Academic Year 2023/2024

Organizational Network Analysis of Actinvision: Insights from Network Science

Master Thesis

Written by

Zied Kebir

Presented on the 2nd of September 2024 in front of the jury members

Julien Larcher

Julien Pénin

Pierre Pelletier

Contents

1 Apprenticeship **5**
1.1 Introduction of the Company 5
1.2 My role in Actinvision 7
1.3 challenges and lessons 10

2 Background **12**

3 Literature review **13**

4 Network Science Related Notions **15**
4.1 Types of Networks 15
4.2 Node centrality measures 17
4.3 Edge Importance Measures 20
4.4 Overall Measures 20
4.5 Network Models 21

5 Implementation of the network structure **23**
5.1 Data Collection 23
5.2 Data Cleaning and Preprocessing 24

6 Network Science Analysis **26**
6.1 Network Visualization and Overall Analysis 26
6.2 Managers and salesmen importance 30
6.3 Gender importance 33
6.4 Forecasting the social structure in 2027 35
6.5 Network's robustness 38
6.6 Challenges, Limitations, and Extentions 40

7 Conclusion **43**

List of Figures

1 New Managerial Framework Structure 6
2 Ticket's Content 7
3 Power Apps Interface 9
4 KPI - Customer Satisfaction 10
5 Undirected and Directed Networks 15
6 Bipartite Network 16
7 Centrality Measures 18
8 Network's Density 21
9 Users' Dataset 24
10 Chat Logs Dataset 24
11 Messages Dataset 25
12 Network Pre and Post Restructuring 26
13 Distribution of Node Importance Measures 27
14 Distribution of node importance for managers and salesmen 30
15 Distribution of node importance for managers and salesmen 32
16 Distribution of node importance for Male and Female before and after restructuring 33
17 Impact of Restructuring on the Distribution of Nodes Importance 34
18 Power Law Distribution of Nodes' Degrees 35
19 Probability of a Link Being Created per Year 36
20 Distribution of Node importance in 2023-2024 and 2027 37
21 Networks Robustness: Random Attack 39
22 Networks Robustness: Targeted Attack 40

List of Tables

1 Network Metrics Before and After September 2023 29
2 Network Metrics in 2023-2024 and 2027 38

Thursday 22nd August, 2024

Master 2 - Data Science pour l'Economie et l'Entreprise du futur

contnet

1 Apprenticeship

1.1 Introduction of the Company

Actinvision is a consulting company specializing in digital transformation and business process optimization. It was created in 2014 in Strasbourg and mainly focuses on the following areas:
Digital transformation: Helping businesses adopt new technologies to improve their efficiency and competitiveness.
Data management: Implementation of data management solutions to help companies fully exploit their information.
Strategy Consulting: Helping businesses define and implement strategies to achieve their long-term goals.
Training: Providing comprehensive training programs to help companies maximize the effective use of their tools. Actinvision works with companies of different sizes and across diverse industries among them pharma, insurance, retail, and Industry, to help them innovate and stay competitive in an ever-changing environment. The company has roughly 80 employees and has branches located in 4 French cities, namely Strasbourg, Paris, and Lyon. The company offers services revolving around data analysis and integration and has multiple partners like Tableau, Power BI, Matillion, Snowflake, Azure, ThoughtSpot among others.
The company has an interesting managerial framework called holacracy. It is an organizational management system designed to distribute decision-making authority across an organization. Unlike a traditional hierarchical structure, Holacracy offers a decentralized approach where authority is distributed across clearly defined and interdependent roles. Here are some key points of this framework:
Structure and roles
Circles: The organization is divided into circles, which are autonomous units responsible for certain functions or projects. Each circle manages its own operations and makes decisions to achieve its goals.
Roles: Within circles, tasks and responsibilities are divided into specific roles. A role is a defined function with a purpose, responsibilities, and areas of competence. People can play multiple roles in different circles.
Governance process
Governance meetings: These regular meetings allow circle members to define, assign, and adjust the roles and policies necessary for the functioning of the circle.

Tensions: In Holacracy, a "tension" is a situation where a person sees an opportunity for improvement. Tensions are addressed in governance meetings to adjust structure and roles to better meet the needs of the organization. Business processes

Tactical Meetings: These meetings focus on daily work and resolving operational issues. Members discuss progress, obstacles, and next actions to move projects forward.

Facilitator Roles: Each circle has specific roles like the Facilitator and Secretary who help organize and maintain governance processes and tactical meetings.

Transparency and clarity

Transparency: All policies, roles, and tensions are documented and accessible to everyone in the organization. This promotes great transparency and a clear understanding of everyone's responsibilities.

Autonomy and responsibility: Circle members have the autonomy to make decisions within their roles, but they are also responsible for the results. This encourages a culture of responsibility and initiative. Holacracy aims to make organizations more agile and responsive by distributing decision-making power and encouraging autonomy and innovation at all levels. In September 2023, the company slightly tuned its managerial framework by modifying how the holacracy framework was implemented.

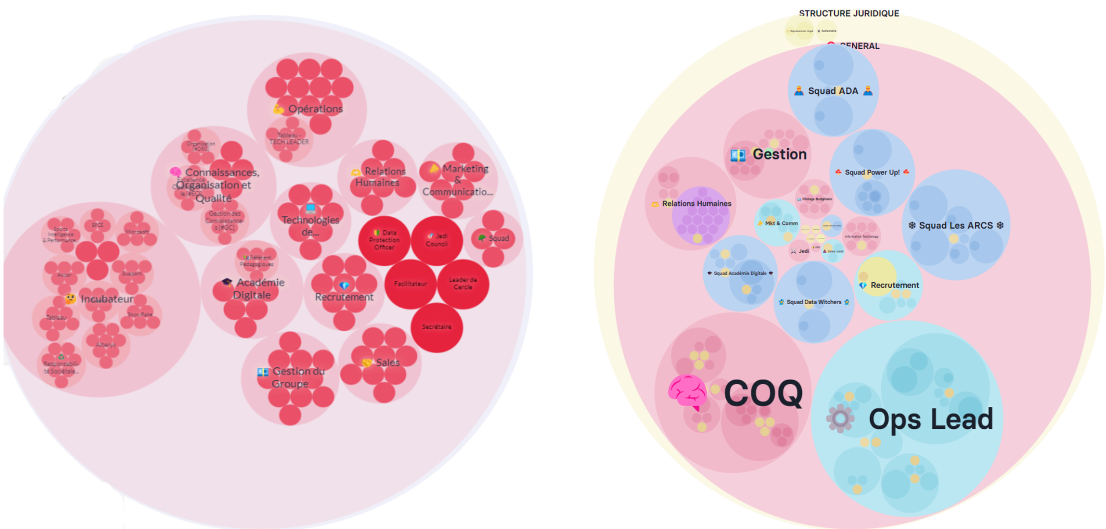


Figure 1: New Managerial Framework Structure

Figure 1 illustrates the structural changes that occurred after September 2023. The figure on the left shows the previous structure, while the figure on the right depicts the new structure. The primary difference is that the former structure was organized by functional groups, or "circles," such as Sales, HR, etc. In contrast, the new structure is more decentralized and consists of teams, referred to as "squads," such as Les Arcs, Ada, Power Up, etc., along with some circles that still represent functions like HR and Management, similar to the old structure.

At the center of each squad, there is a salesperson and a manager. The salesperson's main focus is on securing job offers for their respective team, making each squad relatively independent of the others.

1.2 My role in Actinvision

In Actinvision I was a member of the customer support team. My role was to help and accompany customers in solving their issues. Over 10 clients have subscribed to the customer support service. At any time some of their employees can issue a request or ticket in order to solve a certain issue they faced. Our clients can create tickets via our ticketing system- Azure DevOps Server.

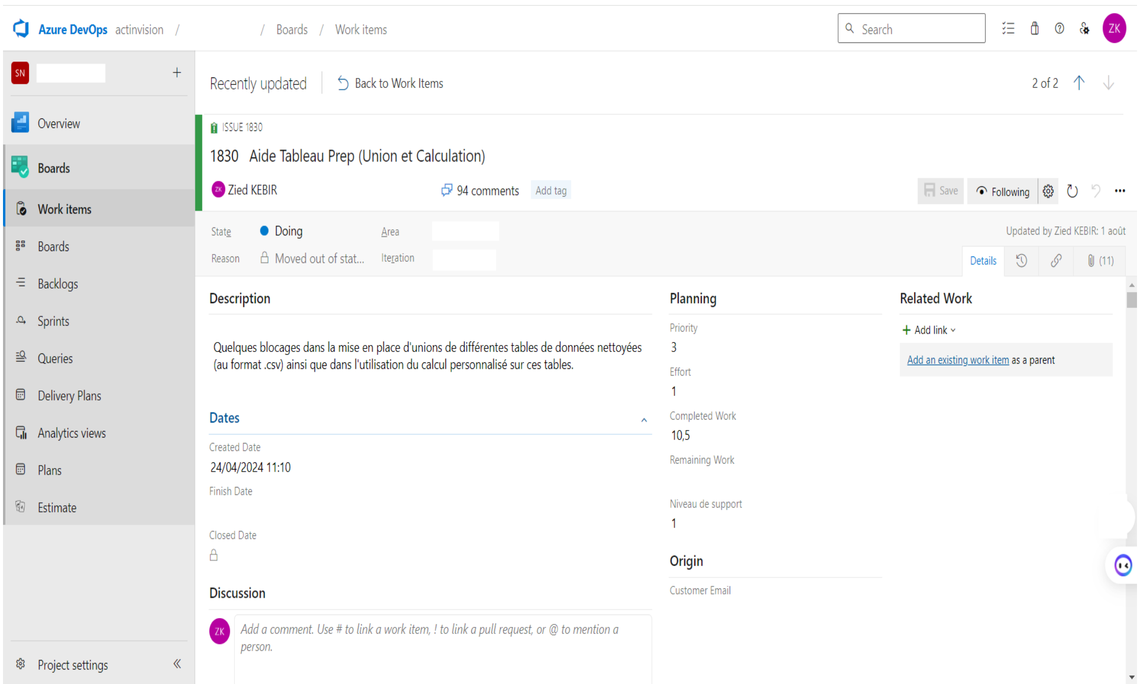


Figure 2: Ticket's Content

As depicted in Figure 2 in a ticket customers can write a description to explain their issue, specify a priority level for the ticket, and contact us at any time via the discussion section. When a client issue arises, I receive a notification and, according to Actinvision's internal policy, I have one hour to respond to the client and assign the ticket to a team member. Several factors influence the decision of who will handle the ticket, with the primary consideration being the availability of the team member. While having the necessary skills to resolve the issue is important, the company values and encourages collaboration and research to find solutions to client problems. Consequently, working on a ticket does not necessarily require deep expertise in the specific area. In fact, I can seek assistance from domain experts at any time or utilize internal communication tools, such as Yammer, to post questions about specific issues. During my year at Actinvision, I had the opportunity to work on a variety of customer issues across a range of topics that were initially unfamiliar to me. In addition, I was also involved in internal projects. Below is a brief overview of the tasks I was involved in. **Data Visualization**

Plenty of our customers are relying on dashboards in order to evaluate their different KPIs. Most of the data visualization tasks I worked on revolved around modifying already existing dashboards in order to satisfy a new need or correct a mistake. This goes from adding filters, modifying graph types, or adding plots to solving server crash issues or optimizing the loading time of a dashboard.

In the data visualization tasks, I primarily worked with Tableau and, to a lesser extent, with Power BI. I completed two fundamental training courses and worked and presented to internal experts two case studies to demonstrate my ability to effectively use these tools.

Access Management

As part of its services, Actinvision provides Tableau licenses along with comprehensive management support. For some clients, I also served as administrator of their Tableau Server systems. This role gave me the opportunity to resolve access issues, such as when employees forgot their passwords or required elevated permissions for specific dashboards.

Data Integration and Database Management

I particularly enjoyed tasks related to data integration and management. These tasks often involved addressing issues that occur upstream in the data flow. For instance, when data appeared incorrect or missing in dashboards, I would methodically trace the data flow upstream, examining databases to identify where the issues originated and working to resolve them. Additionally, I sometimes needed to optimize workflows to ensure that data refresh jobs did not overwhelm the server, preventing crashes caused by simultaneous large data refreshes. Working on such issues allowed me to hone my skills in SQL and learn new tools such as Tableau prep, SQL Server Integration Services (SSIS), and SQL Server Management Studio (SSMS).

API- Data Cleaning and Extraction

In addition within Actinvision I had the opportunity to assist my coworkers who were working on building dashboards but needed help in collecting and structuring the data. In this context, I was a couple of times involved in collecting customer data through API requests using Python. I also had to restructure the data which was stored in JSON and XML files into CSV files so that it could be exploited by my colleagues on Power BI.

Development of HR Management Interface

In the objective of the company to streamline the process of recruiting new talents, I have been involved in an internal project of developing an interface using power apps that allows managers to register talents (Figure 3), see the interviews they have, and schedule new ones, conduct the interview by selecting a list of questions to ask, grade the candidate and write comments, observe the grades and remarks the interviewee got from the whole selection process.

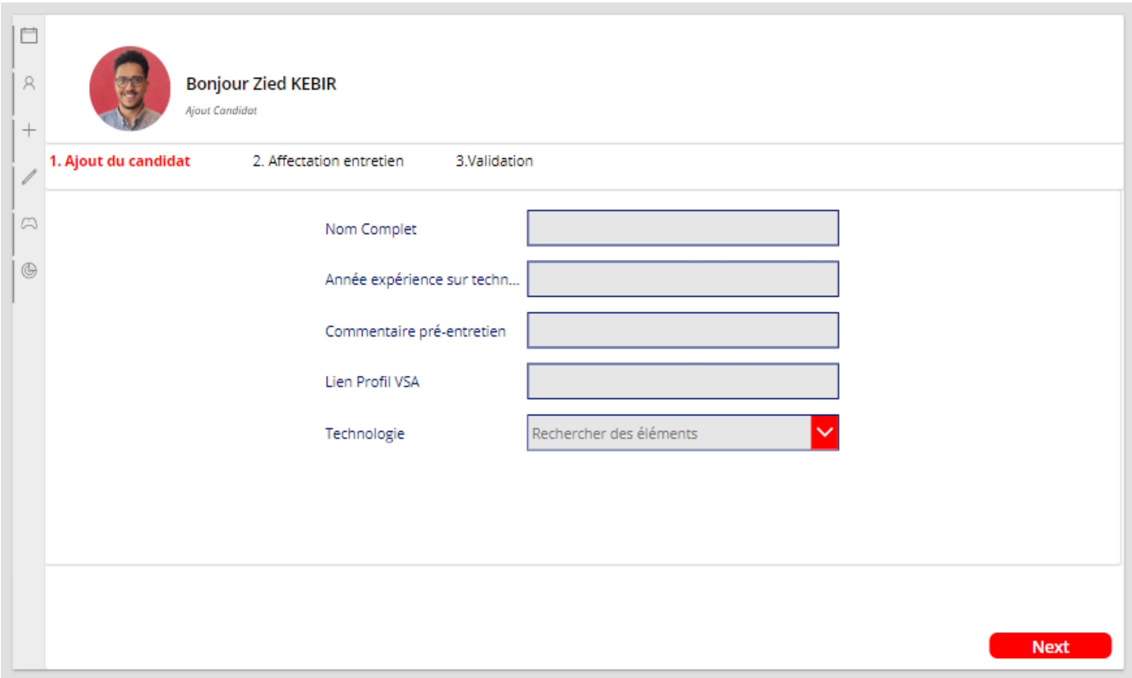
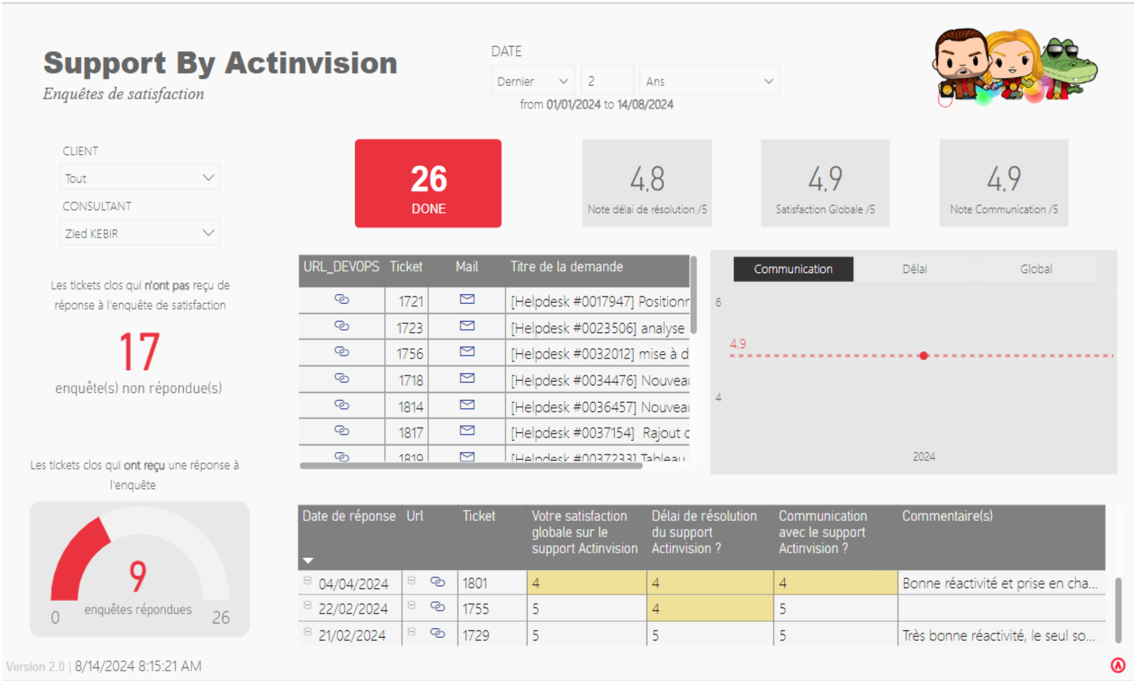


Figure 3: Power Apps Interface

Figure 3 illustrates one of the interfaces I created. It allows users to insert a new candidate. The inserted data is going to be stored in a SharePoint list.

As a result of my apprenticeship within Actinvision, as illustrated in Figure 4, I ended up solving 26 tickets and had pretty good overall customer satisfaction with scores higher than 4.8/5 in solving time, global satisfaction, and communication.



URL_DEVOPS	Ticket	Mail	Titre de la demande
	1721		[Helpdesk #0017947] Positionr
	1723		[Helpdesk #0023506] analyse
	1756		[Helpdesk #0032012] mise à d
	1718		[Helpdesk #0034476] Nouveau
	1814		[Helpdesk #0036457] Nouveau
	1817		[Helpdesk #0037154] Rajout c

Date de réponse	Url	Ticket	Votre satisfaction globale sur le support Actinvision	Délai de résolution du support Actinvision ?	Communication avec le support Actinvision ?	Commentaire(s)
04/04/2024		1801	4	4	4	Bonne réactivité et prise en cha...
22/02/2024		1755	5	4	5	
21/02/2024		1729	5	5	5	Très bonne réactivité, le seul so...

Figure 4: KPI - Customer Satisfaction

1.3 challenges and lessons

In this part, I am going to detail the challenges I faced during my apprenticeship within Actinvision. During the few months within Actinvision, During my first few months at Actinvision, I lacked the confidence in my skills and knowledge to effectively and efficiently resolve tickets. Customer requests were being sent on a daily basis about tools and technologies I had never used before. As a result, I struggled to take the initiative in handling such requests on my own. Throughout my academic journey, I've consistently taken the initiative to independently find solutions to things I don't understand, dedicating time to research and problem-solving. However, in the context of customer service, I had to reply and suggest solutions to the clients as fast as possible. I soon enough realized that often solving a customer issue does not require being an expert in the topic. It is possible and even encouraged to reach out to field experts within the Actinvision in order to find a solution. As a result, I started accepting tickets more and more and trying to solve them, and I ended up learning and improving my skills with new tools such as Tableau, SQL, Tableau prep, SSIS, and SSMS. I have also faced challenges in talking with nontechnical people who often cannot easily formulate or even do not know what they need. Moreover, sometimes I went through overwhelming phases where I had several tickets to work on at the same time. As a result of these challenges in addition to learning hard skills, I improved my time management skills and teamwork. As sometimes I helped or requested help from my coworkers to solve tickets. Moreover, I improved my communication skills because I often had to summarize what I had been discussing with a client to coworkers who were not at all acquainted with the issue at hand.

2 Background

This study comes in the context of my one-year apprenticeship at Actinvision, a year during which I had to write my master's thesis. Different factors and interesting aspects have led me to analyze the network structure of Actinvision. First, during my first days in the company, I was introduced to the management framework of the company- holacracy - and how they recently changed the way the framework was implemented within the company. Moreover, the dynamic of the company highly emphasizes social interactions. In fact, people to a certain extent are identified by their skills in a specific technology like Tableau, Power BI, or Snowflake among others. Some employees are identified as tech leaders and are known as references to whom anyone should refer to in case of any need or help regarding the use of a specific tool. As a result, the peculiar managerial framework, my keen interest in the field of network science as well as the rising interest nowadays in leveraging employee data to discover and comprehend how an organization works [14] enticed me to implement a network analysis on the social structure of the company. The aim of my analysis is to first model the social structure at Actinvision, contrast the network and flow of information before and after the restructuring of the company, checking whether there are key individuals or even bottlenecks. Such analysis could be used to furthermore increase the cohesion and efficiency of the plethora of social interactions happening on a daily basis within the company.

3 Literature review

Network science can be defined as the study of the analysis, visualization, collection, and management of relational data [10]. It is a relatively recent field that gained a lot of consideration in the last two decades. It has been considered as an academic field only recently in 2005 after the US National Research Council categorized it as belonging to the field of basic research [27]. However, network science is rooted in graph theory which is a field that falls as far back as 1736. In fact, during that year Euler applied fundamental network science concepts to solve the Koningsburg bridge problem [1]. However, the subsequent 200 years didn't result in any advances in the field until the 20th century when pillars of network science among them Erdos, Reny, Barabarasi, Albert, Gilbert, or Milgram rekindled scientific interest in the discipline. Among their major breakthroughs, there is the social study of Stanley Millgram in his paper "The small world problem", in which he introduced the concept of the small world problem stating that two individuals are connected through a small number of intermediaries disregarding the size of the whole population. He managed to empirically demonstrate through the 6 degrees of separation phenomena that the distance between two individuals in the United States selected at random doesn't exceed 6 intermediaries [26]. In 1999, the physistics Albert-Laszlo Barab´asi and Reka Albert in 1999 when analyzing the growth of the World Wide Web network, coined the term preferential attachment. This concept illustrates the "rich getting richer" phenomenon. In a network context, it means that the more a node is connected the more likely it is to receive more connections in the future [2]. The issue of the concept of preferential attachment has led to the rise of a new kind of network called scale-free networks. They are non-random networks, in which nodes' links follow a power-law distribution. In other words, they are characterized by a few nodes, called hubs that are highly connected and many nodes that have relatively few connections. The application of network science in political science, sociology, biology, and a lot of other fields suggested that a lot of real-world networks are scale-free, thus furthermore exacerbating the interests of scientists in unveiling the truth behind the formation of observable networks [24]. While the findings mentioned above are not exhaustive they give an idea of the major breakthroughs in the field of network science and how it increasingly gained interest in the research field. Nowadays, research and applications of network science are applied in fields like computer

[1]The Königsberg Seven Bridges Problem seeks to determine whether there is a path back to the starting point by crossing each bridge in the city only once.

science, physics, engineering, mathematics, telecommunications, neuroscience, and psychology... [27] and can efficiently model complex phenomena and structures that affect our daily life [18]. For example, the spread of epidemics and how different vaccination strategies may impede the development of the disease [30, 11], the structure of the financial market, how the different institutions are connected, and how the whole financial system evolves in time of crises [21]. In a nutshell, as the famous Australian sociologist, Duncan Watts, said "Networks are important because if I don't understand networks, I can't understand how markets function, organizations solve problems, or how societies change."[10]. In the context of this study I will focus on a particular application of network science called organizational network analysis (ONA). ONA represents a group of methods and theories that help grasp and improve organizational collaboration and interactions. It leverages communication data to structure a network that highlights the connections between individuals and teams and explains the mechanism through which organizational processes are executed [28]. Nowadays, traditional tools for analyzing a company's efficiency are inefficient and cannot explain the dynamics behind an organization's flow of knowledge, resources, risks, and tasks. In addition, a lot of studies show that social relationships and the networks they form play a big role in how knowledge is created, shared, understood, and used [31]. This has led companies to shift to an anticipatory approach to management and to analyze people's interactions to unveil the mechanisms behind a company's overall performance [19].

Concretely, organizational network analysis helps identify key players or inversely isolated members. It can ensure a more efficient flow of information within teams. In addition, it can exacerbate interaction and collaboration between people with different backgrounds and expertise, thus increasing the sense of responsiveness by allowing individuals to know the right people for a certain issue or opportunity in addition to promoting innovation [13]. In this study, I will model the network structure of a consulting firm composed of up to 80 individuals that switched to a new management framework based on holacracy in September 2023. I will use network science to analyze the impact of such change on the robustness, structure, and flow of information within the company. In addition, following the company's expansion strategy of reaching 300 employees in 2027, will forecast the impact that such a strategic endeavor would have on its social network.

4 Network Science Related Notions

In this section, I will detail some important concepts related to network science that will help us better understand the analysis I conducted during this study.

4.1 Types of Networks

There are different types of networks each having specific characteristics and each type of network is designed to meet specific needs. I will talk about 4 different types of networks, namely directed, undirected, weighted, and bipartite networks.

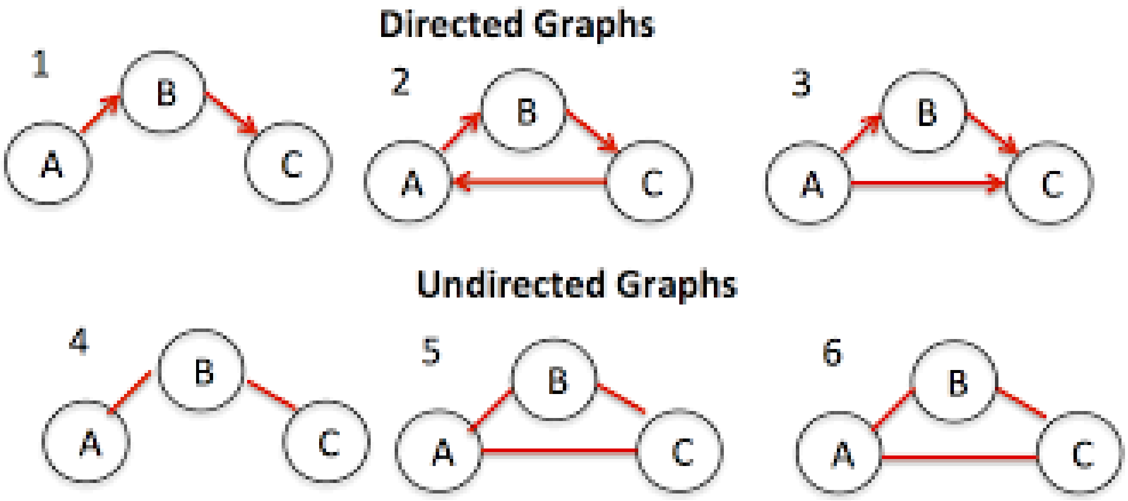


Figure 5: Undirected and Directed Networks

Source: Medium - Network Science Series 002

Undirected Networks

A directed network also called a directed graph or digraph, denoted as G is defined by a pair of sets G=(V,E). V is a finite, non-empty, and countable set called vertices or nodes. E represents a set of unordered pairs of distinct vertices, called edges or links. Formally, $V = v_1, v_2, ..., v_n$ and $E \subseteq \{\{u,v\} \mid u,v \in V \text{ and } u \neq v\}$. In an undirected network, each edge is an unordered set of vertices meaning that $\{v_1, v_2\} = \{v_2, v_1\}$ and two vertices joined by an edge are called neighbors, connected or adjacent nodes. Moreover, adjacency matrices often denoted A_{ij}, are sometimes used in the literature to represent a network. The adjacency matrix keeps track of all links available in a network. For example, if a link exists between node i and j then the adjacency matrix would have $A_{ij} = 1$ and $A_{ji} = 1$. Thus by definition, the adjacency matrix of an undirected graph is always symmetric [5].

In a network the total number of nodes is called the cardinality of V or size of the graph, denoted N. On the other hand, the number of edges or cardinality of E is denoted by M. Moreover, In an undirected network, the maximum number of edges is computed by $\frac{N*(N-1)}{2}$ assuming that the network is a simple graph, thus no node is its own neighbors [8].

Directed Networks

A directed graph, denoted D is also defined by a a non-empty and countable set of nodes V. However, the edges, are directed meaning that they are a set of ordered pairs of vertices. Formally $\{v_1, v_2\} \neq \{v_2, v_1\}$. As depicted in Figure 5, edges are illustrated by arrows indicating the direction of the connection between two adjacent vertices in contrast to undirected graphs whose edges are represented by simple lines [8]. Similarly to an undirected graph, a directed network can also be formalized using an adjacency matrix. However, in this case, it is not symmetric.

Weighted Networks

Weighted networks are graphs whose edges have weights that represent the significance, or strength of a relation [3]. Formally a weighted graph is defined by the set WG = (V,E,w) where w represents a weight function that assigns a value to every edge in the network [1]. The adjacency matrix of a weighted graph would store the weight w_{ij} of the relationship between node i and j if it exists [7]. This kind of network integrates the idea that edges carry a certain amount of information. Thus a weight can represent different attributes, like the distance between nodes, to what extent two nodes are similar, or a quantity moving between two adjacent vertices. Such a network can be applied in modeling trading, transportation, or recommendation systems for example [17].

Bipartite Networks Bipartite networks are special types of graphs whose nodes are divided

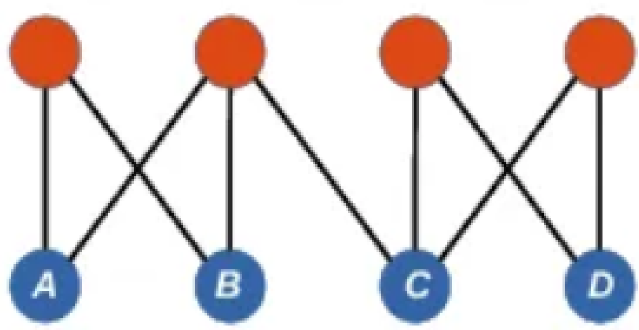


Figure 6: Bipartite Network

Source: Medium - Network Science Series 002

into two groups and links are only possible between nodes in different groups. Formally this kind of network is defined as $G = (V_1 + V_2, E)$ with V_1 and V_2 being two independent sets of nodes [8]. Such graphs have multiple applications such as modeling the relation between buyers and products when doing a market basket analysis or the relation between individuals and possible recommendations in a recommender system.
In the context of the study, I have decided to model the network of the company using an undirected graph. The reason behind this choice is mainly the time constraint. In fact, given that the project has been conducted over a short period of 6 months implementing an undirected and unweighted network seemed to be the easiest type to implement and was already able to return some interesting and informative results.

4.2 Node centrality measures

Different measures have been developed to quantify the importance or centrality of a node within a network. The importance is usually measured to what extent a certain node is connected to other nodes, its position within the network, or the importance of its neighbors. Usually, the metrics I will use in this research focus on quantifying the importance of a node based on one criterion at a time. Even though some studies developed more complex and multi-criteria metrics to quantify the centrality of a node [36, 37] I will keep things simple by focusing on standard node importance measures, namely the degree, betweenness, closeness, eigenvector, and PageRank centrality. Moreover, as mentioned previously in this study I modeled the social structure of the company using an undirected and unweighted graph, thus all the measures I will formalize and explain in this, and the further sections will be related to an undirected network.

Degree Centrality

Degree centrality simply measures the number of connections a node has. As a result, this measure assumes that the importance of a node is measured by the number of neighbors it has. It is formulated as,

$$k_i = \sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ji} \tag{1}$$

It is typically normalized by dividing it by N-1 to reduce the influence of the number of nodes in the network. The issue with this measure is that it is a local measure that doesn't take into consideration the rest of the network. As a result, nodes with the highest degree might be

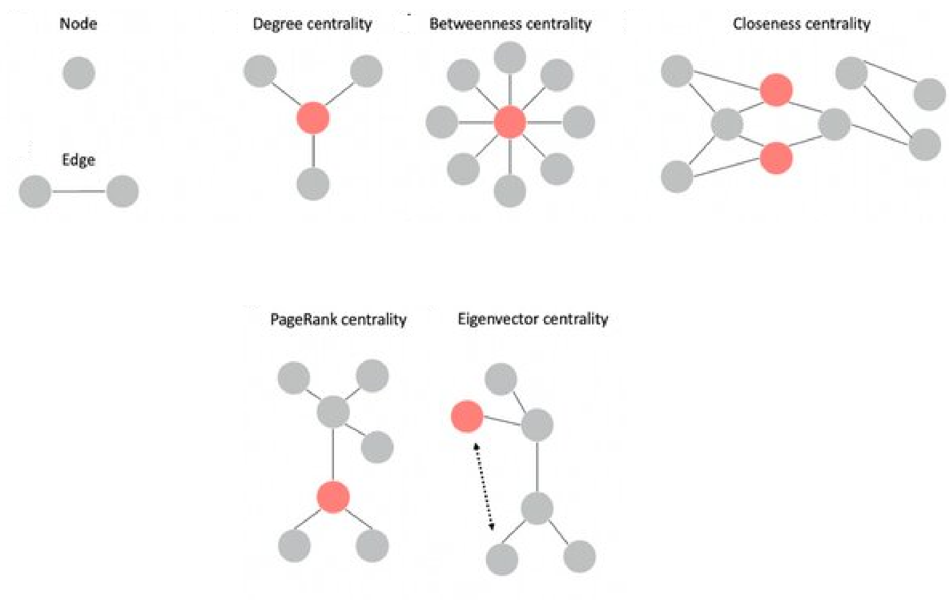


Figure 7: Centrality Measures

Source: [34]

located in the periphery and not at the center.

Closeness Centrality

Closeness centrality of node i denoted as C_i, measures to what extent a node is close to other nodes in the network. The closeness centrality of node i is denoted as C_i and is computed as

$$C_i = \frac{1}{d_i} \tag{2}$$

$$d_i = \frac{1}{N-1}\sum_{j=1}^{N} d_{ij} \tag{3}$$

where d_i is the average distance between node i and all the other nodes in the network and d_{ij} is the distance between node i and j. A distance between two vertices is the length of the shortest path between them [25]. In contrast to degree centrality, closeness centrality is not a local structure since it uses a more holistic approach. However, it is more computationally expensive ($O(n*m)$ with n being the number of nodes and m the number of edges) [32]. In addition, as it relies on computing distances such a measure assumes that the network is fully connected meaning that two nodes are always connected by at least one path. Moreover, closeness centrality gives equal importance to all the nodes even those far away.

Betweenness Centrality

Betweenness centrality measures the node importance by checking to what extent a node acts as a bridge between two other nodes. In other words, how often a node appears along the shortest path connecting two nodes. It is formally defined as,

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4}$$

with σ_{st} measuring the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ indicating the total number of shortest paths that pass through v. The major issue with betweenness centrality is that it has a relatively high time complexity $O(n^3)$ [9].

Eigenvector centrality

The major idea behind eigenvector centrality is that a node is more important if it is connected to important nodes [6]. The eigenvector of node i denoted x_i is computed by aggregating the the eigenvector of all the adjacent nodes. Formally

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} A_{ij} x_j \tag{5}$$

Where A is the adjacency matrix of the network and λ is a constant. The equation can also be written as

$$Ax = \lambda x \tag{6}$$

This corresponds to the eigenvector equation with λ being the eigenvalue of A and x the eigenvector. This equation has only a unique solution for which all values in x are different than 0 [33]. While this node importance measure adopts a global approach and redefines the way the centrality of a node is computed by looking at how important its neighbors are [4]. it is highly impacted by the presence of large hubs in the network. In fact, following the definition of eigenvector centrality a node adjacent to a hub would have a higher centrality thus furthermore increasing the centrality of the hub [29].

PageRank

PageRank centrality is a measure that is based on the eigenvector centrality and mitigates the effect of large hubs. It was developed by Larry Page and Sergey Brin, the founders of Google and it is used to determine the relevance of a web page in search engine results. Formally the PageRank vector is the result of the convergence of the vector π through an iterative process that models a Markov chain which is formally defined as,

$$\pi^{(k+1)T} = \pi^{(k)T} G \tag{7}$$

Where G is called the Google matrix and is computed as,

$$G = \alpha S + (1-\alpha)\frac{1}{nee^t} \tag{8}$$

where N is the number of nodes e is a vector of ones, and S is a matrix containing all the probabilities of jumping from a node to one of its adjacent nodes. In S nodes that are completely disconnected from the network, which are called dangling nodes, will have a probability $\frac{1}{N}$ of moving to any other node within the network. α, called the damping factor, represents the probability of jumping from a node to any other node in the network even a non-adjacent one. This equation highlights the idea behind how people navigate across web pages. In fact, a random person when accessing web page 1 can either move to another page through the multiple links within page 1 or he can type a new address in the research. The former action is denoted by the transition probabilities within the S matrix while the latter is denoted by α [23].

4.3 Edge Importance Measures

Betweenness Centrality

The betweenness centrality of an edge is similar to that of a node. It is formulated as,

$$EBC(e) = \sum_{s \neq t \in V} \frac{\sigma_{st}(e)}{\sigma_{st}} \tag{9}$$

where $\sigma_{st}(e)$ is the number of shortest paths between node s and t that pass through edge e, and σ_{st} is the total number of shortest paths between node s and t. A node with high betweenness centrality is able to connect different parts of a network.

4.4 Overall Measures

In this subsection, I will present a few global network science measures, which are neither node nor edge-specific. The measures I will present are the density, average shortest path, and diameter.

Density

The density measures the percentage of possible connections in the network that are realized. Formally it is defined as

$$D = \frac{\text{Number of edges}}{\text{Total number of edges possible}} = \frac{2M}{N(N-1)} \tag{10}$$

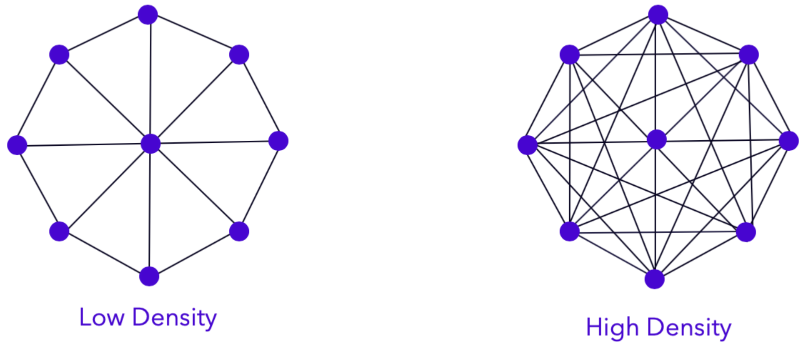


Figure 8: Network's Density

Source Linkedin - Network Effects Total Guide

Where M is the number of edges in the network and $\frac{N(N-1)}{2}$ is the maximum number of possible edges in the network.

Average Shortest Path

It is simply the average of all the shortest distances between two nodes. Formally it is defined as,

$$APE = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{11}$$

where d_{ij} is the shortest path between node i and j, and $\frac{N(N-1)}{2}$ is the total possible combination of node pairs.

Diameter

Diameter is defined as the shortest path between the two most distant nodes. Formally it can be defined as,

$$D = max_{i,j \in V} d_{ij} \tag{12}$$

4.5 Network Models

Network models are constructs that try to capture and reproduce real-world networks. Overall such models are used to answer questions such as how the network evolved to reach a certain state ? or what rules lead to a certain network structure when taking into consideration specific constraints? [20]. I will focus on explaining two famous network models namely Erdős–Rényi Model, or the random graph model, and Barabási–Albert Model, also called the scale-free network model. Such models have been applied for example to evaluate the growth of the World Wide Web network, to analyze power grids in order to reduce the risks of blackouts, or to figure out strategies to reduce the spread of a disease during a pandemic [12].

Erdős–Rényi model: Random model

The Erdős–Rényi model was invented by the two famous Hungarian mathematicians Pál Erdős and Alfréd Rényi. A random graph is defined as G(N,L) where N indicates the number of nodes and L is the number of edges. In this model, L edges are added randomly to connect two nodes. We first start with a network with N nodes and no edges. Then, L random pairs of nodes are chosen with equal probability, and edges are added to the network to connect these selected nodes [16, 15].

Barabási–Albert model: Scale-free model

The Barabási–Albert model is inspired by the preferential attachment concept. Recall that preferential attachment implies that nodes that are highly connected, referred to as hubs, are more likely to be connected to a new node that enters the network.

To generate a network using this model, we first start with a small network with m_0 nodes and edges selected arbitrarily as far as each node has at least one edge. Then a new node is added to the network and can connect to any other node i with a probability $\pi(k_i)$

$$\pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{13}$$

where k indicates the degree of a node. We can see that the probability of a new node being connected to node i depends on the degree of node i.

5 Implementation of the network structure

In this section, I will explain the steps taken prior to the network analysis, which mainly consisted of two parts: data collection and data cleaning.

5.1 Data Collection

The aim of the data collection part was to obtain enough historical data with good quality that contains enough information in order to be able to build a network. The data should have been able to a certain extent to illustrate the relationship between two coworkers. As the company mainly uses Microsoft Teams as the internal tool of communication, I decided to extract chat logs of the messages sent via Microsoft Teams using Microsoft Graph API. A chat log is a record that stores information, such as sender, receiver, date ..., of a conversation. As a result, through chat logs by leveraging on information such as the sender, receiver, and the date a message was sent I have enough information to be able to model the social network of Actinvision. The overall data collection process consisted of first extracting all the users with their relevant information, including the full name, the ID, and the email address. When extracting the users a lot of entries had to be removed given that some of them were external clients that were not relevant in the context of the study. I had to filter these individuals out by checking whether the domain of their email address corresponds to '@actinvision' or not. After doing that I moved to the next phase consisting of extracting all the chat IDs for each user. Basically, I parsed through each of the Actinvision users and extracted all the chat IDs they were involved in. In this step, I made sure to only focus on chat logs involving a one-to-one conversation. All messages sent in group chats were filtered out. This resulted in a data frame including one entry indicating the name and ID of 2 users involved in the chat as well as the ID of that chat. Following this step, I parsed through all the chat IDs to get more granular information about the communication between the two users. As a result of this step, I ended up with a large data frame including message information such as the name and IDs of the users involved in the message as well as the date the message was sent, and the chat ID of the conversation. It is important to note that all the data I extracted after executing calls to the API were stored in CSV files. To sum up, as a result of the data collection phase I ended up with three CSV files containing data that will be used later on to build the networks. The first data frame included information about employees at Actinvision (Figure 9)

The second data frame included overall chat logs information (Figure 10). For later uses, I

node_id	user_id	Squad	hiredate	Nom	jobTitle	email	Team_Lead	Sales	Ancienneté	genre
48	fd042edd-c74	Squad Power Up!	2023-09-04T00:00:00Z	Zied KEBIR	Apprenti BI/Big DATA	zkebir@actin	No	No	236	m

Figure 9: Users' Dataset

decided to add more information in the first dataset. I first manually included a column that indicates whether a user is a manager or not. In addition, I included information about the gender of the users. Since I did not have any access to personal information I used a database from data.gouv that included over 15,000 first names and genders associated with them. Based on this I matched the first names with the corresponding gender. Moreover, I added a column that includes the seniority of a person. To measure that I used a proxy corresponding to the difference between the date of the last message sent and the first message sent.

chat_id	user_1_id	user_1_Name	user_1_email	user_2_id	user_2_Name	user_2_email
19:5808019c-b719-4e1b-a10b-61a089c27710_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	5808019c-b719-4e1b-a10b-61a089c27710		
19:041a7c0c-57ca-469b-9517-4c22ba3100cf_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	041a7c0c-57ca-469b-9517-4c22ba3100cf		
19:cf8cd409-6f5a-4f26-b2fe-bf9fdce76ec3_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	cf8cd409-6f5a-4f26-b2fe-bf9fdce76ec3		
19:c3990175-b28b-486a-8d06-81bf2ff6f1e9_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	c3990175-b28b-486a-8d06-81bf2ff6f1e9		
19:615277d3-3de3-49e2-8f6b-4daee60495ee_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	615277d3-3de3-49e2-8f6b-4daee60495ee		
19:7a925fcf-9b8a-4254-97b4-b0e91466b024_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	7a925fcf-9b8a-4254-97b4-b0e91466b024		
19:b332c9a5-a6e4-48fd-97dc-5b113ec301c9_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	b332c9a5-a6e4-48fd-97dc-5b113ec301c9		
19:8adc32a0-9b0b-4feb-9c0d-a1fe51adf334_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	8adc32a0-9b0b-4feb-9c0d-a1fe51adf334		
19:b09e3fcd-0b67-4001-916c-f7bc7336cb2e_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	b09e3fcd-0b67-4001-916c-f7bc7336cb2e		
19:65426112-89f5-4e84-80b9-0451e82a2e85_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	65426112-89f5-4e84-80b9-0451e82a2e85		
19:d6a7c661-1d54-4a2c-a321-e0283e263bb2_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	d6a7c661-1d54-4a2c-a321-e0283e263bb2		
19:d9ddead9-8e11-4a7c-ad00-5967fb16621b_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	d9ddead9-8e11-4a7c-ad00-5967fb16621b		
19:20cfefb4-af20-4cb0-bb74-8a0f0e8e5736_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	20cfefb4-af20-4cb0-bb74-8a0f0e8e5736		
19:4b68c2a2-9da1-4cdb-9ab9-76f259684ada_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	4b68c2a2-9da1-4cdb-9ab9-76f259684ada		
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	9603dc53-0d36-4888-9572-67b86e0058fe		
19:54f45727-ce9d-471f-8d9b-591e2bc78dd7_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR	zkebir@actinvision.com	54f45727-ce9d-471f-8d9b-591e2bc78dd7		

Figure 10: Chat Logs Dataset

The third dataset (Figure 11) included details about the messages sent in each chat. For each chat ID, I have multiple logs each log indicating when the messages were sent and what users were involved.

5.2 Data Cleaning and Preprocessing

Before being able to build the network and actively analyze it, I went through a time-consuming phase of data cleaning. This phase included a lot of sub-steps. First, I unified the user IDs and user names so that each user is identified by exactly one ID and one user name. The reason behind multiple IDs or user names might be explained by the fact that sometimes individuals within the company have two laptops one from Actinvision and another one from a client. As a result, these users when being involved in a new conversation with Actinvision

chat_id	message_content	createdDateTime	deletedDateTime	lastModifiedDateTime	message_sender_id	ame_user_	id_user_1	name_user_2	id_user_2
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-04-12T09:16:21.342Z		2024-04-12T09:18:27.547Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-04-12T07:51:02.754Z		2024-04-12T07:51:02.754Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-03-15T15:25:43.95Z		2024-03-15T15:25:43.95Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-03-15T15:23:09.779Z		2024-03-15T15:23:09.779Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-03-15T14:18:10.135Z		2024-03-15T14:18:10.135Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-03-15T11:05:57.107Z	2024-04-05T12:39:44.816	2024-04-05T12:39:44.816Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-03-15T08:26:27.508Z		2024-03-15T08:26:27.508Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-12T09:27:40.654Z		2024-02-12T09:27:40.654Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-12T09:07:13.549Z		2024-02-12T09:07:38.817Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-10-25T10:02:11.36Z		2023-10-25T10:02:11.36Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-10-25T09:47:28.052Z		2023-10-25T09:47:28.052Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-10-25T09:46:35.896Z		2023-10-25T09:46:35.896Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-10-23T07:45:36.713Z		2023-10-23T07:45:36.713Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:0a697fe2-f10e-4df5-8ebf-5eb284078cfa_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-10-23T07:44:54.285Z		2023-10-23T07:44:54.285Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			0a697fe2-f10e-4df5-8ebf-5eb284078cfa
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-22T12:07:21.04Z		2024-02-22T12:07:21.04Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-22T12:05:04.467Z		2024-02-22T12:05:04.467Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T13:34:11.653Z		2024-02-21T13:34:11.653Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T13:18:12.902Z		2024-02-21T13:18:12.902Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:59:02.365Z		2024-02-21T12:59:02.365Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:53:37.625Z		2024-02-21T12:53:37.625Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:45:58.815Z		2024-02-21T12:45:58.815Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:45:25.553Z		2024-02-21T12:45:25.553Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:44:54.997Z		2024-02-21T12:44:54.997Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2024-02-21T12:18:03.747Z		2024-02-21T12:18:03.747Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-28T09:32:40.387Z		2023-09-28T09:32:40.387Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-28T09:31:41.807Z		2023-09-28T09:31:41.807Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-28T09:30:40.11Z		2023-09-28T09:36:05.53Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-25T16:01:03.024Z		2023-09-25T16:01:03.024Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-25T16:00:37.757Z		2023-09-25T16:00:37.757Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-25T15:59:01.462Z		2023-09-25T15:59:01.462Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-25T15:54:02.895Z		2023-09-25T15:54:02.895Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe
19:9603dc53-0d36-4888-9572-67b86e0058fe_fd042edd-c743-4d5f-90ba-cf10bf636063@unq.gbl.spaces		2023-09-25T15:47:57.78Z		2023-09-25T15:47:57.78Z	fd042edd-c743-4d5f-90ba-cf10bf636063	Zied KEBIR			9603dc53-0d36-4888-9572-67b86e0058fe

Figure 11: Messages Dataset

users would have a different ID and username. Second, I had to manually delete all non-human users, such as "actinsupport" with an "@actinvision" email address, which represents entities that send automated messages. Third I extracted information about individuals who left the company prior to the start of this project. In fact, data of employees who left the company are deleted after a period of time. Such users do not appear when an API call to extract all users within the company is executed. However, they do appear when they are involved in conversations with people who are still in the company. To illustrate this with an example, let's suppose I have user A, who was still in the company at the start of the project, and user B who left Actinvision in 2022. At the time the project started, in 2024, no data could be directly extracted by requesting all the chat messages of user B. However, if A and B have exchanged messages at a certain point they are stored in the message logs of user A. So I was able to extract data about users who quit by looking at data from currently active users. But as you might have noticed such implementation only allows to partially circumvent the issue. In fact, if I have a user C who also left the company in 2022. Then it is impossible to collect the messages sent between users B and C. I decided to simply disregard this case as it was impossible to capture such data. Fourth I fragmented the dataset containing the detailed information about messages into two parts; messages sent between actinvision's coworkers prior to September 2023 and messages sent after that period. The reason behind this was to analyze the impact that the managerial restructuring within the company had on the network structure.

6 Network Science Analysis

In this section, I will perform an analysis of the two networks I mentioned above and contrast their results

6.1 Network Visualization and Overall Analysis

As a result of the cleaning phase, I ended up with three datasets that I will use to build the network. Namely, a dataset containing the users and their relevant information (ID, Name, Email, Gender ...), and two datasets containing the logs of messages sent among Actinvision's users before and after the restructuring of the company. To build the network I first started by creating a node for each user. Then I counted per day, for each conversation, which is uniquely identified by a chat ID, how many messages were sent. After that, I grouped by chat ID in order to get the average number of messages sent per day for each conversation. I then averaged over the average number of messages sent per day in order to get a threshold starting from which a connection between two individuals would be deemed relevant enough so that a link can be created to connect the two nodes corresponding to these two users. In other words, I looped through all the chat IDs and checked whether the count per day of that conversation was larger than the threshold. If yes, I would create a link connecting the two individuals involved in the conversation otherwise, I would evaluate the next chat ID. I repeated this process twice one time for data prior to September 2023 and one time for the data after the restructuring. For both cases the threshold used was 16, meaning that if a link between two nodes exists then on average per day 16 direct messages are exchanged on Microsoft Teams between the two nodes. Figure 12 below displays the obtained networks.

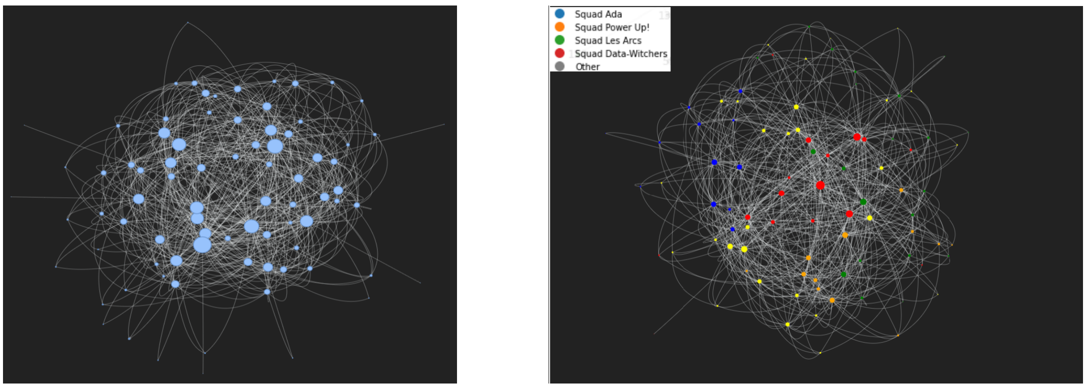


Figure 12: Network Pre and Post Restructuring

The left network displays the social structure within Actinvision prior to the managerial restructuring that was completed roughly in September 2023. On the other hand, the right plot displays the network structure following the implementation of the restructuring. The size of the node in both graphs captures the node degree. Thus, the bigger the node the more connected it is. The colors in the right plot illustrate to what squad or team a node belongs. Visually, if I contrast the two networks, I first notice that prior to September 2023, there were more peripheral nodes connected to only one individual. These nodes could be easily cut off from the network, unlike in the current network. Moreover, looking at the second network I notice a high connectivity in between different squads however I notice that individuals belonging to the same team are often closer to each other.

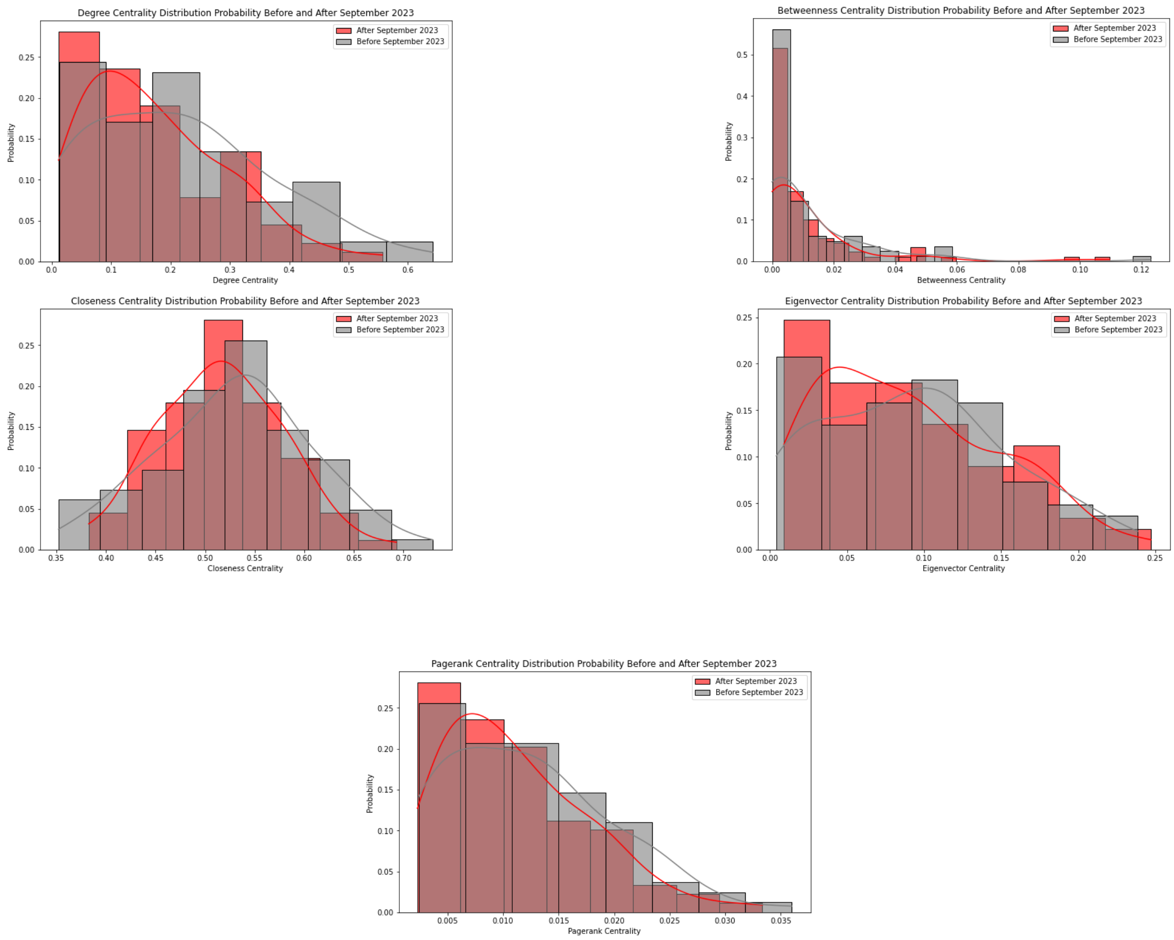


Figure 13: Distribution of Node Importance Measures

Figure 13 contrasts the distribution of the different node importance measures between the two networks. The red graphs represent the distribution for the current network while the grey one displays the distribution for the network prior to the restructuring of the company. Below I will contrast the results of the two networks for each centrality measure.

Degree Centrality

The two networks seem to have really few hubs and several nodes with a low number of connections. If I focus on the density plots I see that after the restructuring of the company, the proportion of nodes with low degree centrality is higher as compared to before the restructuring. On the other hand, the network structure of Actinvision post-restructuring displays more nodes with a larger degree centrality compared to the current one.

Betweenness Centrality

Both networks have close betweenness centrality measures. Before September 2023 the network displays slightly larger betweenness centrality measures reaching 0.12. Moreover, both networks share the same distribution trend with a lot of nodes with high betweenness centrality and relatively few nodes with middle or high values.

Closeness Centrality

The distribution of the closeness centrality measure differs from other measures. For both networks, I observe that relatively fewer nodes have high values, while several nodes have closeness centrality values between the two extremes. The network prior to September 2023 exhibits a wider range of closeness centrality, spanning from 0.35 to above 0.70, in contrast to the current network, which has a slightly narrower range between 0.4 and below 0.70. Moreover, the former network has more nodes with relatively high and low closeness centralities compared to the current network.

Eigenvector Centrality

I notice that, after September 2023, the distribution is more concentrated towards low Eigenvector centrality values, with a higher probability around 0.25. In contrast, before September 2023, the distribution is more spread out, with a greater probability of slightly higher Eigenvector centrality values. This suggests that after the restructuring in September 2023, nodes in the network tend to have lower Eigenvector centralities, indicating an overall decrease in the relative influence of nodes in the network.

PageRank Centrality

I observe that, after September 2023, the distribution is more concentrated towards the very low centrality values of PageRank. In contrast, before September 2023, the distribution is more spread out with a relatively higher probability for PageRank centrality values up to around 0.035.

Overall the different measures do not capture the same notion of node importance. This fact is highlighted by the divergence in the distributions of the different node measures. However, from the 5 plots, the restructuring of the company has led to the reduction of the node's importance. As a result, in the network structure following September 2023 nodes to a certain extent have fewer connections, and occupy fewer central positions in the network. This change might be explained by the new managerial framework. In fact, the issue of squads, which act independently and have a manager and a salesman at their head might explain this shift. Under the new managerial framework, individuals within the company are clustered in small groups and mainly connect with each other thus significantly impeding the communication flow and leading two nodes who are central locally.

	Average Shortest Path	**Density**	**Diameter**
Network Before September 2023	1.93	0.22	4
Network After September 2023	1.97	0.17	4

Table 1: Network Metrics Before and After September 2023

Table 1 displays a more holistic comparison between the two networks by comparing the average shortest path, density, and diameter of both structures. Looking at the table I do not notice a large difference. In fact, prior to September 2023, the network has a smaller shortest path and higher density. However, both networks have the same diameter value. As a result, the restructuring of the company seems to have reduced the overall number of realised connections, and slightly increased the distance between the coworkers.

This might slightly impede the flow of information and can be explained again by the new managerial framework which includes the formation of clusters represented by squads. These results are also linked to the insights I got from the node importance measures analysis. In fact, as the network has fewer central nodes and fewer hubs the path the average path needed to connect two nodes might be slightly higher and the number of links within the network should decrease thus explaining the rise in the average shortest path and the decrease in the network density.

6.2 Managers and salesmen importance

Following the shift in the managerial structure some people became managers of squads. In addition, for each squad, one salesman started having a central role in finding clients mainly for their own team. As a result, I found it worthwhile to first examine whether the restructuring had any impact on the various important measures for both salesmen and managers. Moreover, I checked whether managers and salesmen are more important on average than other individuals within the company.

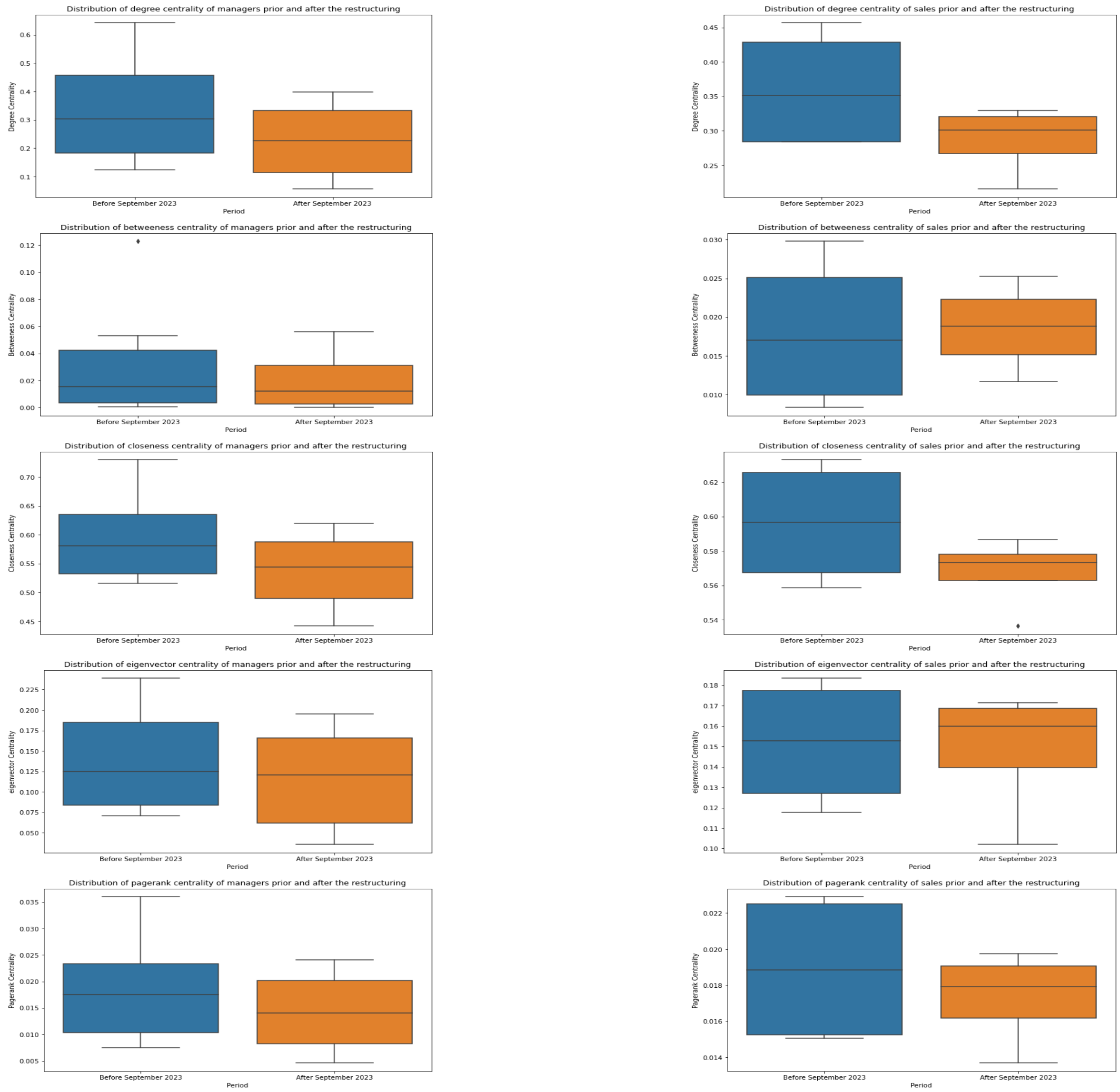


Figure 14: Distribution of node importance for managers and salesmen

Figure 14 displays the distribution of the node importance measures for managers and salesmen before and after the restructuring. The left plots are for managers and the right plots are for salesmen. I observed that following the restructuring the median of node importance as measured by degree, betweenness, closeness, eigenvector and PageRank centrality decreased for managers. However, for salesmen only the median of degree, closeness, and PageRank centrality decreased after the restructuring. As a result, after the structural shift managers became less central in the network. On the other hand, the impact of this change on salesmen was slightly different. In fact, salesmen seem to have fewer connections and are less connected with hubs in the network. In addition, they seem to act more as intermediaries and are on average farther from their coworkers. These insights can directly be extracted from the change in the median value of betweenness and closeness centrality, respectively. The divergence in the impact that the restructuring had on the managers and salesmen might be due to the fact that managers are central in their squad they mostly interact with individuals within their own team. On the other hand, while salespeople are also part of the squad they tend to have more broad interactions. In fact, if a salesman does not find a consultant suitable or available for a mission in his own team he/she can look for talents within other teams. Moreover, after September 2023 the results of the centrality analysis reveal that sellers occupy central positions in the network. The medians of degree, betweenness, closeness, PageRank, and eigenvector centrality are higher for salesmen compared to managers. This indicates that salesmen have more direct connection to other nodes in the network and that they play a crucial role as intermediaries, thus facilitating communication and information flow between different parties in the network. In addition, their higher closeness centrality emphasizes that they are closer to other nodes, allowing them to quickly disseminate information and interact effectively with the rest of the network. Moreover, they are more connected to nodes that are influential themselves.

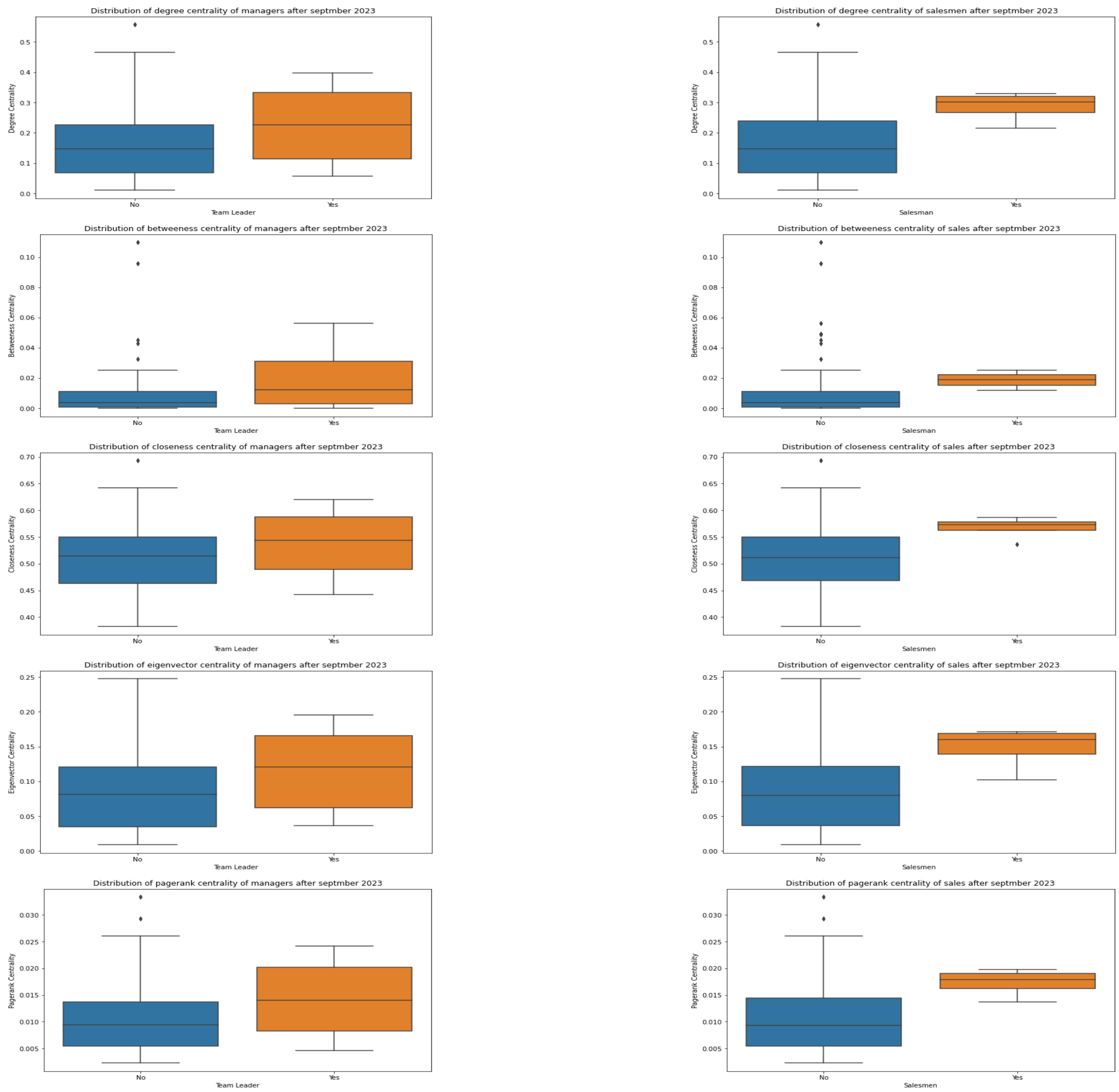


Figure 15: Distribution of node importance for managers and salesmen

However, while managers' and salesmen's importance, in a network sense, decreased as depicted in Figure 14 they are still more central than the other individuals within the company as illustrated by Figure 15. In fact, after September 2023, the median for salesmen and managers is higher compared to that of other coworkers. This interesting result might be explained by the fact that prior to the implementation of the new managerial structure salesmen and people who became afterward managers and are most likely the most experienced people in the company were highly solicited by a lot of their coworkers however in the current network they are mostly responsible of a subgroup of people within the company.

6.3 Gender importance

In this part, similar to the analysis I performed for managers and nonmanagers I will contrast the centrality of nodes representing males and females. In addition, I will compare the impact that the restructuring had on the node importance of both genders.

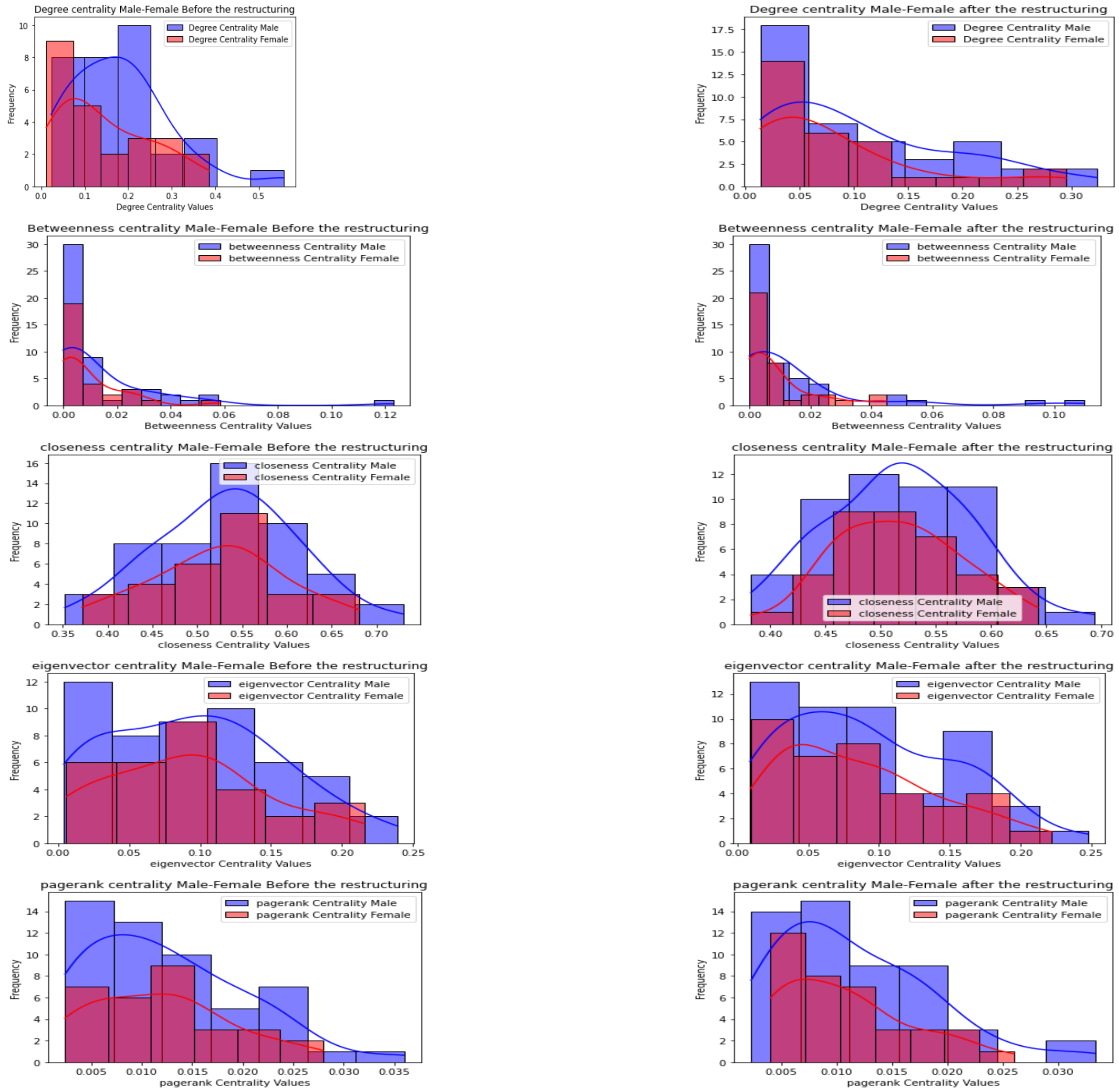


Figure 16: Distribution of node importance for Male and Female before and after restructuring

Figure 16 contrasts the distribution of node importance measures for males and females pre (plots to the left) and post (plots to the right) restructuring. The first thing I noticed is that males are more central than females in the network both before and after the restructuring. However, I notice that following the restructuring the gap in the distribution of degree centrality and betweenness centrality between males and females decreased.

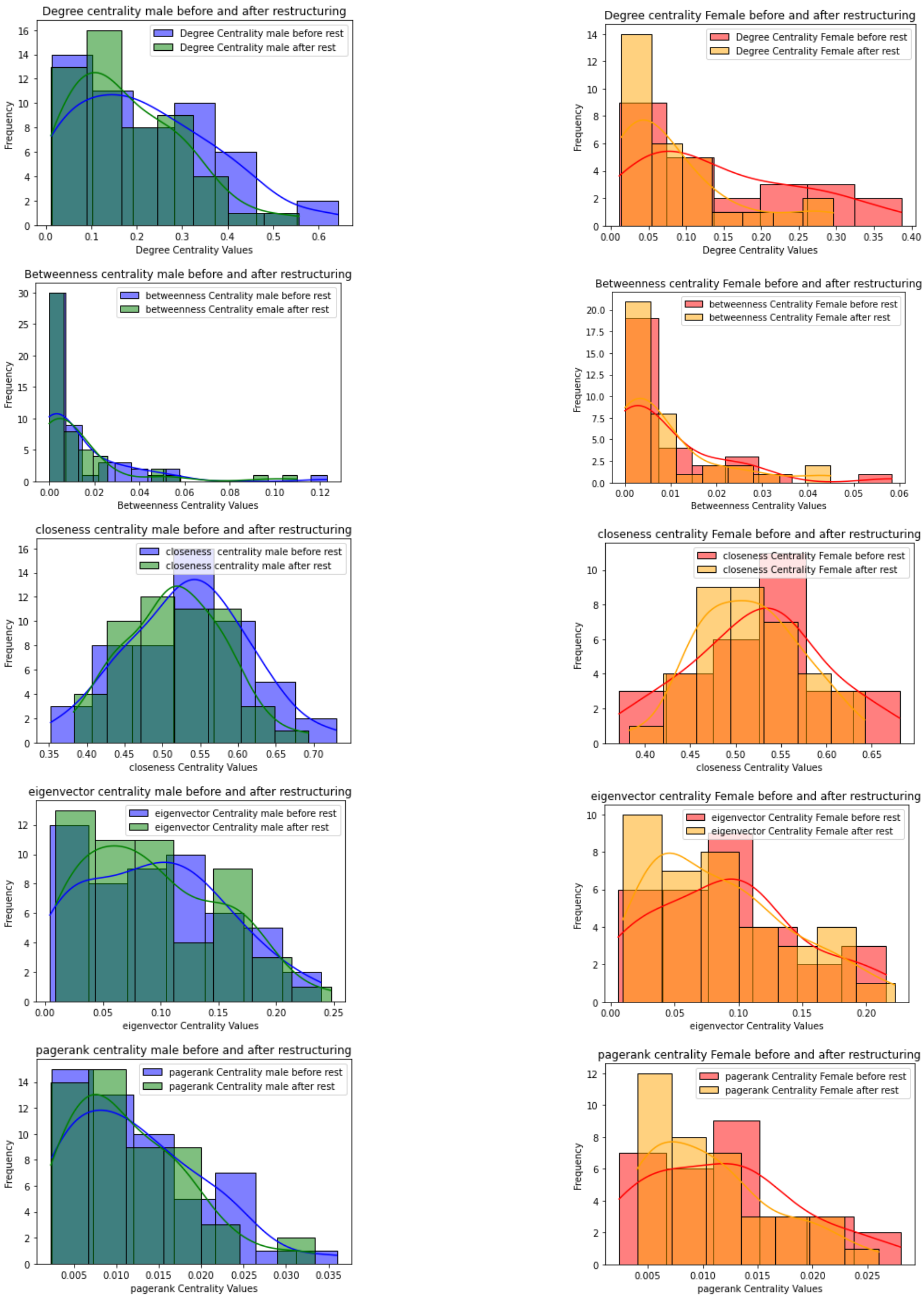


Figure 17: Impact of Restructuring on the Distribution of Nodes Importance

Figure 17 evaluates the impact of restructuring on the node importance for both males and females. The left column in the figure contrasts the node importance measures for males before and after the restructuring, while the right column presents the same comparison for females. Overall for both genders, I notice a slight decrease in node importance. looking at the distributions I observe a shift to the left of most of the distributions. In fact, the maximum value of the degree, betweenness, closeness, eigenvector, and PageRank centrality is higher before than after the restructuring. Moreover, there are more nodes with relatively high importance before September 2023 than after. As a result, the restructuring flattened the gap in importance between the different users by grouping individuals in teams as for both genders following the new managerial structure I observed more nodes with relatively low importance and centrality measures.

6.4 Forecasting the social structure in 2027

Actinvision is an ESN with roughly 80 employees that is growing rapidly. It is adopting an expansion strategy by implementing branches in several locations and attracting and retaining talent. The top management planned to reach a total of 300 employees in 2027. Given this information, I decided to start with the current network and expand its size by adding nodes and links artificially following some rules. To build this network I copied how networks are artificially built in the barbarasi-albert scale free model. Recall that such a model assumes that hubs have a higher probability of being connected to a node that is added to the network. The use of such a model is justified by the distribution of degrees in our network. In fact, as I see in Figure 18, the degree distribution is similar to a power law distribution with several nodes with a low number of connections and only a few nodes.

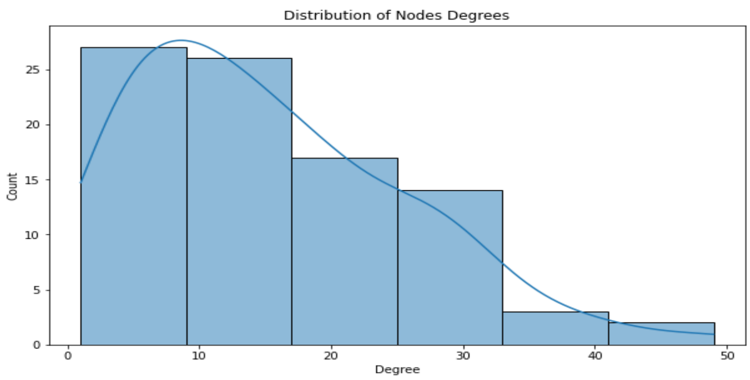


Figure 18: Power Law Distribution of Nodes' Degrees

Following the same process I artificially added nodes until reaching a network size of 300. For each new node, I generated an edge with another node with a probability that is proportional to the degree centrality of the existing node. For example, let's suppose a new employee A is hired. I first create a node representing individual A. At first, this node is completely disconnected from the network. Then I parse through the combinations of node A and all the other existing nodes i in the network. For each node i I compute $\pi(k_i)$ and generate a random number from a uniform distribution between 0 and 1. If the generated number is lower than $\pi(k_i)$ an edge is created between node A and i. This process is repeated until I reach a network with 300 nodes. Moreover, while new employees are hired new connections might also appear between the old nodes. So following the same process I artificially added links between the already existing nodes in the network. However in this case I did not rely on preferential attachment to measure the probability of a new connection being created. I relied on historical data to measure on a yearly basis the probability that a connection between two individuals taken at random is created as depicted in Figure 19.

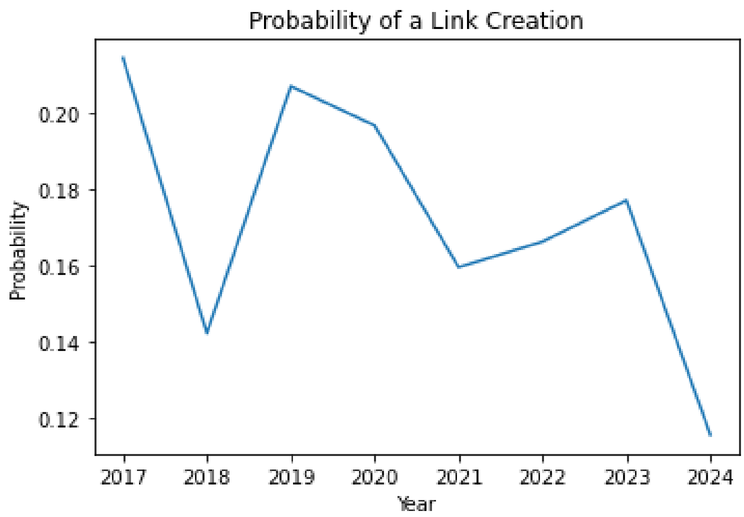


Figure 19: Probability of a Link Being Created per Year

Then I averaged over all the yearly probabilities and used the resulting value as a threshold that is used to evaluate in a similar process as the one described above whether a new connection should be issued between two existing nodes. In other words, when deciding whether a new link should be issued in year T between two existing nodes I used the threshold 0.17 to determine whether an edge should be created instead of using $\pi(k_i)$ Moreover, I also took into consideration the fact that on a yearly basis some employees might leave the company. For this, I used an arbitrary threshold of 10% as the probability of an individual quitting Actinvision. After forecasting the network of 2027 I measured the node importance measures and contrasted them with those of the current network. The results are depicted in Figure 20.

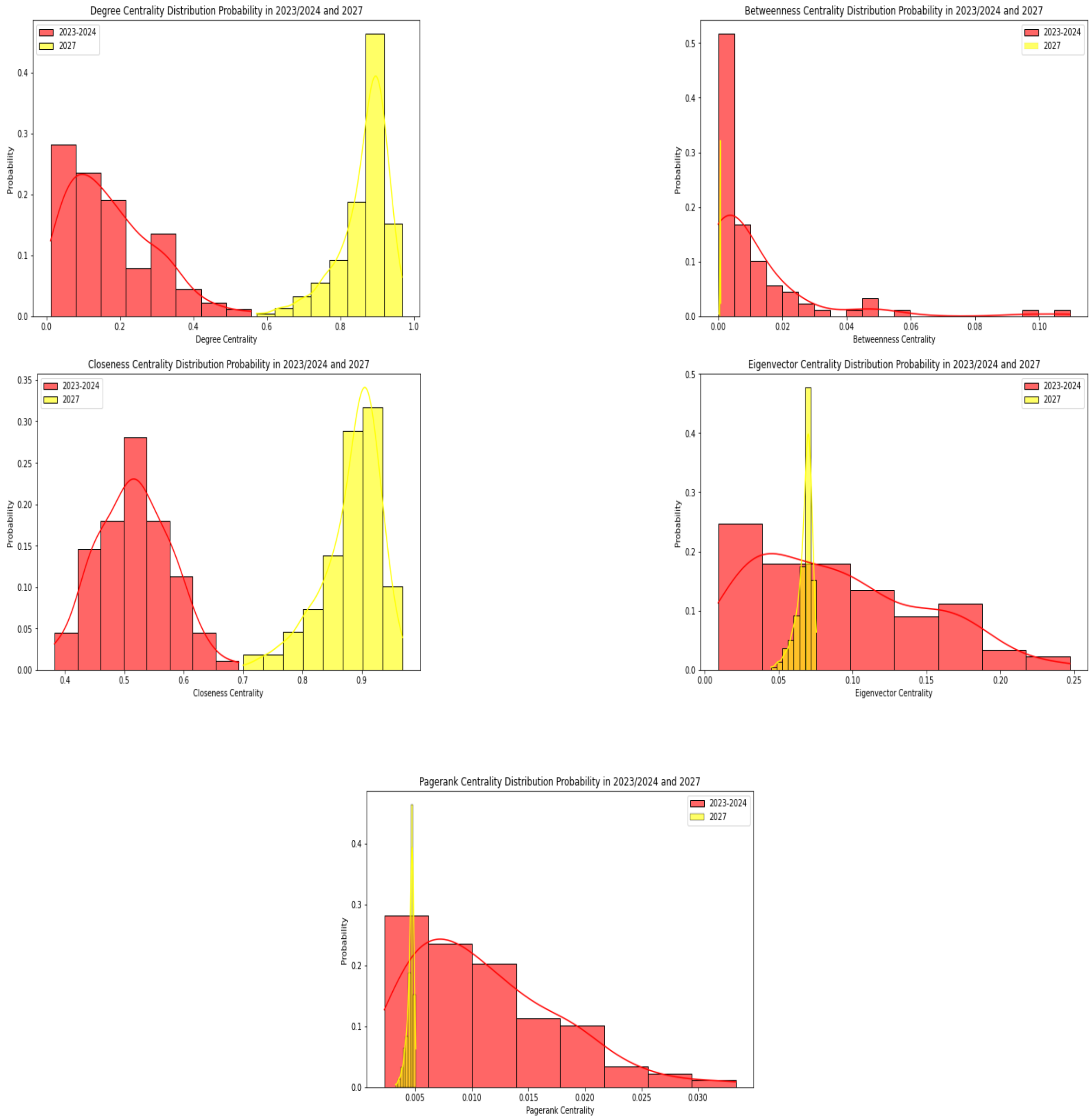


Figure 20: Distribution of Node importance in 2023-2024 and 2027

In Figure 20, the red bars indicate the node importance measures for the current network while the yellow ones illustrate those of the projected network. Interestingly, I notice that the current network has lower closeness and degree centrality while betweenness centrality remains unchanged. This indicates, that in the future, more nodes are expected to have a higher number of direct connections and that nodes will be more central and closer to each other. The invariability in the betweenness centrality indicates that the roles of individuals as bridges or brokers would remain consistent thus the critical paths through which information flows would not change significantly. On the other hand, the range of values for eigenvector and PageRank centrality for the current network is larger than the one of 2027. I might deduce from this that the old network structure has a greater diversity of connections among nodes with a larger gap between influential and non-influent nodes.

	Average Shortest Path	**Density**	**Diameter**
Network metrics in 2023-2024	1.97	0.17	4
Network metrics in 2027	1.14	0.86	2

Table 2: Network Metrics in 2023-2024 and 2027

Moreover as illustrated in Table 2 the future network would have a more efficient flow of information as on average, it will take fewer steps to connect any two nodes in the network in 2027 compared to 2023-2024. Moreover, the network in 2027 will be much more interconnected thus suggesting a more collaborative environment where nodes are more likely to be directly connected to one another.

6.5 Network's robustness

In this part, I will evaluate and compare the robustness of the three networks I generated throughout this project. To do this I will proceed in two ways. First I will remove nodes randomly. Second, I will proceed by performing a targeted attack by removing the nodes with the highest importance and centrality. Assessing the robustness of a network by removing nodes, whether randomly or targeting those with the highest centralities as measures with degree, betweenness, closeness, PageRank, and eigenvector measures, is important for several reasons. First, this evaluation helps determine the most crucial nodes whose loss might damage the network significantly. In addition, such analysis evaluates to what extent our networks are weak against high turnover rates.

Figure 21 illustrates the robustness of the network when nodes are removed randomly. To construct this plot I removed a node randomly and each time I measured the size of the biggest connected component (BCC). The biggest connected component is the sub-network that contains the largest number of nodes in which all nodes are reachable from each other. In our plots since the three networks I am going to compare do not share the same number of nodes I measured the BCC as a percentage of the whole network size. In Figure 21, I

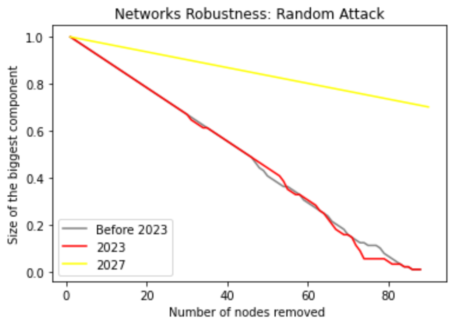


Figure 21: Networks Robustness: Random Attack

observe that when nodes are removed randomly networks before and after the restructuring are more impacted than the projected one in 2027. As I can see, after removing 40 nodes the two old networks have a BCC roughly equal to 50% of the whole network while that of the future network is above 80% of the whole network size. On the other hand, the company's restructuring doesn't seem to have a significant impact on the network's robustness when the nodes are removed randomly.

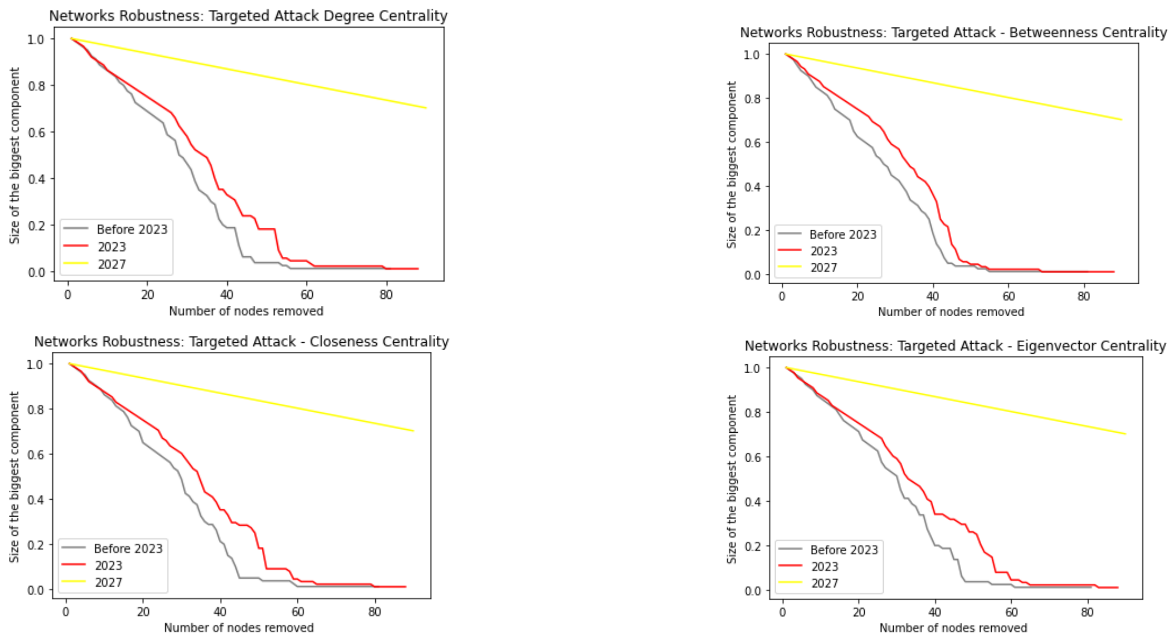


Figure 22 depicts the results when the networks are subject to targeted attacks, which means that each time I remove a node with the highest centrality. I observe that the network in 2027 is more robust as the BCC decreases steadily when more and more highly central nodes are removed compared to the other two structures. Moreover, I notice that in the case of targeted attack, the restructuring seems to increase the robustness of the network. In fact, I noticed lower values of BCC for the network prior to September 2023 compared to after the restructuring of the company. The divergence in robustness between these three

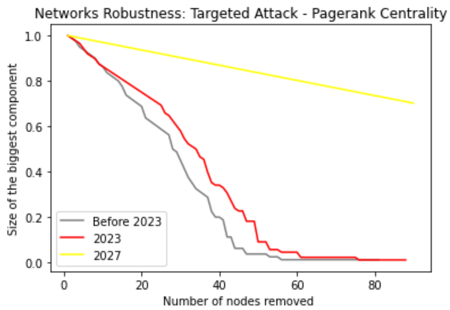


Figure 22: Networks Robustness: Targeted Attack

networks is linked to the results found previously. In fact, in 2027 the social structure of Actinvision is expected to have more central nodes and hubs, fewer disparities between the nodes' importance, and a high density. As a result, the removal of the most central node would not have a significant impact on the size of the largest component. Prior to September 2023, the network was characterized by more central nodes compared to the network following the restructuring. This might explain why the post-restructuring network is more robust to targeted attacks than the older one.

6.6 Challenges, Limitations, and Extentions

Since my lecture on network science at the University of Strasbourg I have always been interested in this field I really enjoyed working on this project as part of my master's thesis. However, working on this project wasn't exempt from challenges. In fact, as my thesis was a personal project that was not really needed by the company or linked to my daily tasks I had to personally allocate time in order to work on it. Moreover, at the university, I had a generic course about network science. Thus, while having a broad idea about the topic I did not have a full picture of what results and insights could be extracted when conducting an organizational network analysis. I had to read academic papers in order to gain a deeper understanding and effectively develop strategies for conducting a comprehensive organizational network analysis. Moreover, collecting the data was a challenging step of the project because the API Microsoft Teams was not really intuitive to use. In addition, it was really time-consuming to extract the whole data as some chat logs and messages dated back to 2017.

In this analysis through the use of network science, I was able to assess how a change in a managerial framework might impact the social structure, flow of information, and the main actors of a company.

In fact, looking at the change in the node importance measures before and after the restructur-

ing reveals several significant changes in the structure of connections between coworkers.
Centrality measures show that restructuring has overall reduced the importance of nodes within the network. More specifically, the results indicate a decrease in the number of connections and the centrality of nodes, which reflects a reorganization of communication flows within the company.
Before the restructuring, the network had a broader distribution of centralities, with nodes having relatively high. Besides, this new organization also had the effect of slightly increasing the average distance between employees, which could potentially slow down the flow of information within the company.
Moreover, looking at the impact that the restructuring had on key influencers such as managers and salesman as well as on node importance of males and females we notice that the company's network before and after restructuring reveals significant changes in the relative importance of managers, and salespeople. After the restructuring, although the importance of managers has decreased in terms of centrality, they nevertheless remain more central than other individuals in the company. Salesmen, meanwhile, have seen an increase in their relative importance, becoming more connected and playing a crucial role as intermediaries in the network. Regarding the gender distribution, although men remained more central than women, the centrality gap between the two genders narrowed after the restructuring.
In addition, the study of Actinvision's network, taking into account the planned expansion by 2027, reveals significant developments both in terms of the structure and robustness of the network. By simulating the addition of new employees and new connections according to the Barabási-Albert model, I found that the projected 2027 network is expected to be more connected, with more central nodes and more efficient information flow. This setup suggests a more collaborative work environment, where employees are closer to each other in terms of connections.
However, the robustness of the network was also evaluated against random node removal and targeted attacks on the most central nodes. It appears that the future network is more resilient, particularly in the face of targeted attacks, thanks to a better distribution of centralities and a higher density. On the other hand, the 2023 restructuring has already made it possible to strengthen the robustness of the network compared to its previous configuration, by reducing dependence on certain central nodes.
While the study was able to unveil some interesting results our study is not flawless. In fact, an important part of our analysis revolved around contrasting the network structure before

and after the restructuring. However, the amount of data retrieved in both periods is largely imbalanced. In fact, prior to the data prior to the restructuring included data going up to 2017, while data after the restructuring included data from only 9 months (From September 2023 to May 2024). As the managerial change has been recently introduced only part of its impact could be captured. In addition, in this analysis, I forecasted the state of the network in 2027. However, by doing this I kept things simple and disregarded some factors that might highly influence the projected network structure. First, I assumed that the same connection patterns would prevail and that no external impact like a crisis or another managerial restructuring would impact our network. In addition, when computing the probabilities of nodes being created I didn't take into consideration that individuals belonging to the same team would have a higher probability of connecting.

As a result, certain enhancements could be implemented to refine and improve the outcomes of this analysis. First, I can redo the analysis with more data post-restructuring. Moreover, probabilities can be recomputed to take squad membership into consideration. Taking into consideration structural shifts when building the future network is hard as I do not have enough visibility on what might happen in the next 3 years. Thus, we should keep in mind that this projected network is just a proxy of how the social structure within the company would look if everything is kept constant.

Finally, the different implementations in this study are not the only ones that can be done in the context of an organizational network analysis (ONA). In fact, some patterns present in the network structure might determine which employees might come up with innovative ideas, have more influence on their coworkers, and which teams are more efficient, and innovative [28]. Moreover, I mainly used simple and widely used node importance measures that focus on a single criterion to evaluate the centrality. However, Some studies have developed weighted measures that have been proven to be effective in ranking nodes by importance in complex networks [35]. Thus, a possible extension would be to use such a measure to evaluate the importance of nodes in the network. Besides, another possible extension would be to contrast the degree of connectivity within and between squads. This can help identify which squads demonstrate the highest level of teamwork and which individuals within those squads collaborate most effectively with one another. Finally, our analysis did not take advantage of the longitudinal data I have. In fact, I can leverage Actinvision's historical data to analyze the evolution of the number of messages sent, also called the amplitude of communication, and whether a pattern exists for individuals who left the company [22] or whether factors like gender, age, background influence the speed at which nodes connect to the network. Another possible extension would be to model weighted or bipartite networks. In the former case, the weights could represent the number of messages sent between individuals. In the latter case, we could evaluate the interactions between individuals and the company's clients

7 Conclusion

To sum up, my experience within Actinvision was full of challenges through which I have significantly improved not only my soft but also my hard skills. I have learned to cooperate, work under pressure and time constraints, and quickly evolve in the use of specific tools in order to be productive and efficient as soon as possible.
I have also had the opportunity to use and hone some of the hard skills I have learned at the University of Strasbourg like SQL and Python. In addition, thanks to the company's flexibility in encouraging internal projects I was able to work independently on a data science project in the field of network science that I have been introduced to at the University of Strasbourg. The code used to implement the organizational network science analysis can be accessed on the Github repository via the following link.

References

[1] Luis C.S. Afonso, Alexandre Xavier Falcão, and João Paulo Papa. Chapter 2 - theoretical background and related works. In Alexandre Xavier Falcão and João Paulo Papa, editors, *Optimum-Path Forest*, pages 5–54. Academic Press, 2022.

[2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *nature*, 401(6749):130–131, 1999.

[3] Mary Attenborough. 19 - graph theory. In Mary Attenborough, editor, *Mathematics for Electrical Engineering and Computing*, pages 461–478. Newnes, Oxford, 2003.

[4] Seyed Mojtaba Hosseini Bamakan, Ildar Nurgaliev, and Qiang Qu. Opinion leader detection: A methodological review. *Expert Systems with Applications*, 115:200–222, 2019.

[5] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[6] Phillip Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29:555–564, 10 2007.

[7] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[8] Katy Börner, Soma Sanyal, Alessandro Vespignani, et al. Network science. *Annu. rev. inf. sci. technol.*, 41(1):537–607, 2007.

[9] Ulrik Brandes. A faster algorithm for betweenness centrality*. *The Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[10] Ulrik Brandes, Garry Robins, Ann McCranie, and Stanley Wasserman. What is network science? *Network science*, 1(1):1–15, 2013.

[11] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003.

[12] A.C.C. Coolen, A. Annibale, and E.S. Roberts. 207Applications of random graphs. In *Generating Random Networks and Graphs*. Oxford University Press, 03 2017.

[13] Rob Cross, Tim Laseter, Andrew Parker, and Guillermo Velasquez. Assessing and improving communities of practice with organizational network analysis. *The Network Roundtable at the University of Virginia*, pages 2–22, 2004.

[14] Michael DiClaudio. People analytics and the rise of hr: how data, analytics and emerging technology can transform human resources (hr) into a profit center. *Strategic HR Review*, 18(2):42–46, 2019.

[15] Paul Erd6s and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.

[16] P ERDdS and A R&wi. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.

[17] Swati Goswami, Asit K. Das, and Subhas C. Nandy. Sparsity of weighted networks: Measures and applications. *Information Sciences*, 577:557–578, 2021.

[18] Shlomo Havlin, Dror Y Kenett, Eshel Ben-Jacob, Armin Bunde, Reuven Cohen, H Hermann, JW Kantelhardt, J Kertész, S Kirkpatrick, Jürgen Kurths, et al. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *The European Physical Journal Special Topics*, 214:273–293, 2012.

[19] Sarah Barriage Sam Abramovich Laurie Bonnicci Youngok Choi Heting Chu Yunfei Du Hong Huang Lili Luo Valerie Nesset Heather L O'Brien Soo Young Rieh Laura Saunders Sei-Ching Joanna Sin Luanne S. Sinnamon Amy VanScoy Yan Zhang June Abbas, Caryn Anderson. *Library Information Science Research*. Elsevier, Oklahoma, 2024.

[20] Michael Kaufmann and Katharina Zweig. Modeling and designing real–world networks. In *Modeling and Designing Real–World Networks*, pages 359–379, 06 2009.

[21] Dror Kenett and Shlomo Havlin. Network science: A useful tool in economics and finance. *Mind Society*, 14, 05 2015.

[22] Abdel-Rahmen Korichi, Hamamache Kheddouci, and Taha Tehseen. *Modeling and Analysis of Organizational Network Analysis Graphs Based on Employee Data*, pages 354–367. 05 2023.

[23] Amy N. Langville and Carl D. Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.

[24] Ted G Lewis. *Network science: Theory and applications*. John Wiley & Sons, 2011.

[25] Sorin Matei. Analyzing social media networks with nodexl: Insights from a connected world by derek hansen, ben shneiderman, and marc a. smith. *Int. J. Hum. Comput. Interaction*, 27:405–408, 02 2011.

[26] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[27] Roland Molontay and Marcell Nagy. Two decades of network science: as seen through the co-authorship network of network scientists. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 578–583, 2019.

[28] Talha Oz. Improving company performance with organizational network analysis (ona). Technical report, Humanyze, 2018.

[29] Romualdo Pastor-Satorras and Claudio Castellano. The localization of non-backtracking centrality in networks and its physical consequences. *Scientific reports*, 10(1):21639, 2020.

[30] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[31] Corey Phelps, Ralph Heidl, and Anu Wadhwa. Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, 38:1115–1166, 07 2012.

[32] Akrati Saxena, Ralucca Gera, and SRS Iyengar. A faster method to estimate closeness centrality ranking. *arXiv preprint arXiv:1706.02083*, 2017.

[33] Leo Spizzirri. Justification and application of eigenvector centrality. *Algebra in Geography: Eigenvectors of Network*, 2011.

[34] Onur Tanglay, Nicholas Dadario, Elizabeth Chong, Si Tang, Isabella Young, and Michael Sughrue. Graph theory measures and their application to neurosurgical eloquence. *Cancers*, 15:556, 01 2023.

[35] Yuanzhi Yang, Lei Yu, Xing Wang, Zhongliang Zhou, You Chen, and Tian Kou. A novel method to evaluate node importance in complex networks. *Physica A: Statistical Mechanics and its Applications*, 526:121118, 2019.

[36] Yuanzhi Yang, Lei Yu, Zhongliang Zhou, You Chen, and Tian Kou. Node importance ranking in complex networks based on multicriteria decision making. *Mathematical Problems in Engineering*, 2019:1–12, 01 2019.

[37] Jie Zhang, Xiao-Ke Xu, Ping Li, Kai Zhang, and Michael Small. Node importance for dynamical process on networks: A multiscale characterization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):016107, 03 2011.