06/23/2024

NETWROKS THEORY PROJECTAS

MARIANO MIRANA SANCHEZ DATA 8°B

SOCIAL NETWORK ANALYSIS

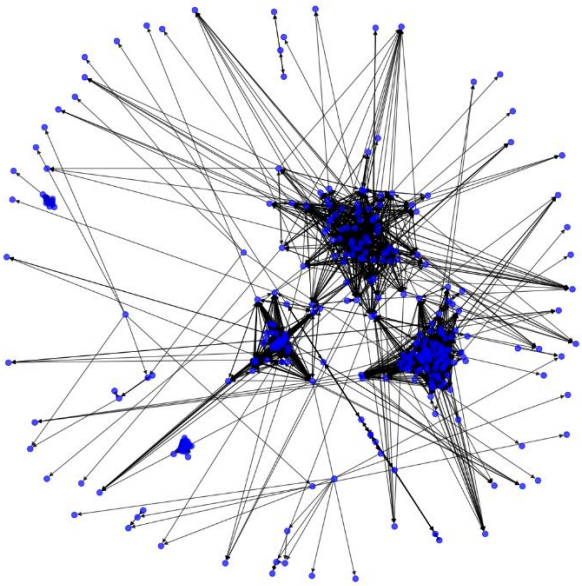
DIDIER OMAR GAMBOA ANULO
Univeria Politecnica de Yucatan

Abstract

The project uses graph theory and social media analysis techniques to analyze an organization's email network. The network consists of 309 nodes and 3031 links, indicating a high density of connections. The average length of the path on the network is 2.19, indicating rapid information propagation between nodes. The grouping coefficient of 0.57 indicates the formation of clusters inside the network. The centrality analysis identified key nodes, with node 279 standing out for its high centrality of intermediation and eigenvector. The degree distribution follows a power law, with just a few highly connected nodes. The detection of communities using the Girvan-Newman method revealed a number of well-defined subcommunities. These findings provide a detailed understanding of the structure and dynamics of the network, providing strategies to improve and optimize organizational communication.

Introduction

The network used in the project is an email network within a department. This network is highly relevant as it allows for the analysis of interactions and the flow of information within the organization.



(If use your imagination looks like a pokeball)

Some of the key points to highlight from the dataset are the internal communication (unlike my ex), understanding how department members communicate. The identification of central nodes shows the people who act as hubs within the network.

The detection of communities reveals subgroups that communicate frequently among themselves. Optimization of information flow results in better efficiency in internal communication between nodes.

Betweenness centrality within the network is significant, with nodes having high betweenness centrality acting as critical bridges. These nodes facilitate communication between different parts of the network and are crucial for the network's cohesion.

This network is classified as directed, due to the direction in which emails must be sent, having a clear destination from the sender to the recipient. The links between nodes, which in this case are the people using the emails, have a direction. This means that the flow of information is unidirectional in each case.

Network Characteristics

In the network analysis, 309 existing nodes with 3031 links were found, with an average path length of 2.19. This indicates that any node can be reached by another within two steps on average. The mean distance of the network is equal to the average path length, with a value of 2.19, confirming the strong connectivity of the network and the rapid propagation of information between nearby nodes. However, more distant nodes cannot connect with all existing nodes. The clustering coefficient of the network is 0.57, indicating a tendency for nodes to form clusters and suggesting the presence of subgroups within the network where many neighbors are also connected to each other.

The diameter of the network is nonexistent since it is not strongly connected, meaning not all nodes are reachable from other nodes, which prevents direct calculation of the diameter.

The network's eccentricity varies in a range of 3 to 5 in the largest connected components, which is relatively small. The most common eccentricities in the network are 3 or 4, with few nodes having a value of 5, such as nodes 37, 46, 105, 146, etc. These nodes are connected to the most isolated nodes in the network. Finally, the radius of the network is 3, being the smallest value of node eccentricity, occurring mostly in the central nodes of the network.

Centrality Measure

The centralization measures for this network are betweenness centrality and eigenvector centrality.

Betweenness centrality is important because, in a network with extensive communication among nodes, the bridges between nodes in different groups are crucial for the constant flow of information and network connectivity. This measure also highlights the roles of certain nodes that connect with the more isolated nodes in the network or with other subgroups.

- Betweenes Centrality results

```
[(279, 0.015615617568068592),
 (234, 0.008098834930110613),
 (252, 0.007985088547564766),
 (299, 0.005723504539919011),
 (57, 0.00493906170079727),
 (7, 0.00481592790250476),
 (228, 0.004740735185574033),
 (26, 0.00392251272821579),
 (278, 0.0037281323202759645),
 (90, 0.0032407173320805585)]
```

Node 279 has the highest value of 0.0156, serving as a crucial bridge in the network, facilitating the transfer of information across different parts of the network. Nodes 234, 252, and 290 are equally important as they have high values and maintain greater connectivity between nodes, facilitating cross-communication.

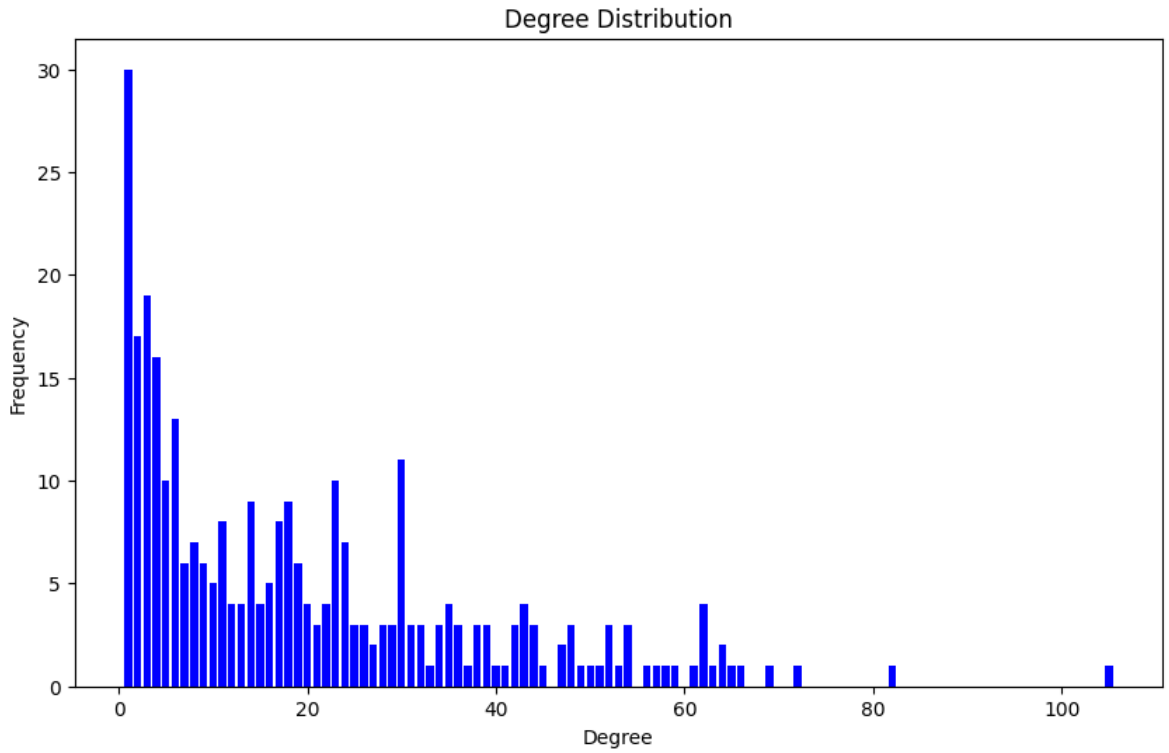
Eigenvector centrality is also useful as it measures the global influence of a node, considering not only direct connections but also connections to other influential neighbors. Nodes with high eigenvector centrality are ideal for rapid and wide dissemination of information because they have strong connections with other important nodes, whether within the same group or with others.

- Eigenvector Centrality results

```
[(279, 0.27435997687451624),
 (234, 0.2303160079887136),
 (299, 0.22725642433587326),
 (228, 0.22442701307146307),
 (135, 0.20208126938736778),
 (278, 0.19264384184286645),
 (241, 0.19207400529812493),
 (202, 0.1818783243364019),
 (242, 0.1803660496166467),
 (116, 0.17369614704844885)]
```

The results of Eigenvector centrality are similar to those of Betweenness centrality, largely confirming the conclusion obtained from the first measure.

Degree Distribution

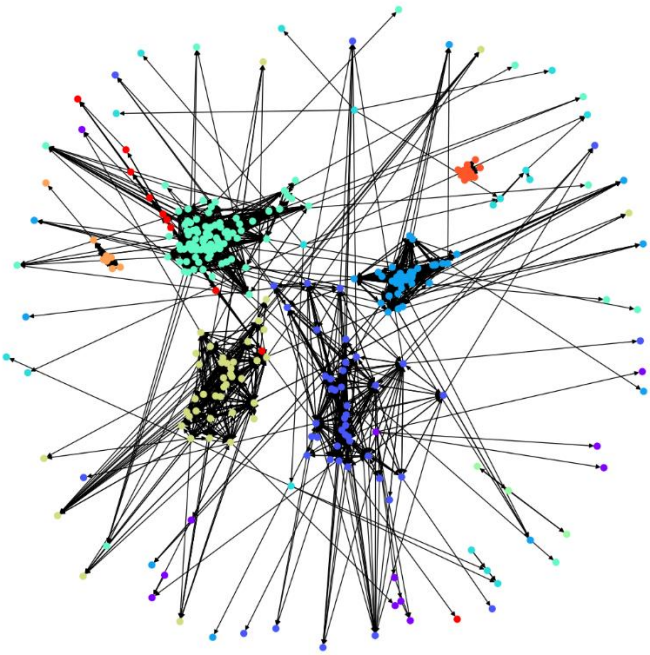


The network sample shows the frequency where, in the case on the left, there is a low number of connections with a high frequency of nodes with small degrees, ranging between 0 and 10. This indicates that the majority of the nodes in the network are sparsely connected.

However, there are nodes with a significantly high number of connections, having more than 100 links, known as 'hubs', which play an important role in the network by facilitating connectivity.

The results in the middle of the graph are nodes with degrees between 10 and 30, indicating that the network is not composed solely of isolated nodes or those with few existing connections.

Community Detection



The graph shows each community with a different color, highlighting the points that act as bridges between communities as well as the isolated nodes. Dense communities are easier to observe due to higher interaction among their nodes, as seen with the cyan and blue communities, which have a greater number of nodes acting as bridges between isolated nodes and communities. Additionally, there are two groups of nodes whose main function is to serve as bridges between specific nodes, such as the red and orange communities.

Conclusion

The network represents interesting aspects such as the groups that form, the importance of the bridges for communication between these groups, as well as the isolated nodes. These groups have a strong connection due to the short distance between nodes, making the movement of information faster and more consistent compared to the isolated nodes.

The main nodes, which are the central ones between groups, are crucial because they handle most of the information exchange between communities. These nodes tend to have larger timespans as they manage a greater amount of information.

References

- Centrality Analysis in Networks: An Approach from Game Theory. (n.d.). https://www.researchgate.net/publication/228451062_Analisis_de_Centralidad_en_Redes_Un_enfoque_desde_la_Teoria_de_Juegos
- Social Network Analysis: Basic Concepts and Applications. (n.d.). https://www.researchgate.net/publication/228451062_Analisis_de_Centralidad_en_Redes_Un_enfoque_desde_la_Teoria_de_Juegos
- Visualization and Analysis of Complex Networks with Python. (n.d.). https://nbviewer.jupyter.org/github/networkx/networkx/blob/master/examples/drawing/plot_ego_graph.ipynb
- Degree Distribution in Complex Networks. (n.d.). https://www.researchgate.net/publication/228451062_Analisis_de_Centralidad_en_Redes_Un_enfoque_desde_la_Teoria_de_Juegos