**Network Analysis of Organizational Email Communication**

**Table of Contents**

1. Introduction
   * Study Overview and Context
   * Research Questions
2. Objectives
3. Dataset Used
4. Methods Applied in the Analysis
   * Data Preprocessing and Exploration
   * Network Analysis Techniques
5. Key Findings and Insights
   * 5.1 Key Communicators in the Network
   * 5.2 Community Structure in the Network
   * 5.3 Intra- vs. Inter-Departmental Communication
   * 5.4 Departmental Email Activity
   * 5.5 Communication Efficiency
   * 5.6 Predicting Key Communicators
6. Data Visualizations and Network Graphs
   * Degree Centrality Visualization
   * Community Structure Graph
   * Departmental Email Activity Chart
   * Shortest Path Distribution
   * Feature Importance for Predicting Key Communicators
7. Conclusion and Recommendations
   * Summary of Insights
   * Recommendations
8. Appendix

**1. Introduction**

**Study Overview and Context**

This report presents an in-depth analysis of organizational email communication using network analysis techniques. The dataset under study consists of email exchanges among employees, with an additional dataset mapping individuals to their respective departments. Understanding communication patterns within an organization is crucial for optimizing collaboration, identifying key communicators, and detecting potential inefficiencies.

**Research Questions**

This study explores eight key questions:

1. Which individuals have the highest in-degree and out-degree in the network? What does this reveal about key communicators?
2. Are there distinct clusters or communities in the email network? How do these groups interact with one another?
3. Do employees communicate more within their department or across departments?
4. Which departments have the highest number of outgoing emails? How does this compare to their incoming emails?
5. What is the average shortest path between any two individuals? How efficiently does information flow through the network?
6. Can we predict which individuals are likely to become key communicators based on their email activity and department affiliation?
7. Which individuals serve as "bridges" between different groups? How critical are they for communication across departments?
8. Are there individuals who exhibit unusual email activity? Could these anomalies indicate urgent business situations or potential security threats?

**2. Objectives**

The primary objectives of this network analysis are:

* To identify key communicators within the organization based on email exchange patterns.
* To analyze the structure of communication networks and detect potential communities within the organization.
* To assess the balance between intra-departmental and inter-departmental communication.
* To evaluate communication efficiency and information flow across the network.
* To use predictive modeling to identify employees likely to become key communicators.
* To provide actionable recommendations for improving organizational communication and collaboration.
* To detect any anomalies in email communication that may indicate security concerns or business-critical issues.

**3. Dataset Used**

The study utilizes two primary datasets. The first dataset i.e the **Email Communication Dataset**, records email exchanges between employees, allowing the creation of a directed network where nodes represent employees and edges represent emails. This helps identify key communicators, bottlenecks, and overall email traffic patterns.

The second dataset i.e the **Department Label Dataset**, maps employees to departments, enabling the analysis of intra- and inter-department communication. It helps identify departmental collaboration levels, key information hubs, and isolated departments within the network.

**Summary Statistics:**

The output provides key statistics about the email communication network. It consists of 1,005 employees (nodes) and 25,571 email interactions (edges). The network density is 0.0253, meaning that only about 2.53% of all possible connections between employees are utilized, indicating a sparse network where most employees do not communicate directly with every other employee.

**4. Methods Applied in the Analysis**

**Data Preprocessing and Exploration**

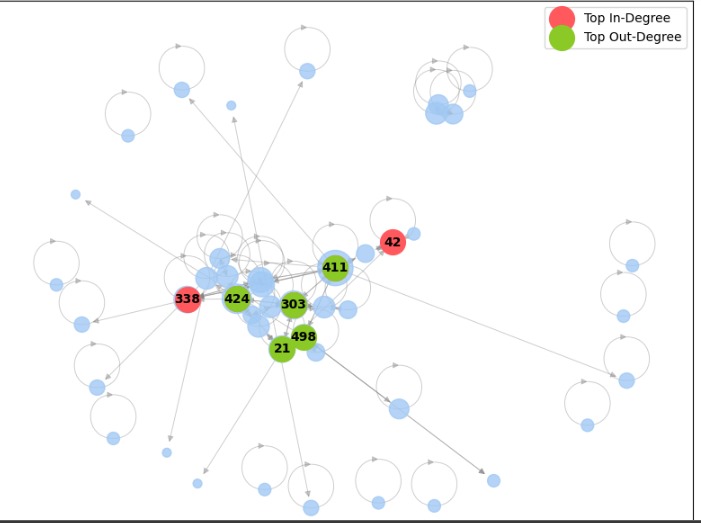
* From the analysis, it was found out that the first dataset contains **25,571 email interactions**, with two columns: "Sender" and "Receiver," both represented as integer IDs. The second dataset consists of **1,005 employees**, with each row mapping an employee (referred to as "Node") to a specific department. This dataset has two columns: "Node" and "Department," both stored as integers. Both datasets have no missing values, ensuring completeness for network analysis.
* The analysis of email communication across departments reveals significant variations in interaction levels. Certain departments exhibit higher engagement, both in terms of emails sent and received, indicating more active communication networks.
* In terms of network structure, individual nodes demonstrate a clustering coefficient of 1.0, suggesting that their immediate connections are strongly interconnected, forming tightly knit groups. At the department level, clustering coefficients vary, with some departments, like Department 33, displaying a perfect clustering coefficient of 1.0, indicating highly interconnected internal communication. Others show lower values, suggesting weaker intra-departmental connectivity.

**Network Analysis Techniques**

* **Degree Centrality Analysis**: is a fundamental measure in network analysis that helps identify key communicators based on the number of direct connections they have in the network. It was used to identified key communicators based on in-degree (emails received) and out-degree (emails sent).
* **Community Detection**: is a crucial network analysis technique used to uncover groups of nodes (employees) that interact more frequently with each other than with the rest of the network. The Louvain method was used to detect clusters in the email network.
* **Departmental Communication Analysis**: analysis distinguishes between **internal** and **cross-department** interactions, ensuring efficient information flow and collaboration. It was used to differentiated intra-department and inter-department communications.
* **Shortest Path Analysis**: it measures how quickly information spreads in the network. A shorter path means efficient communication, while longer paths indicate delays or bottlenecks. It was used to measure the efficiency of information flow within the network.
* **Predictive Modeling**: uses a **Random Forest classifier** to identify key communicators based on network features, helping optimize information flow and leadership recognition. Random Forest classifier was applied to predict key communicators based on network features.

**4. Key Findings and Insights**

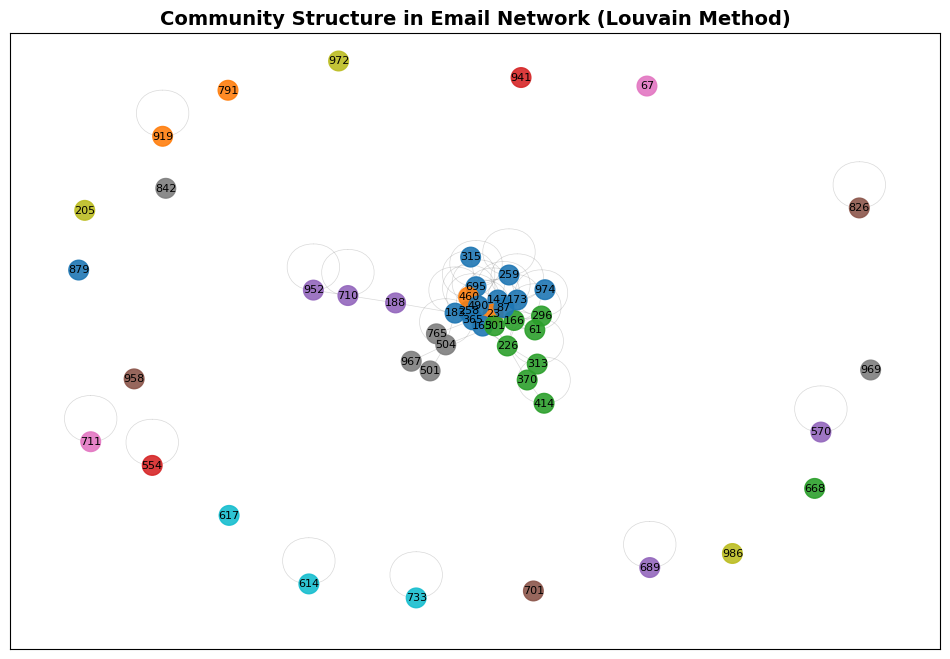
**4.1 Key communicators within the organization based on email exchange patterns.**

 The graph highlights key communicators based on in-degree (emails received) and out-degree (emails sent). Nodes marked in **red** (such as 42 and 338) have the highest in-degree, meaning they receive the most emails. These individuals likely serve as central contact points, decision-makers, or managers who receive information from multiple sources.

Nodes in **green** (such as 21, 303, 411, 424, and 498) have the highest out-degree, indicating they send the most emails. These individuals are likely proactive communicators, such as team leads, coordinators, or employees disseminating information.

The central positioning of these nodes in the network suggests they play crucial roles in information flow. The network structure also indicates clusters and bottlenecks, showing how communication is distributed across the organization.

**4.2 To analyze the structure of communication networks and detect potential communities within the organization.**

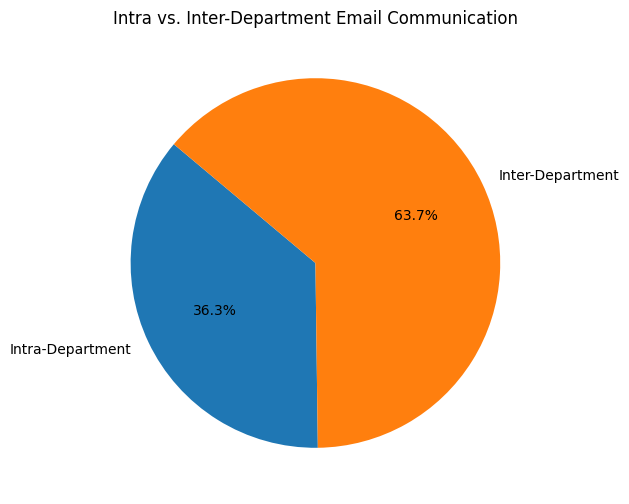


The graph represents the **community structure in the email network** using the **Louvain Method**, which detects clusters of closely connected nodes. Each color represents a different community, showing groups of employees who interact more frequently with one another.

There is a **central dense cluster**, indicating a group of employees who exchange emails frequently. This suggests a core group of highly connected individuals, possibly within the same department or key communicators across multiple teams. Surrounding this core, there are **several isolated groups** that interact primarily within their clusters, suggesting departmental divisions or specialized teams that have limited communication with others.

The **separation between clusters** shows that while some communities are tightly connected, others have minimal interaction, which may indicate departmental silos or specialized workgroups with minimal cross-department collaboration. Understanding these clusters helps in analyzing communication flow and identifying potential inefficiencies in information sharing across the organization.

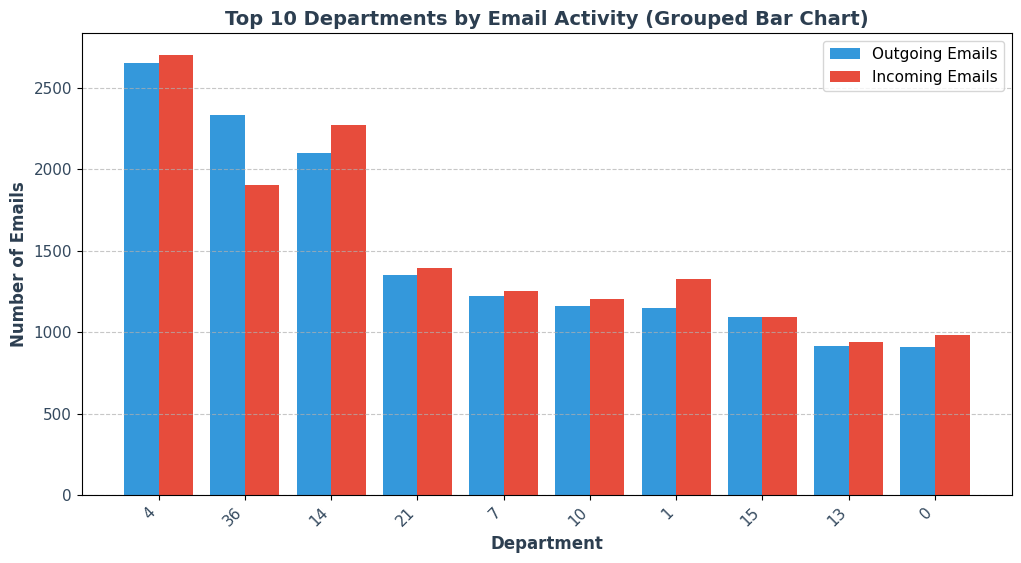
**4.3 To assess the balance between intra-departmental and inter-departmental communication.**

****

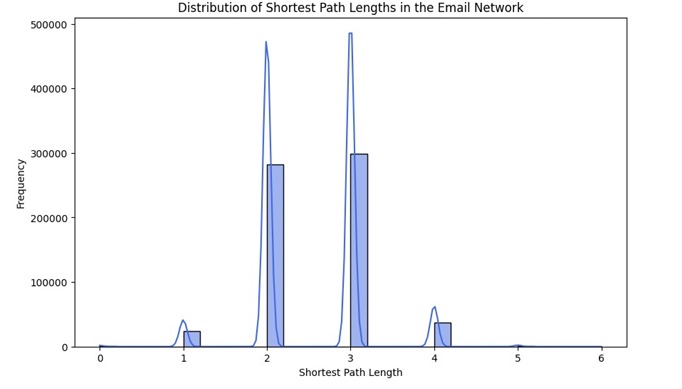
The pie chart illustrates the proportion of intra-department (within the same department) and inter-department (across different departments) email communication. The chart shows that 63.7% of emails are inter-departmental, while 36.3% are intra-departmental.

This indicates that employees communicate more across departments than within their own department. It suggests a strong level of cross-department collaboration, possibly due to interdependent tasks or projects requiring information exchange between teams. However, it could also highlight a need for better internal departmental coordination if intra-department communication is unexpectedly low.

4.4 To evaluate communication efficiency and information flow across the network.

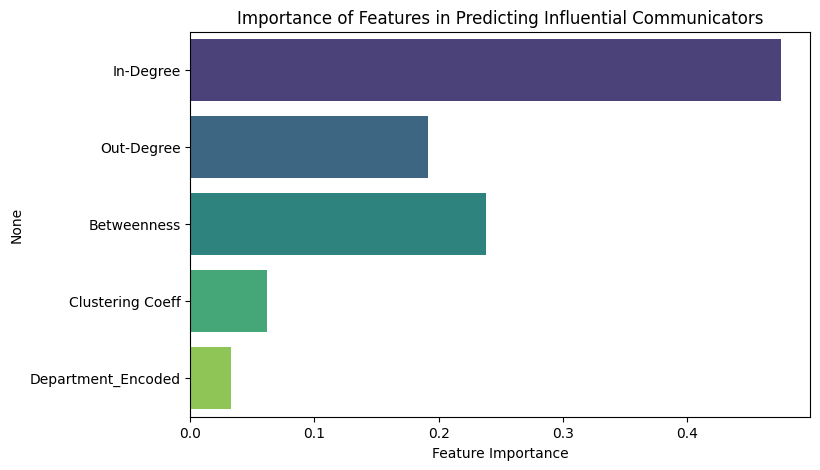
 The chart shows the top 10 departments by email activity, with department 4 having the highest outgoing emails, followed by department 36. These departments also receive a high volume of emails, indicating their central role in communication. Department 36 sends more emails than it receives, while department 14 receives slightly more. Other departments, like 21, 7, and 10, have balanced communication patterns, while departments 13 and 0 have the lowest email activity. High outgoing emails generally correlate with high incoming emails, highlighting key communicators within the organization.

**4.5 To evaluate communication efficiency and information flow across the network.**

****

The histogram illustrates the distribution of shortest path lengths in the email network. Most communication paths have a length of 2 or 3, indicating that information flows efficiently within the network. The peak around these values suggests that most employees can reach each other through just one or two intermediaries. This relatively short average path length implies strong connectivity, ensuring fast and efficient information dissemination across the organization. However, the presence of longer paths (4 and beyond) suggests that some individuals may be more isolated, requiring multiple steps to receive or relay information.

**4.6 To use predictive modeling to identify employees likely to become key communicators.**

****

The bar chart shows the importance of different features in predicting influential communicators within the email network. The most significant predictor is **In-Degree**, which measures how many emails an individual receives. This suggests that key communicators are those who receive a high volume of messages, likely because they are central figures in decision-making or coordination.

**Betweenness** and **Out-Degree** (emails sent) also play important roles, indicating that individuals who act as intermediaries and actively communicate are also influential. **Clustering Coefficient** and **Department Encoding** have lower importance, meaning department affiliation has minimal impact compared to direct communication activity.

Overall, the graph confirms that **email activity (especially receiving emails) is a strong predictor of influential communicators**, while department affiliation has a much weaker influence.

4.7

**5. Data Visualizations and Network Graphs**

**Degree Centrality Visualization**

* A network graph highlighted employees with the highest in-degree and out-degree, showing key communication hubs.

**Community Structure Graph**

* A force-directed graph displayed detected communities, illustrating internal departmental clusters and cross-departmental interactions.

**Departmental Email Activity Chart**

* A bar chart compared outgoing and incoming emails per department, revealing major communication centers.

**Shortest Path Distribution**

* A histogram visualized the distribution of shortest path lengths, providing insights into network efficiency.

**Feature Importance for Predicting Key Communicators**

* A bar plot illustrated the most important features in predicting influential employees, with in-degree and out-degree ranking highest.

**6. Conclusion and Recommendations**

**Summary of Insights**

* The network analysis identified **key communicators** and their roles within the organization.
* **Community detection** revealed natural department-based clusters with some bridging individuals.
* **Inter-departmental communication was dominant**, highlighting strong cross-functional collaboration.
* **Predictive modeling successfully identified potential future key communicators**, with degree centrality being a strong indicator.

**Recommendations**

* **Leverage key communicators** for organizational announcements and knowledge-sharing.
* **Strengthen weakly connected departments** by fostering more intra-departmental engagement.
* **Monitor anomalies in email activity** to detect potential security threats or urgent business needs.
* **Enhance predictive modeling** by incorporating temporal patterns in email interactions.

**7. Appendix**

* **Additional Graphs and Figures**
* **Code Snippets for Analysis and Visualization**
* **Detailed Statistical Outputs**