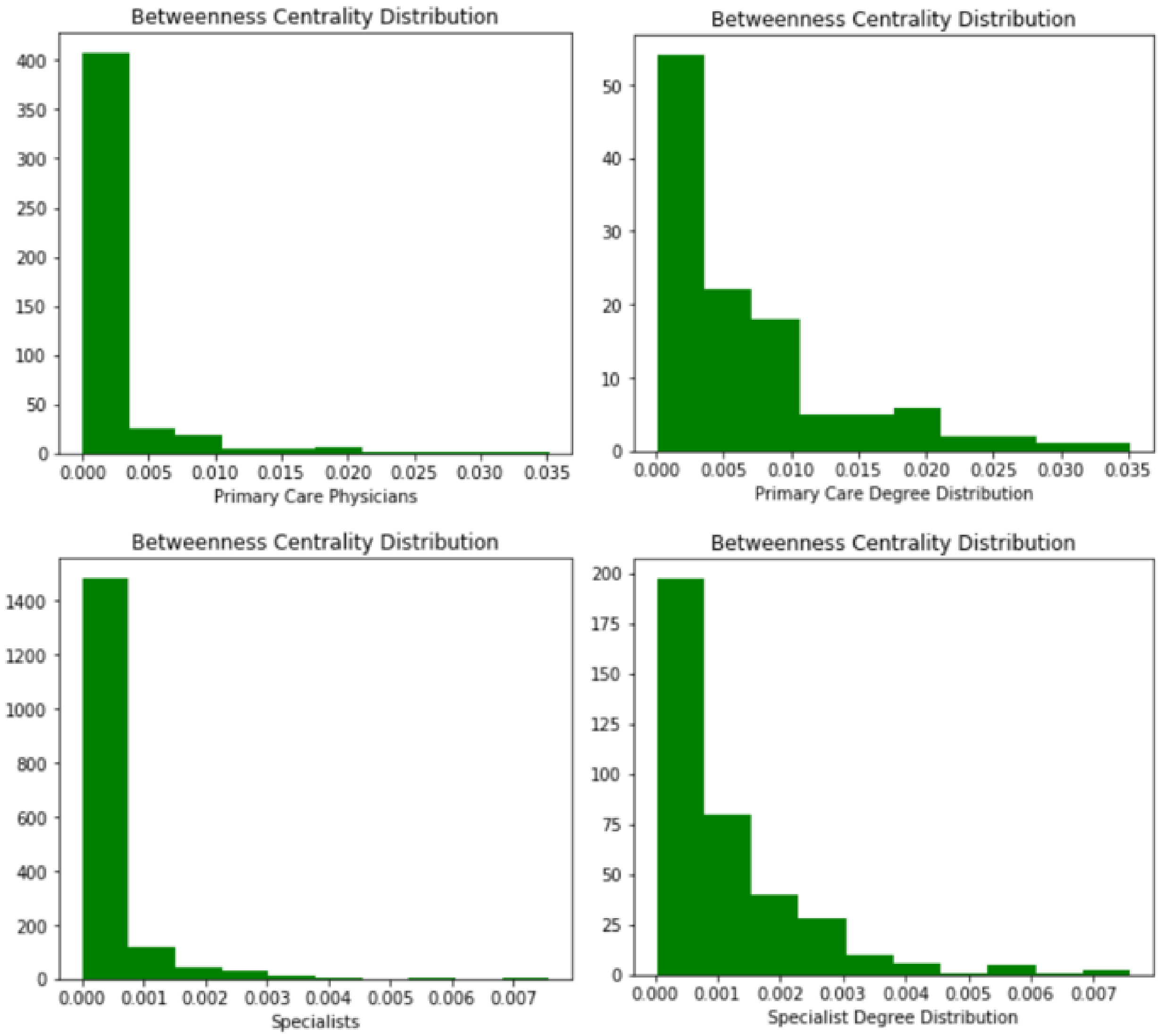
Community 6

Community 7

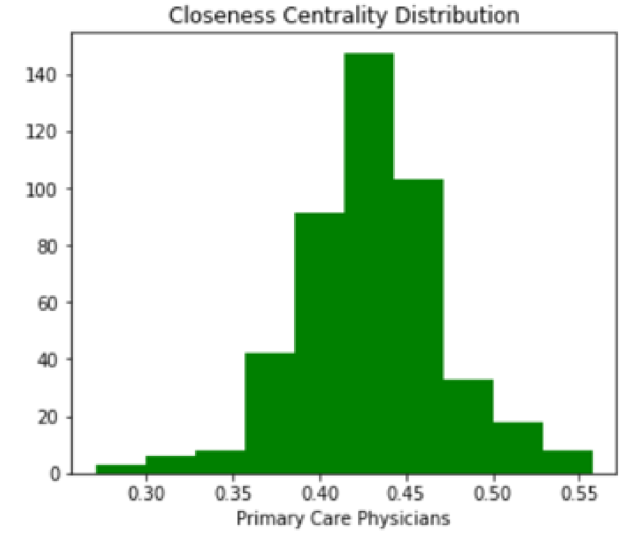
9.11. COMMUNITIES STRUCTURE

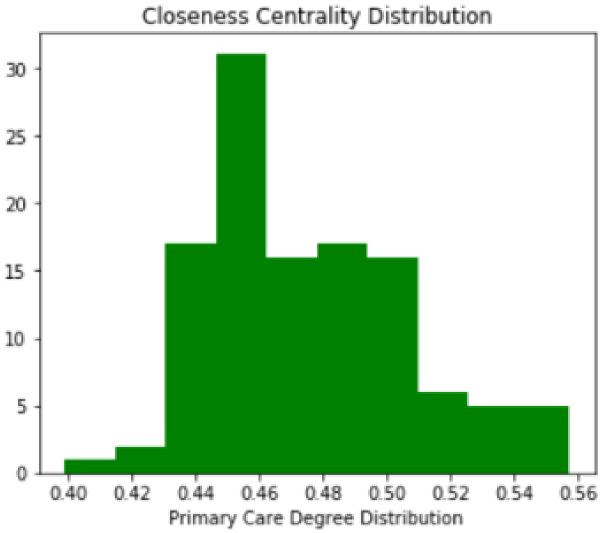
Community	1	2	3	4	5	6	7
Doctors							
# Unique Doctors	634	80	191	198	384	322	137
Gender							
Male	55.84%	43.75%	25.66%	52.53%	50.26%	53.10%	47.45%
Female	44.16%	56.25%	74.34%	47.47%	49.74%	46.90%	52.55%
Doctor Classification							
Primary Care	155	13	95	30	86	57	23
Differentiated Care	479	67	96	168	298	265	114
Hospital							
2	2	19	32	80	75	313	49
5	0	7	21	28	36	28	121
7	634	1	1	1	1	3	0
8	0	25	168	54	370	67	41
11	0	76	34	16	52	34	19
12	0	3	8	116	12	19	8
15	1	4	43	95	32	14	7
Specialty							
Top Specialty	Ginecologia e Obstetrícia; Pediatria; Ortopedia; Oftalmologia; Medicina Interna;	Ginecologia e Obstetrícia; Medicina Geral e Familiar; Radiologia; Gastroenterologia; Oftamologia;	Pediatria; Anestesiologia; Otorrinolaringologia; Oftalmologia; Imunoalergologia;	Ginecologia e Obstetrícia; Oftalmologia; Anestesiologia; Ortopedia; Medicina Geral e Familiar;	Ginecologia e Obstetrícia; Medicina Interna; Medicina Geral e Familiar; Ortopedia; Radiologia;	Medicina Interna; Cirurgia Geral; Anestesiologia; Ortopedia; Otorrinolaringologia;	Anestesiologia; Medicina Geral e Familiar; Ortopedia; Ginecologia e Obstetrícia; Urulogia;
Bottom Specialty	Psiquiatria da Infância e da Adolescência; Cirurgia Maxilo-Facial; Cirurgia Cardíaca; Nefrologia; Anatomia Patológica;	Otorrinolaringologia; Nefrologia; Psiquiatria; Pneumologia; Pediatria;	Cirurgia Plástica Reconstutiva e Estética; Medicina Geral e Familiar; Neurocirurgia; Anatomia Patológica; Psiquiatria;	Imunoalergologia; Neuroradiologia; Nefrologia; Oncologia Médica; Anatomia Patológica;	Neuroradiologia; Nefrologia; Medicina Desportiva; Imunoalergologia; Doenças Infeciosas;	Cirurgia Torácica; Imunohemoterapia; Neurorradiologia; Psiquiatria da Infância e da Adolescência; Cirurgia Pediátrica;	Imunoalergologia; Medicina do Trabalho; Peneumologia; Reumatologia; Cirurgia Cardíaca
Age							
Minimum	27	32	29	32	28	28	28
Average	48	51	48	51	50	52	51
Maximum	77	76	77	80	78	77	76
Clinical Expertise							
Minimum	0	0	0	0	0	0	1
Average	6	7	8	9	9	10	8
Maximum	24	16	16	26	21	39	32

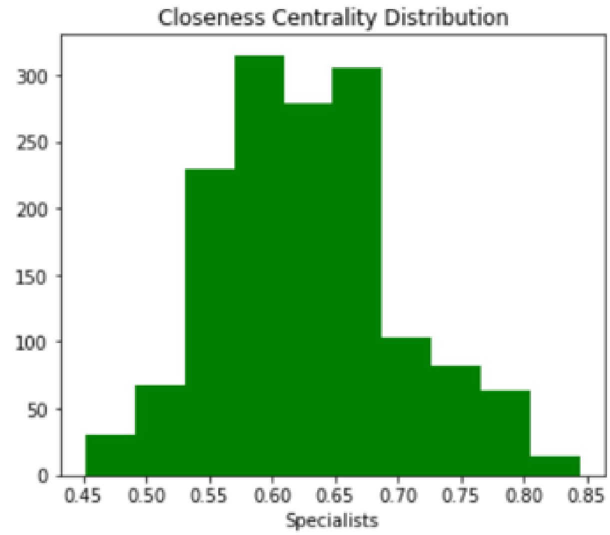
9.12. Doctors with More Activity Outside their Community Betweenness Centrality Distribution

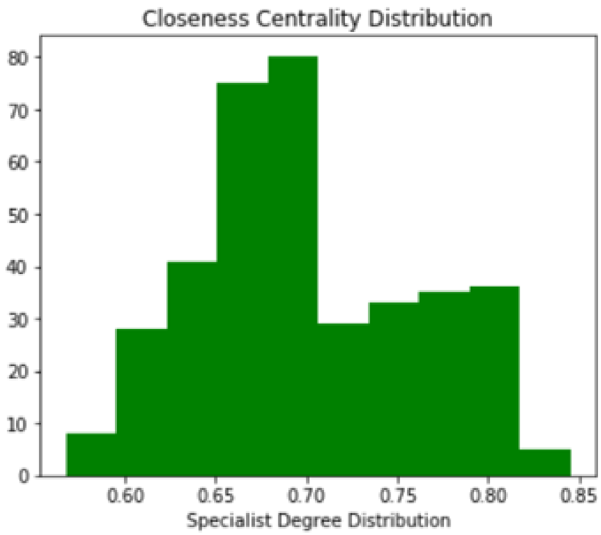


9.13. Doctors with More Activity Outside their Community Closeness Centrality Distribution









9.14. SPECIALISTS REGRESSION ANALYSIS

OLS Regression Results

Dep. Variable:	referral_rate	R-squared:	0.341
Model:	OLS	Adj. R-squared:	0.321
Method:	Least Squares	F-statistic:	16.75
Date:	Sat, 02 Nov 2019	Prob (F-statistic):	1.99e-114
Time:	11:42:39	Log-Likelihood:	1141.1
No. Observations:	1698	AIC:	-2180.
Df Residuals:	1647	BIC:	-1903.
Df Model:	51		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
age	-0.0019	0.000	-6.019	0.000	-0.002	-0.001
years_of_practice	-0.0007	0.001	-0.955	0.340	-0.002	0.001
closeness	0.1264	0.018	7.193	0.000	0.092	0.161
degree	-0.0300	0.058	-0.515	0.607	-0.144	0.084
gender_F	0.0049	0.007	0.698	0.485	-0.009	0.019
number_of_hospitals_2.0	0.0485	0.090	0.536	0.592	-0.129	0.226
number_of_hospitals_1.0	0.0946	0.091	1.046	0.296	-0.083	0.272
number_of_hospitals_3.0	0.0391	0.091	0.429	0.668	-0.140	0.218
number_of_hospitals_4.0	0.0254	0.093	0.273	0.785	-0.157	0.208
number_of_hospitals_5.0	0.0217	0.107	0.203	0.839	-0.188	0.232
number_specialties_1.0	-0.0222	0.030	-0.748	0.455	-0.080	0.036
specialty_Oftalmologia	0.0217	0.091	0.237	0.812	-0.157	0.201
specialty_Estomatologia	-0.0371	0.092	-0.403	0.687	-0.218	0.144
specialty_Ortopedia	0.1004	0.091	1.104	0.270	-0.078	0.279
specialty_Medicina Nuclear	0.0157	0.110	0.142	0.887	-0.201	0.232
specialty_Ginecologia/Obstetrícia	0.0229	0.091	0.253	0.801	-0.155	0.201
specialty_Cirurgia Maxilo-Facial	0.0080	0.096	0.084	0.933	-0.181	0.197
specialty_Otorrinolaringologia	0.0606	0.092	0.660	0.509	-0.119	0.240
specialty_Urologia	0.0626	0.092	0.684	0.494	-0.117	0.242
specialty_Radiologia	0.1224	0.092	1.334	0.182	-0.058	0.302
specialty_Pneumologia	0.0549	0.094	0.587	0.557	-0.129	0.239
specialty_Cardiologia	0.0523	0.092	0.568	0.570	-0.128	0.233
specialty_Gastrenterologia	0.0069	0.092	0.075	0.940	-0.174	0.188
specialty_Oncologia Médica	0.0001	0.094	0.001	0.999	-0.185	0.185
specialty_Reumatologia	-0.0193	0.095	-0.204	0.839	-0.205	0.167
specialty_Cirurgia Geral	0.0610	0.091	0.667	0.505	-0.118	0.240
specialty_Cirurgia Plástica Reconstrutiva e Estética	0.0762	0.092	0.825	0.410	-0.105	0.258
specialty_Dermato-Venereologia	0.0288	0.093	0.310	0.757	-0.153	0.211
specialty_Endocrinologia e Nutrição	-0.0088	0.095	-0.093	0.926	-0.194	0.177
specialty_Psiquiatria	-0.0120	0.094	-0.128	0.898	-0.196	0.171
specialty_Angiologia e Cirurgia Vascular	0.0474	0.093	0.512	0.609	-0.134	0.229
specialty_Anestesiologia	0.0431	0.091	0.473	0.637	-0.136	0.222
specialty_Anatomia Patológica	-0.0155	0.098	-0.159	0.874	-0.207	0.176
specialty_Neurocirurgia	0.0230	0.092	0.250	0.803	-0.158	0.204
specialty_Neurologia	0.0155	0.093	0.167	0.867	-0.166	0.197
specialty_Cirurgia Cardíaca	-0.0136	0.116	-0.117	0.907	-0.241	0.214
specialty_Hematologia Clínica	-0.0005	0.127	-0.004	0.997	-0.250	0.249
specialty_Neurorradiologia	0.1264	0.107	1.181	0.238	-0.083	0.336
specialty_Imunoalergologia	0.0165	0.094	0.176	0.860	-0.167	0.200
specialty_Nefrologia	0.0576	0.100	0.576	0.565	-0.138	0.254
specialty_Cirurgia Cardiotorácica	0.0289	0.104	0.278	0.781	-0.175	0.233
specialty_Cirurgia Pediátrica	0.2481	0.096	2.596	0.010	0.061	0.436
specialty_Cardiologia Pediátrica	0.1106	0.111	1.000	0.318	-0.106	0.327
specialty_Cirurgia Torácica	-0.0133	0.127	-0.105	0.916	-0.262	0.236
specialty_Radioterapia	-0.0038	0.101	-0.038	0.970	-0.202	0.194
specialty_Psiquiatria da Infância e da Adolescência	0.0202	0.101	0.200	0.842	-0.178	0.218
specialty_Medicina do Trabalho	0.0189	0.116	0.163	0.871	-0.209	0.247
specialty_Patologia Clínica	0.0728	0.155	0.469	0.639	-0.231	0.377
specialty_Imunohemoterapia	0.0109	0.156	0.070	0.944	-0.295	0.316
specialty_Doenças Infecciosas	0.0072	0.127	0.057	0.955	-0.242	0.257
specialty_Genética Médica	0.0408	0.155	0.263	0.793	-0.263	0.345

Omnibus:	1683.494	Durbin-Watson:	2.060
Prob(Omnibus):	0.000	Jarque-Bera (JB):	64196.413
Skew:	4.900	Prob(JB):	0.00
Kurtosis:	31.484	Cond. No.	1.04e+04