Organizational Network Analysis (ONA)

----------------------------Mayur G. Dushetwar---------------------------

Abstract

Mayur Inc. (fictional name) is a an automobile manufacturers headquartered in US. The company is also expanded in UK, France, Germany, and Brazil. Recently, due to COVID outbreak, many employees are required to work from home. Amid recent challenging time, the fear of loss in productivity and organisational knowledge is more apparent than ever. In order to avoid knowledge loss and loss in productivity due to virtual communication, top management along with Technical Communication department are keen to understand the current communication network and want to know who are the influencers in the organisation. The insights and recommendations will be used to improve the networking in the organisation. The project uses the Python programming language to perform exploratory data analysis and hypothesis testing to extract valuable insights and results using the network analysis.

Introduction:

Working from home prior to the COVID 19 outbreak had been a luxury. However, organizational environment has now changed. Thanks to the continuous advancement in the field of digitalization, the shift to a more knowledge-based economy have made the option of remote work and hybrid work more realistic for employers and employees. Working from home has become a decisive element for choosing the employment (Justina Alexandra Sava, 2022). One of the top challenges that current organizations face includes alignment of organizational goals, monitoring and control, and cultivating shared values (Sen, Deb, & Kumar, 2021). As a result, analyzing organizational network and identifying influencing factors is one of the top priorities for the organizations.

What is networking? Some scholars view networking behaviors that focus on building relationships that will help individuals advance in their careers. Gould and Penley (1984) provided one of the first empirical definitions of networking where they described networking as "the practice of developing a system or 'network' of contacts inside and/or outside the organization, thereby provided relevant career information and support for the individual" (Gibson, Hardy III, & Buckley, 2014). Organizational social network research has achieved a prominent position. When we talk about networking through mobile, more than half of population is already connected through mobile devices and around 40% population is within the footprints of mobile broadband but do not use it (GSMA, 2022). However, when we talk about social networks in the organizations, it is more complex and sometimes, beyond networking between people or people and parts which usually benefits from the entity. Thus, understanding organizational network becomes critical in designing effective organizational frameworks (McDowell, Horn, & Witkowski, 2016). Researchers started exploring the field of network analysis around 1970s. The general shift from individualists understanding towards more relational, contextual, and systemic understandings was peeked in the second half of the

20th century. If we look at graph of the publications containing phrase 'social network' in the abstract or title, it has risen exponentially from 1970 to 2000 (Borgatt & Foster, 2003). Since then, there has been many new trends evolved in the field such as merging social networks, understanding cognitive and personality perspectives, and analyzing large networks configuration.

The project is focused on understanding the organizational network of given datasets from the organization. While the paper explains the organizational network, it also parallelly demonstrates practical application on a network analysis on the datasets of the organization (source and credit: datacamp.com) using python. The dataset includes information of senders and receivers. For privacy reasons, name of employees kept anonymous. Based on the results and insights obtained from the analysis, the paper also provides recommendations to HR and technical communication departments in order to improve collaboration among the employees.

Structure of the Paper:

In the first section, the paper defines important terminologies and concepts used in an organizational social network. Once the network graphs are created, it is crucial to understand the influencers in the organization which will be targeted to bring the changes or improvement in the overall organizational communication framework. Second section is dedicated to understanding the centrality concept and finding different forms of centralities. The section also covers the key roles in an organizational network. In final section, the paper ends with possible recommendations in order improve collaborations among the employees.

SECTION I: ORGANIZATIONAL NETWORK

The term 'Organizational network' appears to be self-explanatory. Nonetheless, if one needs to define an organizational network, what would be the answer? The organizational network is an extended concept of social network analysis. The organizational network is not always limited to people. It takes the form of resources, tasks, knowledge, and many other elements which may constitute network nodes (Ujwary-Gil, 2016). In its simplest form, an organizational network is a group of three or more elements or actors that decide to collaborate, share resources, and otherwise work together (Derr, 2021). Using the Neworkx library for python, the network diagram is formed using two main components i.e., nodes and edges. The node is a human being, who is also called an agent or actor in the network. The line joining two nodes is called an edge. Fig 1.1 shows a general structure of the network using nodes and edges.

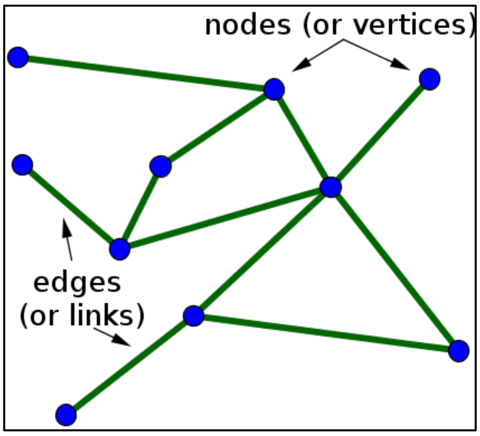

Fig 1.1: Representation of network with nodes and edges

The project is conducted using python programming. Python provides several handy and easy-to-use libraries to deal with complex topics. One such library to map the organizational network is the Networkx. The library is imported, and further analysis is carried out using functions and methods provided in it.

The data is collected randomly from 5 different locations namely, the US, the UK, Brazil, France, and Germany. Employees are selected from the Sales, Operations, Admin, Engineering, Marketing, and IT departments. The dataset includes the employee IDs and ages of the sender and receivers. Additionally, the lengths of the messages are mentioned for each observation along with the timestamp when the message is sent.

	sender	receiver	timestamp	message_length
0	79	48	2021-06-02 05:41:34	88
1	79	63	2021-06-02 05:42:15	72
2	79	58	2021-06-02 05:44:24	86
3	79	70	2021-06-02 05:49:07	26
4	79	109	2021-06-02 19:51:47	73

Fig 1.2: Dataset 1

	id	department	location	age
0	3	Operations	US	33
1	6	Sales	UK	50
2	8	IT	Brasil	54
3	9	Admin	UK	32
4	12	Operations	Brasil	51

Fig. 1.3: Dataset 2

The graphs are plotted each for network among the employees and network among departments. The aim of the project is to visualize the components of the organizational networks and point out the main components or influencers in the organization. Different forms of centralities are used to determine the key roles in the network which will be covered in next section. With the first look from exploratory data analysis, general idea about the activeness of the department is obtained. This can be easily understood from the following graphs.

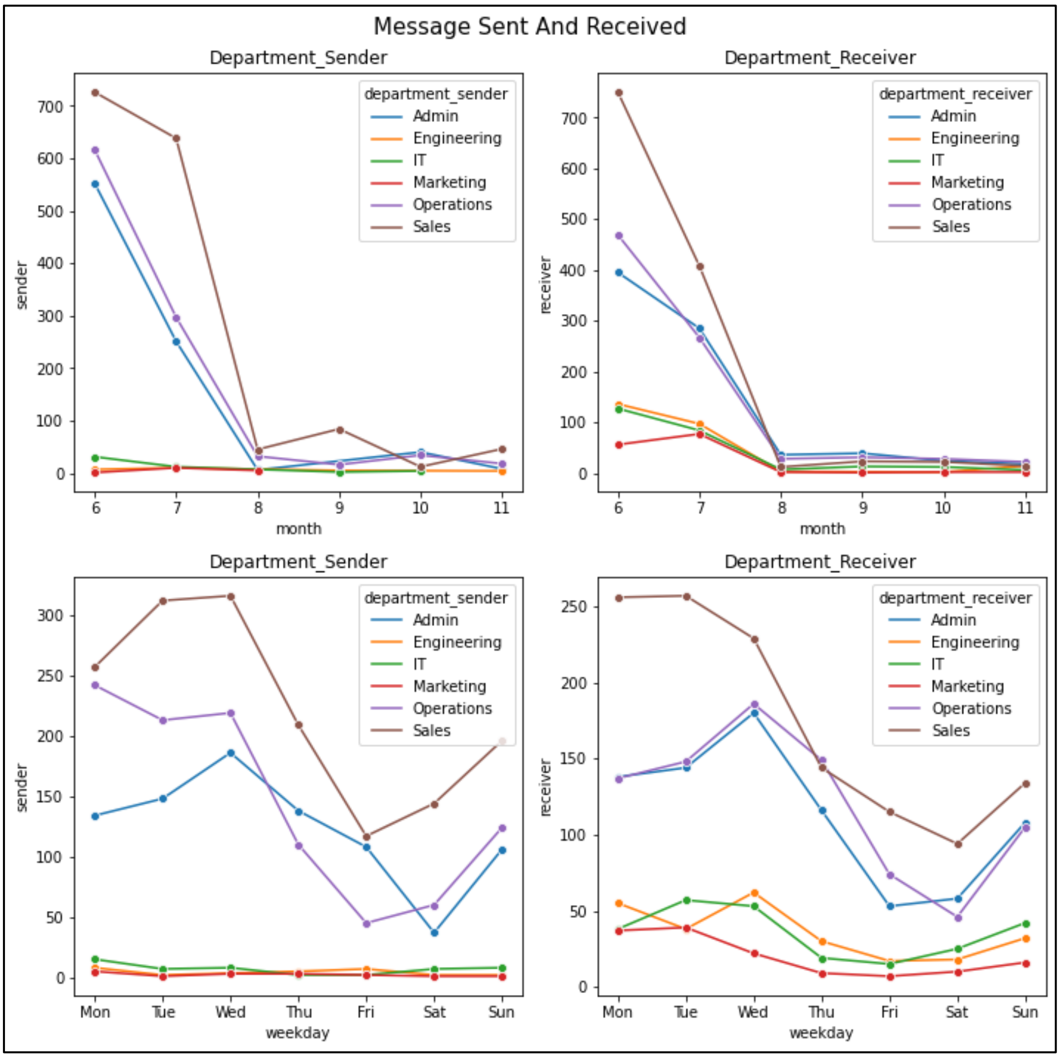

Fig 1.4: Messages sent and received by different departments

The above plots clearly indicates that Sales, Operations and Admin departments are more active than Marketing, Engineering, and IT departments. Although, it seems that Sales department is the most active one, overlapping of trend lines make it somewhat ambiguous and hard to confirm which department is the most and least active among all. To make sure what we see in the graphs is 100% in-line with the facts, well known method of hypothesis testing is utilized. Since, the proportions of messages sent and received by each department is to be checked, the analysis is carried out using t-proportion test on the sample.

The results of hypothesis testing supports that there is enough evidence to say that **Sales department is the most active** among all. When the bottom tier departments are analyzed, the **Marketing and Engineering departments appeared to be the least active**. However, there is not enough evidence to say one is lesser active than other.

Once the general idea about activeness in the different departments is acquired, the project moves ahead with the core part of the analysis which is drawing the graphs and analyzing key components and influencers in the communication network. Please note that the graphs are directed graph to keep tab on two-way communication.

```
# Graph for connection among employees--------------------------
CG =nx.from_pandas_edgelist(comm, source="sender", target="receiver",
                            edge_attr=True , create_using=nx.DiGraph())
# Graph for connection among departments------------------------
DG= nx.from_pandas_edgelist(comm, source= "department_sender",
                            target = "department_receiver",
                            edge_attr=True,
                            create_using=nx.DiGraph())
```

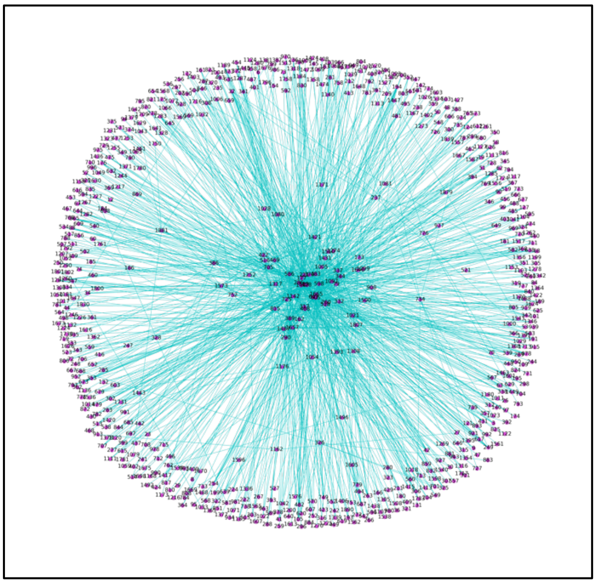

Fig. 1.5: Employee Communication Network

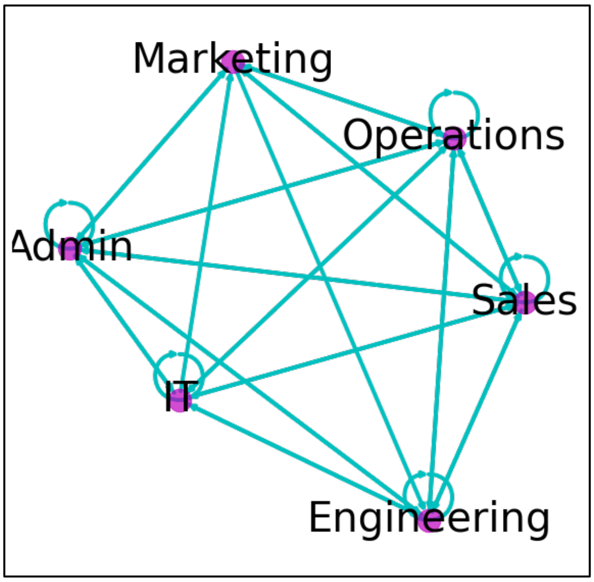

Fig. 1.6: Department Communication Network

Employee communication network vividly indicates that the number of recipients who are lying in the peripheral area of the network is significantly more than those of senders who sit in the middle of the network. After further digging, the number is approximately to be 86% more than that of senders. For departments, it is simply that each department is connected to the remaining 5 departments under observation. Once these graphs are plotted, it encourages finding the key elements and influencers in the given communication framework. It is covered in the following section.

SECTIONS II: CETNTRALITIES AND KEY ROLES

Visualizing and analyzing formal and informal relationships in the network help organizations shape business strategy that maximizes the organic exchange of information, thereby helping the business become more sustainable and effective. All organizations have people as nodes in their respective communication network who serves as main conduits for knowledge transfers. A connection between two nodes delivers value when needed information is exchanged (McDowell, Horn, & Witkowski, 2016). Fig.2.0 shows sample actors in the organizational network.

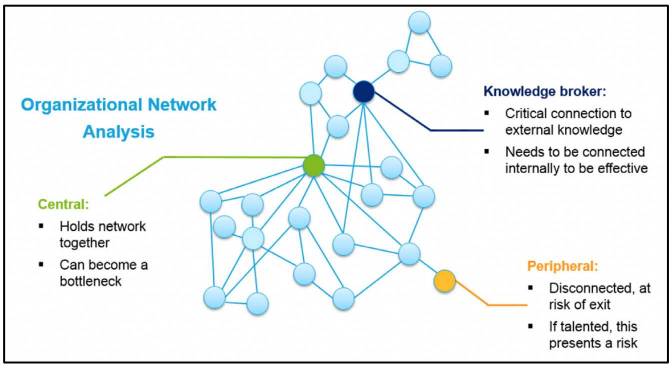

Fig. 2.1: Sample actors in the organizational network (Source: Deloitte)

To answer the question who the key element in the network is or define the role of an influencer require to define centrality metric of the network. Four metrics are mainly used in the network analysis which are Degree, Closeness, Betweenness and Eigenvector centralities. Let's check the meaning of each.

2.1 Degree Centrality:

The degree of a node is defined as the number of connecting edges that it has. In the case of a directed network where edges have direction, degree centrality is normally measured by indegree and outdegree. It can also be interpreted as number of edges associated with the node. In terms of communication network, more the number of connections, more the degree centrality. It is representation of popularity of the element (Wang, 2019).

2.2 Closeness Centrality:

It can be informally thought as 'average distance' to all other nodes. The more central a node is, the closer it is to all other nodes. Degree centrality measures might be criticized because they only consider the immediate ties that an actor has, or the ties of the actor's neighbors, rather than indirect ties to all others. One actor might be tied to many others, but those others might be rather disconnected from the network. In a case like this, the actor could be quite central, but only in a local neighborhood (Hanneman & Riddle).

2.3 Betweenness Centrality:

Betweenness centrality quantifies the number of times a node acts as a bridge (or "broker") along the shortest path between two other nodes. Suppose that I want to buy a new flat in Pune. But, in order to contact the owner, I must go through a Broker. Imagine that I want to get the flat at the discounted price. Here firstly, I must forward my request to the broker. Broker could delay my request or influence it in some way. This gives the broker who lie "between" me and the flat owner the power with respect to me. Having more than one channel makes me less dependent, and, in a sense, more powerful.

2.4 Eigenvector Centrality:

A node is high in eigenvector centrality if it is connected to many other nodes who are themselves well connected. Suppose I have good number of connections on LinkedIn and my connections also have good amount connections, then my eigenvector centrality value is higher. Intuitively, it considers not just "how many people you know," but also "who you know."

Roles in Networking:

Using the idea of centrality, many researchers suggest the various roles for a given organizational network. Researchers such as McDowell, Horn, & Witkowski in 2016, Wang in 2019 and Ezequiel Ortiz Recalde in 2022 define different roles in their articles or papers. Recalde mentioned several roles such as key player for information transfer, essential motivators, emotional supporter or technical referants. Wang jotted down roles based on popularity, influence, centralness and bridge. The i4cp Team defines boundary spanners (who sit in the white spaces between groups or units that would otherwise not be connected), central connectors (who are crucial to performance and yet risk overload and burnout), and energizers (who generate enthusiasm and a sense of purpose amongst their networks).

The aim of the paper is to find who are the influencers in the organisation at both employee and department level. Each centrality value signifies certain form of influence. In addition, each person is different from another in some way. It is rare to find a person transcend all skills in communication networking. Therefore, in place of using weighted average method, the employees are categorised into different categories of influence which is stated as below.

Proactive: These members are widely popular in organizations because of their extroverted nature. Statistically, they have a higher degree of centrality. They may act as key players in information transfer due to their networking skills.

Central: These members are with higher closeness centrality. These members are closer to other members and easy to reach out to. They can be emotional supporters for some people. They are usually the ones with higher closeness centrality value

Knowledge Brokers: These members are intermediatory between the groups. It quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. A bridge in a social network is someone who connects two different social groups. They fill the gap in information transfer between the two groups. Knowledge brokers will have a higher betweenness centrality value

Instrumental: Instrumental influencers are recognized as the ones with higher eigenvector centrality value. They are influential not only because of higher connections but also due to richness of their connections. Since these people have popular and powerful connection, they contribute to the decision-making process.

Now, we have already defined what are the categories for key players in the organizational network. Let's now define who are these key players. For departments, we will use the weighted average method to decide which department influences the most. The department is said to be a networking bellwether in the organization when it works closely, promotes maximum collaboration, and connects other critical departments.

Based on these criteria, we will assign weighing factors to a degree, closeness, betweenness, and eigenvector centralities and calculate the mean value for each department. We will assume that organization provided the preferences as weight factors in the following manner:

- **degree--> 40%**
- **closeness--> 30%**
- **betweenness--> 20%**
- **eigenvector--> 10%**

Calculating values for each centrality for both employees and departments separately is cumbersome. We will make it easier by creating a helper function centrality_dataframe(). It will take the graph as an argument and returns the new data frame containing the values for different centralities. The function also scales the values of each centrality using the maximum value of the respective centrality.

```
def centrality_dataframe(Graph):
    '''the function returns the new dataframe containing different centrality values for
each observation'''

    # Degree-----------------------------------------------------------------
    connections = dict(Graph.degree)
    # Degree centrality value------------------------------------------------
    degree_emp = nx.degree_centrality(Graph)
    # closeness centrality value---------------------------------------------
    closeness_emp = nx.closeness_centrality(Graph)
    # betweeness centrality value--------------------------------------------
    betweeness_emp = nx.betweenness_centrality(Graph)
    # Eigenvector Centrality value-------------------------------------------
    eigen_emp = nx.eigenvector_centrality(Graph)

    # Normalising the values-------------------------------------------------
    max_val_deg = max(degree_emp.items(), key=lambda x: x[1])
    max_val_close = max(closeness_emp.items(), key=lambda x: x[1])
    max_val_between = max(betweeness_emp.items(), key=lambda x: x[1])
    max_val_eigen = max(eigen_emp.items(), key=lambda x: x[1])

    # Creating empty dictionaries
    deg_scaled ={}
    close_scaled ={}
    between_scaled={}
    eigen_scaled ={}

    for key in eigen_emp.keys():
        deg_scaled[key] = degree_emp[key]/max_val_deg[1]
        close_scaled[key] = closeness_emp[key]/max_val_close[1]
        between_scaled[key] = betweeness_emp[key]/max_val_between[1]
        eigen_scaled[key] = eigen_emp[key]/max_val_eigen[1]


    # Creating dataframe of number of connections------------------------------
    df=pd.DataFrame.from_dict(connections.items())
    df.columns=["Employee_id", "Connections"]
    # Creating dataframe of degree centrality---------------------------------
    df1=pd.DataFrame.from_dict(deg_scaled.items())
    df1.columns=["Employee_id", "Degree"]
    # Creating dataframe of closeness centrality------------------------------
    df2=pd.DataFrame.from_dict(close_scaled.items())
    df2.columns=["Employee_id", "Closeness"]
```

```
    # Creating dataframe of Betweeness centrality-------------------------------
    df3=pd.DataFrame.from_dict(between_scaled.items())
    df3.columns=["Employee_id", "Betweenness"]
    # Creating dataframe of Eigenvector centrality------------------------------
    df4=pd.DataFrame.from_dict(eigen_scaled.items())
    df4.columns=["Employee_id", "Eigenvector"]


    # Merging all the dataframes------------------------------------------------
    data = [df1,df2,df3,df4]

    for dataframe in data:
        df = df.merge(dataframe, on="Employee_id")

    return df
```

The output of this code is a new data frame containing different centrality values for each observation. Please note that the values are scaled concerning the maximum value of respective centrality. The results are shown in fig 2.1. The resulting dataframe is sorted for each centrality type and employees with top values in each category are selected as influencers.

	Employee_id	Connections	Degree	Closeness	Betweenness	Eigenvector
0	79	12	0.142857	0.513728	0.040852	2.601756e-01
1	48	2	0.023810	0.462963	0.000000	9.454764e-02
2	63	4	0.047619	0.694143	0.000000	3.509053e-01
3	58	1	0.011905	0.418871	0.000000	8.693082e-02
4	70	1	0.011905	0.418871	0.000000	8.693082e-02

Fig 2.2: Data frame created using helper function

When we sort values for each centrality in descending order to get the top candidates for each centrality. These candidates are then selected as the influencers in the respective categories. The datasets are attached in the appendix. The results are summarized as below:

- An employee with ID **598** from the Operations department has the highest connections. He/she is selected as the 'Proactive' influencer
- An employee with ID **32** from the Sales department has a closer working network than any other employee. He/she is selected as the 'Central' influencer
- An employee with ID **509** from the Admin department has the highest degree of betweenness. It means he/she acts as a medium between more groups by joining and connecting them. He/she is selected as a Knowledge Broker.
- An employee with ID **194** from the Admin department has a richer network than others. In other words, he/she is connected to people who also have a popular and powerful network. Therefore, Employee with ID 194 is selected as an 'Instrumental' influencer

Similarly, the centrality values for departments are created dataframe using the helper function and calculating the weighted average for the centrality values. Following are the main findings:

- Sales, Operations and Admin departments have employees who are more connected, working closely, connecting different groups, and having rich networks. All three departments have a weighted average of 1. Hence, these departments are the Influencers in the organization. It bolsters our previous findings.

- IT and Engineering departments have significant connections with higher eigen values however moderate betweenness. Nonetheless, weighted average of Engineering is more than IT department as former is working more closely and have richer connections.

- The Marketing department has the lowest weighted average. It is lowest in all the categories. While it shows moderate centralities in degree, closeness, and eigenvector, it has absolutely no power in bridging any two departments.

The output of the helper function is attached in the appendix.

SECTION III: RECOMMENDATION

The above section was dedicated to identifying where the centrality of the organization lies. Based on this, the influencers were identified in the organization. One could have valid questions about how we can improve collaboration with the help of these influencers. To answer the question, let’s first understand how the organization can improve collaboration in general. The flow of information is very critical for collaborative efforts. It increases transparency which as result improves employee engagement. Harvard conducted research into the team behavior of 15 multinational companies. The research shows that team members collaborate more easily and naturally if they perceive themselves as being alike. The study found that the higher the proportion of strangers on the team and the greater the diversity of background and experience, the less likely the team members are to share knowledge or exhibit other collaborative behaviors. The investment in face-to-face interaction creates many opportunities for people across the company to see the top executives in action (Gratton & Erickson, 2007).

The influencers are the nodes connecting and motivating the employees in the organizations. HR and communications departments can work with the influencers to improve the information flow by spreading messages through key nodes. While we focused on the influencers, the network analysis helps to identify the peripheral or isolated nodes(employees). The departments can pay close attention to these employees and involve them in networking activities.

The marketing and Engineering departments are identified to be the least active departments in the organization. The HR department needs to focus on these departments and find the reason for their less involvement in networking activities. Since these departments principally work on technical and production areas, they are lesser involved in communication activities than the management and admin departments. However, actions need to be taken to streamline the information flow in and out of these departments which in turn, needs the active engagement of employees. This can be achieved with the help of influencers in the organization. While onboarding new employees, the HR department could pair new employees with the influencers or other key nodes so that new employees get acquainted with the different teams easily with more transparent information flow.

Summary:

Sales department is the most active department whereas **Marketing and Engineering** departments are the least active in the organisation. The sample does not have enough evidence to compare Marketing and Engineering departments as we fail to reject the null hypothesis stating the proportion of two department is equal. The employees with IDs **598, 32, 509 and 194** are collectively selected as the influencers as they possess different attributes such as proactiveness, closeness, acting as a bridge and instrumental networking ability, contributing to increase collaboration. Like activeness, **Sales department is also proved to be the most influential department.** HR and communication departments need to put some efforts to improve networking among Marketing and Engineering departments by organizing more social activities.

Future Work:

The project is based on descriptive analysis. The analysis is done on the data of randomly selected employees from 6 departments for a period of 6 months. Future work can be dedicated to creating machine learning model to predict the links of a given node. One of the examples of usage of this analysis is to supplement the churn models using different surveys.

Bibliography

McDowell, T., Horn, H., & Witkowski, D. (2016). Organisational Network Analysis. Deloitte.

GSMA. (2022). *The Mobile Economy 2022*. Retrieved from GSMA: https://www.gsma.com/mobileeconomy/

Borgatt, S. P., & Foster, P. C. (2003). The Network Paradigm in Organizational Research: A Review and Typology. *Journal of Management*, 991-1013.

Sen, P., Deb, P., & Kumar, N. (2021, July 31). *The Challenges of Work From Home for Organizational Design*. Retrieved from California Management Review: https://cmr.berkeley.edu/2021/07/the-challenges-of-work-from-home-for-organizational-design/

Justina Alexandra Sava. (2022, November 22). *Work from home & remote work - Statistics & Facts*. Retrieved from Statista: https://www.statista.com/topics/6565/work-from-home-and-remote-work/#dossierContents__outerWrapper

Derr, A. (2021, January 20). *Types of Organizational Networks: 6 Ways to Organize a Network*. Retrieved from Visible Network Labs: https://visiblenetworklabs.com/2021/01/20/types-of-organizational-networks/

Ujwary-Gil , A. (2016). Organizational Network Analysis of the Interplay Between Business Model Components.

Hanneman , R. A., & Riddle, M. (n.d.). *Introduction to social network methods.* Retrieved from http://www.faculty.ucr.edu/~hanneman/nettext/C10_Centrality.html#Closeness

Wang, P. (2019, April 10). *Organizational network analysis – an experimental study.* Retrieved from AI Journey: https://ai-journey.com/2019/04/organizational-network-analysis/

Ezequiel Ortiz Recalde. (2022, June 29). *An Introduction to Organizational Network Analysis*. Retrieved from medium: https://medium.com/towards-data-science/an-introduction-to-organizational-network-analysis-cfa91a0e2fda
i4cp. (2021, May 10). *WHAT IS ORGANIZATIONAL NETWORK ANALYSIS? AND HOW DOES IT BENEFIT COMPANIES?* Retrieved from i4cp: https://www.i4cp.com/productivity-blog/what-organizational-network-analysis-is-and-how-it-benefits-companies
Gibson, C., Hardy III, J., & Buckley, M. (2014). Understanding the role of networking inorganizations. *Career Development International*, 146-161.
Gratton, L., & Erickson, T. (2007). *Eight Ways to Build Collaborative Teams*. Retrieved from Harvard Business Review: https://hbr.org/2007/11/eight-ways-to-build-collaborative-teams

Appendix:

Appendix A: Sorted dataframe for Degree centrality value

	Employee_id	department	location	age	Connections	Degree	Closeness	Betweenness	Eigenvector
278	598	Operations	US	38	84	1.000000	0.666458	0.939673	0.592904
53	144	Sales	US	50	78	0.928571	0.320246	0.495242	0.022778
48	128	Sales	France	47	75	0.892857	0.524659	0.596213	0.255030
280	605	Admin	France	31	71	0.845238	0.530300	0.497212	0.369941
271	586	Operations	France	38	65	0.773810	0.410983	0.248176	0.093507

Appendix B: Sorted dataframe for Closeness centrality value

	Employee_id	department	location	age	Connections	Degree	Closeness	Betweenness	Eigenvector
234	509	Admin	US	39	55	0.654762	0.725263	1.000000	0.744123
147	337	Sales	US	37	42	0.500000	0.657572	0.976297	0.570332
278	598	Operations	US	38	84	1.000000	0.666458	0.939673	0.592904
48	128	Sales	France	47	75	0.892857	0.524659	0.596213	0.255030
102	260	Operations	US	48	39	0.464286	0.725263	0.559868	0.778682

Appendix C: Sorted dataframe for Betweenness centrality value

	Employee_id	department	location	age	Connections	Degree	Closeness	Betweenness	Eigenvector
234	509	Admin	US	39	55	0.654762	0.725263	1.000000	0.744123
147	337	Sales	US	37	42	0.500000	0.657572	0.976297	0.570332
278	598	Operations	US	38	84	1.000000	0.666458	0.939673	0.592904
48	128	Sales	France	47	75	0.892857	0.524659	0.596213	0.255030
102	260	Operations	US	48	39	0.464286	0.725263	0.559868	0.778682

Appendix D: Sorted dataframe for Betweenness centrality value

	Employee_id	department	location	age	Connections	Degree	Closeness	Betweenness	Eigenvector
74	194	Admin	US	28	13	0.154762	0.982479	0.000000	1.000000
102	260	Operations	US	48	39	0.464286	0.725263	0.559868	0.778682
525	1283	Admin	US	29	8	0.095238	0.761553	0.000000	0.758304
234	509	Admin	US	39	55	0.654762	0.725263	1.000000	0.744123
42	103	Sales	Brasil	23	8	0.095238	0.954790	0.000000	0.742087

Appendix E: Dataframe for centrality values for departments

	Employee_id	Connections	Degree	Closeness	Betweenness	Eigenvector	Weighted_average
0	Sales	12	1.000000	1.000000	1.000000	1.000000	1.000000
1	IT	11	0.916667	0.833333	0.500000	0.875130	0.804180
2	Operations	12	1.000000	1.000000	1.000000	1.000000	1.000000
3	Marketing	8	0.666667	0.833333	0.000000	0.695621	0.586229
4	Engineering	11	0.916667	1.000000	0.500000	1.000000	0.866667
5	Admin	12	1.000000	1.000000	1.000000	1.000000	1.000000