Computational Social Science

Miniproject 1

Nicolas Bschor Fani Sentinella-Jerbić

June 2023

Contents

1 Introduction

In recent years, the analysis of social networks has become an essential tool for understanding human interactions in various contexts. It has provided valuable insights into the structural and dynamic patterns of relationships between individuals and groups, enabling investigation of different phenomena such as diffusion of information, formation of communities, and opinion polarization. One domain where social network analysis holds potential is the healthcare sector, where interpersonal connections play an important role in patient care, organizational dynamics, and the overall well-being of the healthcare organization.

Thus, our work focuses on the analysis of a temporal social network collected within a hospital setting. What makes this setting exceptionally interesting is the crucial importance of teamwork and collaboration when it comes to the life or death of patients. An in-depth analysis has the potential to reveal how information, expertise, and practices are shared, disseminated, and ultimately implemented in patient care. Furthermore, hospitals represent complex organizations with hierarchical levels, departments, and units where uncovering communication patterns and power structures can help identify areas of improvement and optimize workflows, leading to improved operational efficiency and resource allocation.

2 Dataset Description

For our analysis, we chose a temporal network of interactions from a hospital ward located in Lyon, France. The interactions were detected by the use of wearable sensors, with spatial resolution of approximately 1.5 meters, and temporal resolution of 20 seconds. Meaning, the sensors could detect interactions between individuals within this range. The interactions were tracked in a short-stay geriatric unit of a university hospital, from Monday, December 6, 2010, at 1:00 pm to Friday, December 10, 2010, at 2:00 pm, spanning a duration of four days and nights. Overall, a total of 14,037 contacts were recorded among 46 healthcare workers and 29 patients. Healthcare workers were additionally categorized into four classes according to their work position: medical doctors (physicians and interns), paramedical staff (nurses and nurses' aides), and administrative staff. In the continuation of the document these might be referred to as PAT, MED, NUR, and ADM, respectively.

Furthermore, the data shows interesting insights to investigate the spread of disease, which is explored further in the second mini project of this course. The dataset was actually initially collected for the similar purpose of estimating potential infection transmission routes.

3 Exploration

To gain a comprehensive understanding of the dataset, exploration was the first part of our work. Here we present preliminary findings which inspired our research questions, as well as methods to achieve them.

First, we plot the whole network with distinct nodes colored by class as seen in figure 1. For network layout, we use the Kamada-Kawai algorithm which aims to position the nodes of a graph in a way that minimizes the edge lengths while maintaining a visually pleasing layout. Thus, for nodes positioned in the center, it generally implies that they are more connected with other nodes in the graph. We can see most of the paramedical staff being positioned near the center, as well as patients being positioned more on the outskirts of the graph. This matches the general intuition of nurses being the most *mobile* staff members as they need to take care of many patients, as well as patients being the least connected as they are only temporary visitors in contact with a limited amount of staff.

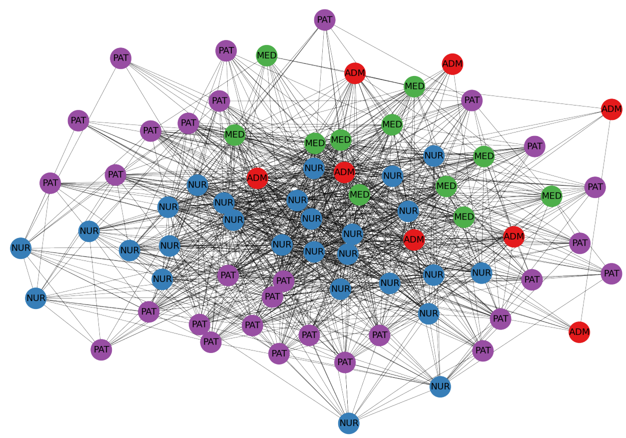

Figure 1: Full dataset graph

Proceeding with this thought we investigate the distribution of the number and frequency of interactions in the network. We can see from figure 2 that some interactions occur way more often than others, indicating some staff members being more connected than others and maybe some patients being in larger need of help than others. Figure 2 covers all interactions including e.g. a nurse visiting a patient twice.

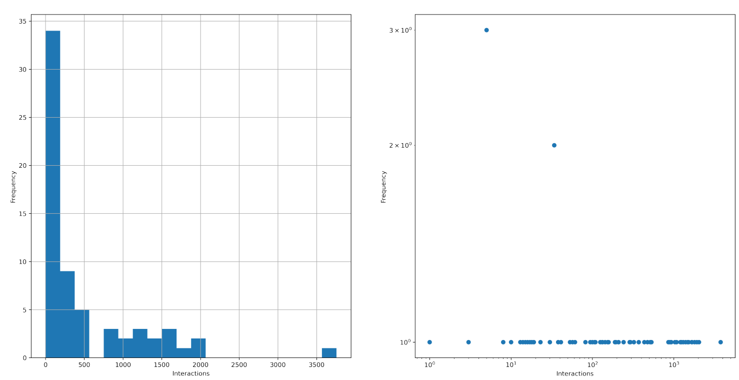

Figure 2: Number of interactions

Next up, we investigate this idea by calculating the degrees – the number of connections each node has with other nodes (time independent duplicated removed). We additionally visualize degrees for each group individually. The histograms are shown in figure 3.

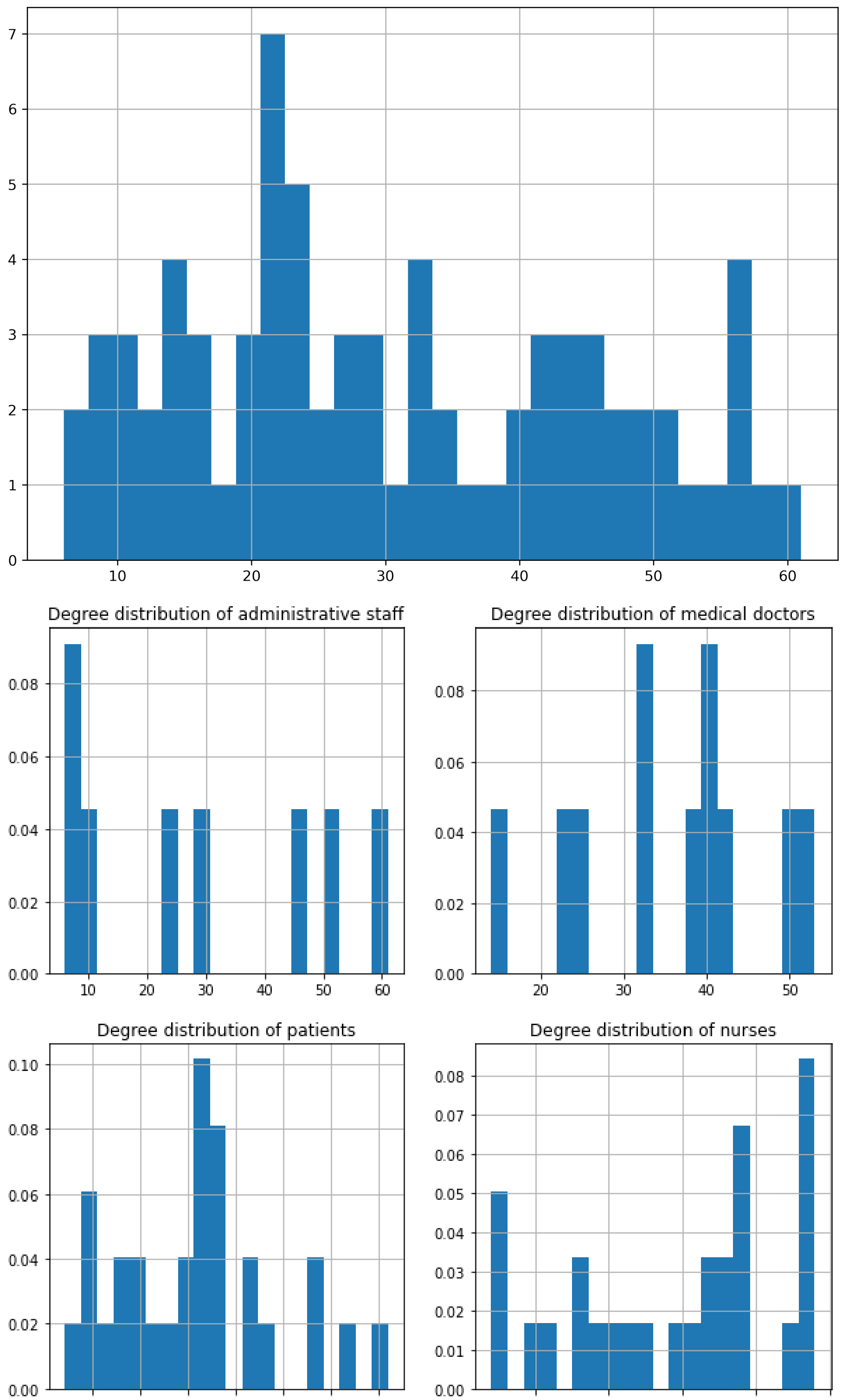

Figure 3: Histograms of degrees

4 Research Questions and Methods

Based on the preliminary results we decided to focus on analyzing **staff dynamics**, as this is something that can help design and distribute future hospital operations. Our investigation revolved around examining **power distribution** among the staff members. We considered the staff power dynamics in two different ways, which we called inter- and intra-group analysis. For this assignment, we decided to focus on **intra-group analysis**. We considered each group independently and compared it with other groups, as well as the whole staff and the whole network. This included investigating the degree of heterogeneity or homogeneity within these groups in terms of centrality. Centrality refers to the measure of how influential an individual is within a network. By assessing the centrality of staff members within different groups, we aimed to ascertain whether power dynamics were evenly distributed or if certain individuals held more influence.

As for the methods to achieve this, we used different centrality measures:

Degree. As already mentioned in the exploration section, the degree of a node refers to the number of connections or edges it has with other nodes. It represents the immediate popularity or connectivity of a node. Nodes with a high degree could be considered more influential within the network.

Degree Centrality. This measure quantifies the importance or centrality of a node based on its degree. It is calculated by dividing the degree of a node by the maximum possible degree in the network. Nodes with a higher degree centrality are considered more central in the network and have a greater potential to influence the flow of information or interactions.

Eigenvector Centrality. This measure takes into account both the number of connections a node has and the importance of those connections. It assigns a score to each node based on the idea that a node is important if it is connected to other important nodes. In other words, the centrality of a node is influenced by the centrality of its neighbors. It is calculated using the eigenvector of the adjacency matrix of the network. Nodes with high eigenvector centrality are not only well-connected but are also connected to other highly central nodes.

Pagerank Centrality. This is a measure inspired by the Google PageRank algorithm, which was originally developed to rank web pages based on their importance. In the context of social networks, PageRank centrality assigns a score to each node based on the idea that a node is important if it is connected to other important nodes. It considers both the number of incoming connections to a node and the importance of those nodes. PageRank centrality uses an iterative algorithm to calculate the centrality scores of nodes in a network. Nodes with higher PageRank centrality are also considered more influential within the network.

Betweenness centrality. This measure quantifies the extent to which a node lies on the shortest paths between pairs of other nodes in a network. A node with high betweenness centrality acts as a bridge or connector between different parts of the network. It captures the idea that nodes with high betweenness have the potential to control the flow of information.

Closeness centrality. Lastly, this one measures how close a node is to all other nodes in terms of geodesic distance (the shortest path length) within a network. It quantifies the efficiency of information or interaction flow from a node to other nodes in the network. Nodes with high closeness centrality can quickly access and disseminate information, making them potentially influential within the network.

For each of these measures we used network and barplot visualizations to show our results.

5 Results

For each staff group as well as the whole staff and the whole network, we showcase the relevant graphs and tables. First, we show the plain network visualization of a group, and then we show the same network with color of nodes corresponding to centrality values from different measures mentioned previously. In the end, an aggregated table of results is shown and visualized using bar plots.

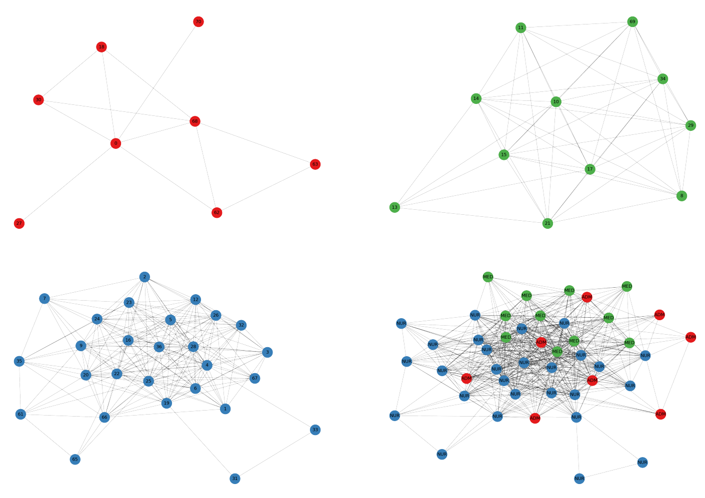

Figure 4: Each group visualized as an individual network – administrative staff, medical doctors, paramedics and all staff together

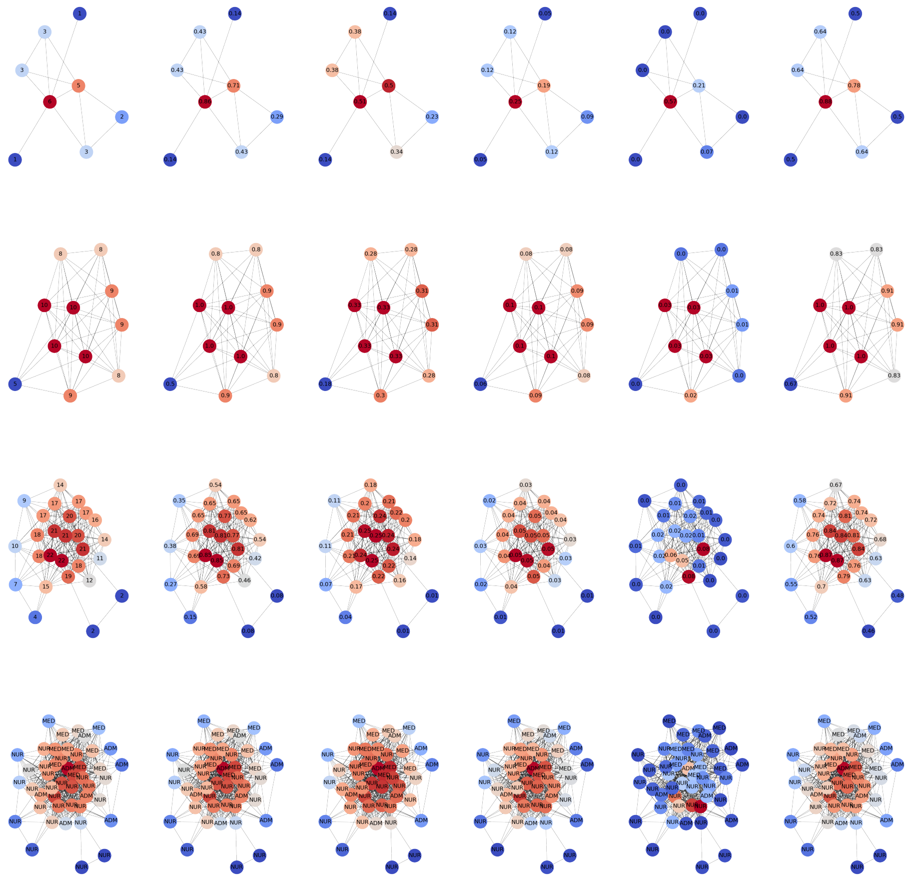

Figure 5: Staff centrality analysis; from top to bottom – administrative staff, doctors, paramedics, whole staff, from left to right – degree, degree centrality, eigenvector centrality, pagerank centrality, betweenness centrality, closeness centrality

graph	MED	NUR	ADM	STAFF	ALL
min_degree	5.00	2.00	1.00	2.00	6.00
avg_degree	8.73	14.96	3.00	24.04	30.37
max_degree	10.00	22.00	6.00	40.00	61.00
min_degree_centrality	0.50	0.08	0.14	0.04	0.08
avg_degree_centrality	0.87	0.58	0.43	0.53	0.41
max_degree_centrality	1.00	0.85	0.86	0.89	0.82
min_eigenvector_centrality	0.18	0.01	0.14	0.01	0.02
avg_eigenvector_centrality	0.30	0.18	0.33	0.14	0.11
max_eigenvector_centrality	0.33	0.25	0.51	0.21	0.19
min_pagerank	0.06	0.01	0.05	0.01	0.00
avg_pagerank	0.09	0.04	0.12	0.02	0.01
max_pagerank	0.10	0.05	0.25	0.03	0.03
min_betweenness_centrality	0.00	0.00	0.00	0.00	0.00
avg_betweenness_centrality	0.01	0.02	0.11	0.01	0.01
max_betweenness_centrality	0.03	0.08	0.57	0.06	0.04
min_closeness_centrality	0.67	0.46	0.50	0.45	0.50
avg_closeness_centrality	0.90	0.71	0.63	0.69	0.64
max_closeness_centrality	1.00	0.87	0.88	0.90	0.85

Figure 6: Aggregated results of centrality analysis per groups

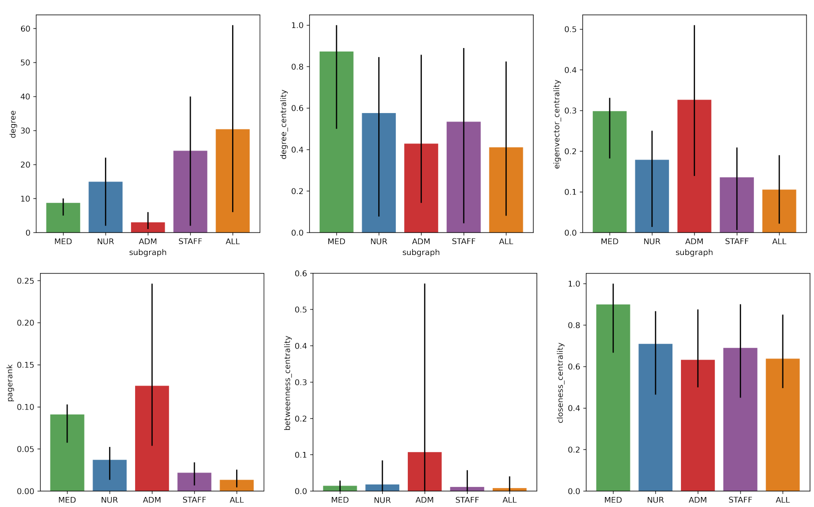

Figure 7: Aggregated results of centrality analysis visualized

6 Discussion, Conclusions and Future Work

To finish off our project, we discuss the results and possible extensions below.

Degree. Administrative staff shows the lowest degree and nurses showcase the highest degree but this measure doesn't serve much information as this it is also affected by the size of the groups.

Degree centrality. Medical doctors show the largest average degree centrality whereas the administrative staff shows the lowest. When looking at the network graphs this makes sense, as the medical doctors seem to be highly connected to each other but administrative staff is highly connected to one specific administrative person.

Eigenvector centrality. Administrative staff have the highest average eigenvector centrality, suggesting they are connected to other highly central nodes within the network. This is consistent with the aforementioned paragraph. Paramedics have the second-highest average eigenvector centrality, indicating they are also more heterogenous as we can also see from the graph that some nurses are only connected to central nodes.

Pagerank. Adminsitrative staff again has the highest average value, meaning they have a high likelihood of being reached, probably by the central administrative person.

Betweenness centrality. Administrative staff again have the highest average value. Here we can see that only one person has a really high value, suggesting they act as a bridge between other nodes, consistent with the other measures.

Closeness centrality. Medical doctors have the highest average closeness centrality, indicating they have the shortest average path length to reach other doctors within the network. When looking at the graph it also makes sense with most doctors having connections with each other.

These results although not fully anticipated, make sense when all things are considered. Administrative staff being more connected makes sense considering there is usually one highly central person, such as the hospital chief. Another surprising finding is that nurses don't seem that much connected as we have expected, maybe suggesting they only seem more connected to people outside the medical field, considering they are the ones dealing with patients the most.

Inspired by our work, we had many ideas on what to investigate further that didn't fit the scope of this project. First up, we would consider inter-group analysis, pairwise per staff groups, especially between medical doctors and paramedics considering the historical context of authority dynamic between the two. It would also be interesting to see if more central groups of staff are also more connected to patients or if it is the opposite case. Furthermore, we would consider performing the same analysis for each day independently to investigate the stability of centrality through time. Lastly, we would consider an ablation of the most influential members in groups, to see how it would affect the network as a whole.