CHALLENGE_2
Daniela BATALHA
Ishita CHAUDHARY

Reformulated Problem:
A business hierarchy is one of the many ways that a company can structure its employees.
A good structuration of the hierarchy in organizations can be essential to manage employees better and gain an advantage of performance within the market [1]. Understanding this hierarchy can be beneficial to solving many internal problems and highlight, for example, a lack of sufficient human resources to accomplish all parts of a given task, the level of productivity, and the quality in the attribution of roles, among others.

As data scientists in a consultant company, we are interested in using graph theory analysis to reveal hierarchies in the organizational structure of our client company to help their Human Resources managers better identify the productivity of employers and improve the quality of attribution of roles and performance.

Decision-maker interested:
Manager of the internal RH of the client company.

Why understanding the hierarchy of an organization is essential?
Understanding a company's hierarchy is crucial because it can help better comprehend the relation between the actual employer's role and the proper position attribution [2]. The social network analysis of corporate emails can be a valuable tool for this task. It can be helpful to address questions such as: Is this employee been productive as he should be? Is this employer been burdened with the job? If the company decides to search internally for a successor, which employee could be a good one? It may also help extract some future problems in the company, like managers avoiding communication with other employees within the organization.

The final objective of our analyses is to reveal the organizational structure and help the company gain an advantage of performance within the market.

Reveal the structure of an organization using the emails exchanged between the employees:
People send hundreds of emails to exchange information within organizations. As an implicit result, each of these interactions forms a link in a social network. Conducting the analysis can reveal groups of employees with similar communication patterns and the organization's hierarchy.

Diving into the Graphs:
We can reformulate our problem as a directed weighted graph G:= (V, E), where we have the organization's employees (in the node-set V). The edge between any two nodes will be directed from the sender towards the receiver. Each edge will have an ID associated with it; the subject and date-time stamp will determine the ID to identify if the same email was sent to many recipients. The weight of edge e_{ij}, from node i to node j is defined as:

w_{ij} = (number of emails sent from i to j)/ (total number of emails sent from node i)

We will use a different color of nodes for different seniority levels of nodes, for observation purposes.
Also, we will calculate a local clustering coefficient for each node, which allows capturing the density of connections between neighbors, as well as two additional features related to cliques:

CC_i = 2(number of pairs of neighbors of a node i)/(d(d-1))
where d is the degree of node i

We will try various classification algorithms to segregate the emails based on different coefficients and degrees of nodes calculated, like Decision Tree, Random Forest, Neural Network, etc.

Suggestions based on possible observations:
As data scientists, we can suggest the company to take the following measures to improve the efficiency:

- If there are too many emails exchanged between two employees of equal seniority level, i.e., edge weight close to one, then we can suggest them to use some internal communication method over not email.
- We can give the ID weights according to the time stamp, i.e., if many employees receive an email outside of which is not during the office working hours, then it is highly possible that the email is not important
- As an exception, if a mail is received outside of office hours to a senior level of employee, then it might be very urgent.

Sources:

Principal

1. https://arxiv.org/abs/1906.09576
2. https://www.researchgate.net/publication/221281813_Matching_Organizational_Structure_and_Social_Network_Extracted_from_Email_Communication

Extras

3. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.2696&rep=rep1&type=pdf
4. https://www.mdpi.com/2076-0760/8/11/306/htm
5. https://dl.acm.org/doi/10.1145/1593105.1593229
6. https://academiccommons.columbia.edu/doi/10.7916/D85T3T9Z/download
7. https://hbr.org/2011/02/the-importance-of-organization
8. https://link.springer.com/article/10.1057/s41266-018-0033-y