**Social Network Analysis of Institutional Email Communication - Group 6**

Alireza Hatami – ah3733@drexel.edu

Robert Lignowski – rml345@drexel.edu

Ram Kishore Karuppiah Nadar Venkateswaran – rk999@drexel.edu

**Introduction**

People communicate through various channels, such as emails, forum discussions, and instant messaging. This exchange of information between individuals forms the foundation of a social network (Kazienko et al., 2009). However, according to Uddin and Jacobson (2013), sending emails is one of the most essential methods of communication in academic settings. It allows individuals to provide detailed information about their topics of interest and include various references, such as files, images, and links, to better convey their intentions. This project examines email as a form of communication, and such a network can be modeled as a directed graph, where nodes represent individuals and edges signify the flow of information from one user to another. Using various methods in social network analysis, we can explore the relationship between different nodes or individuals and evaluate their influence in the social network. One of the most important aspects of social network analysis is evaluating the position of nodes within the network. This evaluation helps identify the most influential individuals in a group or community who may hold the highest social status (Kazienko et al., 2009). Uddin and Jacobson (2013) suggest that the pattern of communication among individuals can be explored by implementing social network analysis, and this will help us to uncover hidden traits and gain valuable insight about the social network. This project begins with a review of existing literature related to email communication. After formulating research questions, there is a methodology section where various methods are used to address these questions. For each methodology, the results are discussed, and finally, recommendations are provided based on the analysis.

**Literature Review**

Social network analysis (SNA) can help us understand how people communicate in organizations such as businesses, government agencies, schools, or research institutes. It helps us examine how communication happens, see how information flows, identify problems in communication, and find ways to fix them. This can help improve communication efficiency within these organizations (Michalski et al., 2011). Michalski et al. (2011) state that identifying key influential nodes is very important because, within organizations, each individual plays a unique role. The position of these individuals may grant them access to information and resources that are vital for the organization's communication. Influential figures can promote a more active role in communication and in disseminating news and information among others and possibly activating other network members. Conversely, individuals who are less influential can be encouraged to take on a more meaningful role in communications.

In addition, Social Network Analysis (SNA) can be implemented to uncover the patterns of communication among the members of organizations (Christidis & Gomez Losada, 2019). Kolli and Narayanaswamy (2013) explained that an individual's position within an organization should be analyzed because each position may hold varying levels of authority, which suggests the presence of hierarchical communication. This communication pattern can be identified using social network analysis methods to reveal the hierarchical nature of communication in the email exchanges among individuals in the organization. While the study by Kolli and Narayanaswamy (2013) investigated the temporal aspects of email communication, it emphasized the significance of individuals' positions and communication patterns within social network analysis.

Certain nodes in a social network hold more crucial positions, meaning that after a specific period, their influence remains relatively stable. These individuals can be identified using various metrics, and their centrality can be assessed to determine their importance within the overall dynamics of the network (Malinka & Schäfer, 2009). Examining nodes as cohesive subgroups offers valuable insights into their interactions and proximity within a network. Identifying K-core structures can effectively reveal tightly knit groups collaborating together (Yee et al., 2005), a concept that can also be applied to the relationships among individuals in an organization. By analyzing internal communication within an organization, we can identify K-core formations. While Yee et al. (2005) suggested using K-core analysis for medium-sized organizations, we believe that this approach can equally help identify cohesive subgroups in research institutions.

Some of the main approaches used to describe how social networks are structured are density and centralization, which are proposed as main ideas by Weare et al. (2007). Density indicates the ratio of actual connections among group members to possible connections, while centralization measures the extent to which communication is concentrated among specific individuals instead of being evenly distributed. In addition, Uddin and Jacobson (2013) emphasized the importance of using density and calculating centrality measures, such as closeness centrality for individual nodes, as well as group centralization. They also highlighted the significance of in-degree and out-degree distributions for analyzing interaction patterns within organizations. The in-degree can help identify the most prestigious individuals, while the out-degree can reveal the most active and influential nodes in terms of information dissemination (De Choudhury et al., 2010).

In this project, we are focusing on the communication of individuals working in a European research institution. By analyzing the email communication network of a research institution, we can understand interaction patterns, collaboration, and departmental relationships. According to Christidis & Gomez Losada (2019), understanding these aspects can provide insights into the members’ level of influence in the research institution, communication efficiency, information flow, and potential bottlenecks. The results of this project's analysis will potentially help improve communication and collaboration within the institution. According to Michalski et al., (2011), combining the results of these methods will enhance interactions among various players in the network even further. However, there is always a need for new approaches in social network analysis. Therefore, in this project, we will explore additional methodologies to analyze the dynamics of communication within research institutes.

**Objectives**

Despite the extensive research on social network analysis (SNA) in email communications, several gaps remain unexplored. Previous studies have primarily focused on degree centrality when identifying key influential figures, overlooking other crucial centrality measures such as betweenness and closeness centrality, which offer deeper insights into the network's structure and information flow. Additionally, research has largely emphasized temporal analysis, while other network properties—such as hierarchical communication detection—remain underexplored. Understanding hierarchical structures in email interactions can reveal power dynamics and decision-making patterns within an organization. Furthermore, while K-core analysis has been widely used to identify cohesive subgroups, other methods such as K-plex analysis (a more flexible alternative to cliques) should be investigated to better capture collaboration patterns. Also, previous studies have not sufficiently examined interdepartmental interaction patterns and their significance in organizational communication. Previous studies on similar datasets did not take network robustness into account. We also wanted to find the structural equivalents, with could be insightful and rewarding for our analysis. By addressing these gaps, this project aims to provide a more comprehensive understanding of email communication within a research institution, ultimately improving communication efficiency and collaboration. Several questions have been formulated to identify key influential figures and better understand the network's structure, information flow, and communication efficiency.

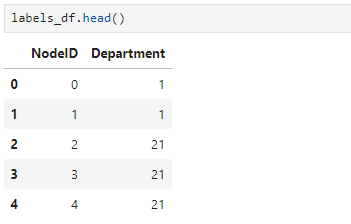
1. **Network Structure and Communication Pattern Among Departments:**
   * What are the basic characteristics of a network, such as mean degree, variance, and density?
   * What does email communication between individuals at a research institution look like (the network visualization)?
   * What is the communication pattern between departments and within each department?
   * Which departments interact with each other the most?
2. **Path Analysis and Efficiency:**
   * What are the shortest paths and geodesic distances within the network?
   * How does the diameter reflect overall communication efficiency?
3. **Identifying Key Players and Examining Group Centralities:**
   * Who are the most influential individuals in the network?
   * How does group centralization vary across departments?
4. **Group Dynamics and Cohesion:**
   * Are there strongly cohesive subgroups (such as k-plexes or k-cores) within the network?
   * Which departments play more important roles in the cohesive subgroups?
5. **Information Flow and Reciprocity:**
   * What is the reciprocity of email communication, and how does it vary across the network?
   * Are there structural imbalances or evidence of hierarchical communication?
   * How robust is the network to node or edge removal? Would removing certain nodes disrupt the communication structure \significantly?
6. **Network Robustness and Node Removal:**
   * How resilient is the network to the removal of different types of nodes, such as those with high degree, high betweenness, or randomly selected nodes?
   * Which removal strategy causes the most fragmentation, and why might that be critical for understanding vulnerabilities?
   * What does the concept of a “network robustness score” reveal about overall connectivity after specific nodes are removed?
   * How can these findings inform strategies to minimize the impact of losing key individuals?
7. **Hierarchical Nodes:**
   * Which centrality measures most clearly identify hierarchical nodes, and why are these measures effective?
   * How do hierarchical nodes facilitate or control communication between otherwise disconnected departments or subgroups?
   * What happens to overall connectivity when hierarchical nodes are removed, and how can the network mitigate this risk?
8. **Structural Equivalence:**

* Which groups of individuals are structurally equivalent, and what roles do they play in institutional communication?
* How does the reduced network representation help in understanding hierarchical structures and information flow?

**Dataset**

For this project, we are focusing on a dataset representing various incoming and outgoing emails sent from researchers at a large European research institution. The data only covers emails sent to other members of the institution, ignoring any emails sent outside the institution. In addition, this dataset also includes information about each member, indicating which department they belong to within the institution (Leskovec & Krevl, 2014). Leskovec and Krevl (2014) coded the names of individuals and departments as numbers for privacy reasons.

In addition, the data was initially given in a text file (txt), so we needed to write a Python program to convert it into a CSV file for easier use. Figure 1 displays the data frame, which shows the nodes and their corresponding departments, while Figure 2 illustrates the sender of the emails (Source) and the recipients (Target). The dataset has 1,005 unique researchers, each represented as a node. Moreover, we have 25,571 instances of emails being sent from one researcher to another.



*Figure 1*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 2*

**Methodology**

**1. Network Structure and Communication Pattern Among Departments**

**1.1 Network Structure**

To better understand the overall structure and communication patterns within the research institution, we conducted an analysis of fundamental network metrics, including degree distribution, variance, density, and isolated nodes. Table 1 provides an overview of these key measures. The mean in-degree and out-degree values are approximately equal (~25), indicating that, on average, the number of emails sent and received by each researcher is balanced in the network. However, the high variance in out-degree (1098.23) suggests that while some researchers communicate frequently, others engage minimally. This discrepancy highlights that certain individuals may play a more significant role in sending emails. Additionally, the total-degree variance (3573.57) is notably large, further confirming that some individuals are far more central within the network, likely holding key positions that facilitate information flow.

Table 1: Network Metrics

|  |  |  |  |
| --- | --- | --- | --- |
| **Measure** | **In-Degree** | **Out-Degree** | **Total Degree** |
| **Mean** | ~25 | ~25 | ~50 |
| **Variance** | 771.95 | 1098.23 | 3573.57 |
| **Density** | ~0.025 | | |

The network's density is approximately 0.025, indicating that it is highly sparse. A low density suggests that most researchers interact selectively rather than maintaining broad communication across the institution. Senior researchers, for instance, may primarily engage with their students or closest collaborators rather than sending email to other individuals working in the research institution. Another important observation is the presence of self-loops. A total of 642 self-loops were identified in the network. While self-loops may occur due to various reasons, such as saving emails for record-keeping or reminders, their presence can increase the degree centrality measurements. Therefore, these self-loops were removed from the network for a clearer network analysis and a more accurate representation of communication patterns.

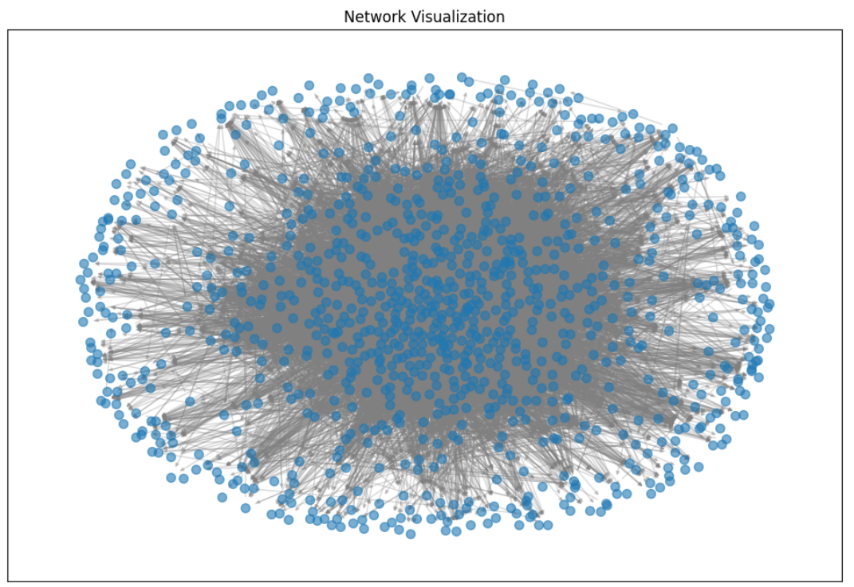


Figure 3: Network Visualization

Additionally, we identified 19 isolated nodes in the network, meaning these researchers do not engage in email exchanges with any other members of the institution. The list of isolated nodes includes individuals such as 580, 633, 648, 653, and others. The presence of isolated nodes suggests that some researchers may be inactive in email communication, either because they primarily use alternative communication methods or because they are no longer actively participating in the institution’s activities. To gain further insights into the network's structure and communication dynamics, we generated a directed network visualization (Figure 3). This graphic shows how emails move between researchers, highlighting the level of connection and centralization in the network. The visualization reveals a dense core of researchers who are highly connected, while others on the edges have fewer connections. This highlights the uneven engagement among researchers.

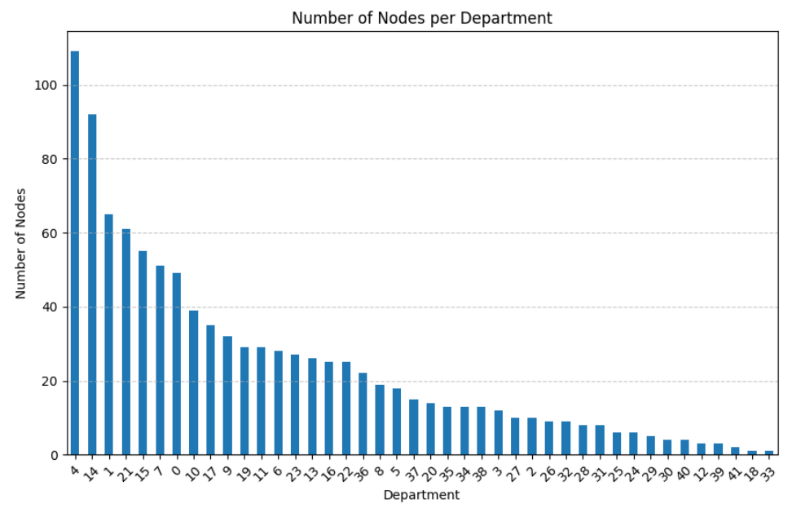


Figure 4: Number of Nodes per Department

Figure 4 presents a visualization of the number of researchers (nodes) assigned to each department within the institution. This distribution helps in understanding the varying sizes of different departments. The x-axis represents the department identifiers, while the y-axis indicates the number of individuals in each department. From the data, we observe that certain departments have significantly higher numbers of members, while others are much smaller. Department 4 has the largest number of researchers, with 109 members, followed by Department 14 (92 members) and Department 1 (65 members). These departments may play a crucial role in the communication network as they have more individuals. In contrast, some departments, such as Department 18 and Department 33, have only one individual each.

**1.2 Degree Distribution Analysis**

Figures 5 illustrates the in-degree (left-hand visualization) and out-degree distributions (right-hand visualization) within the email communication network. Both distributions exhibit a downward trend with a long-tail characteristic, which is a common feature in social and communication networks. In the in-degree distribution (Figure 5, left-hand), most individuals receive only a few emails. In contrast, a small number of nodes (such as department heads, administrators, or key decision-makers) receive a very high number of emails, contributing to the long tail of the distribution. Similarly, the out-degree distribution (Figure 5 right-hand) reveals that the majority of users send only a limited number of emails. At the same time, a small group of individuals exhibit significantly higher out-degree values. These high out-degree nodes may correspond to administrative assistants or automated email systems that broadcast messages to multiple recipients. The presence of these power-law-like distributions suggests an uneven communication structure, where a few key players manage or facilitate most of the institution's email exchanges while the majority of researchers maintain limited direct interactions within the network.

A graph of a graph

AI-generated content may be incorrect.

Figure 5: Degree Distribution

**1.3 Inter-Departmental Interactions and Communication Patterns**

The analysis of inter-departmental email exchanges reveals key insights into communication flow within the research institution. Figure 6 displays the top five departments based on total emails received (in-degree) and emails sent (out-degree). Department 4 is the most active in sending and receiving emails, followed by Departments 14, 36, 21, and 1 for in-degree, and Departments 36, 14, 21, and 7 for out-degree.

A green and purple bar graph

AI-generated content may be incorrect.

Figure 6: Top 5 Departments by Emails Received and Sent

The ranking suggests that these departments play a central role in institutional communication, possibly serving as administrative or leading departments in research projects. Additionally, when examining individual email activity, the top five most active individuals in both in-degree and out-degree belong to Department 36, reinforcing its importance in communication dynamics.

Furthermore, Figure 7 presents the Out-Degree/In-Degree ratio, which helps classify departments based on whether they predominantly send or receive emails. Departments above the red threshold (ratio > 1) send more emails than they receive, aligning with administrative leadership roles, coordination, or project management. In contrast, departments below the threshold (ratio < 1) receive more emails than they send, suggesting they function as recipients of directives, data requests, or administrative updates. These may include support teams, lab-based research groups, or postdoctoral researchers who primarily receive guidance from senior faculty.

A graph of a number of people

AI-generated content may be incorrect.

Figure 7: Departments by Communication Ratio

Figure 8 ranks department pairs based on the number of emails exchanged, highlighting the strongest interdepartmental interactions. The most active communication occurs between Department 4 and Department 36, followed by Department 4 and Department 5, suggesting close collaboration between the individuals working in these departments. Other highly interactive pairs include Departments 21 and 22, 15 and 36, and 0 and 4, indicating potential research collaborations, shared projects, or organizational coordination. Notably, Department 36 appears frequently in the top interactions, reinforcing its central role in institutional communication. These patterns show how departments work together, share important information, and build collaboration networks within the institution.

A graph of purple bars

AI-generated content may be incorrect.

Figure 8: Interdepartmental Email Communication

**2. Path Analysis: Eccentricity and Diameter**

To assess the efficiency of communication within the research institution’s email network, we analyzed eccentricity and diameter, which are key measures of how reachable different nodes are. Eccentricity represents the greatest shortest path from a node to any other node, highlighting how far each individual is from the most distant contact within the network. However, we encountered an issue: the overall graph is disconnected, meaning some nodes cannot reach others. Because of the existence of isolated nodes, computing the network diameter, which is the longest shortest path in the graph, was not feasible for the entire dataset. To resolve this, we focused on the largest connected component (LCC), which allowed us to use the built-in functions in the NetworkX library to calculate diameter and eccentricity.

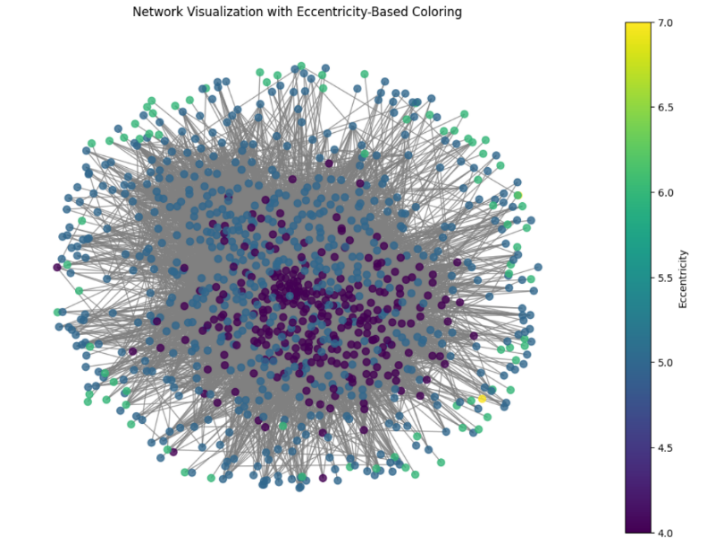


Figure 9: Eccentricity-Based Visualization

As shown in Figure 9, nodes are color-coded based on their eccentricity values. Lower eccentricity values (darker colors) indicate well-connected nodes that can reach others in fewer steps, whereas higher eccentricity values (lighter colors) identify nodes that are on the periphery of the network, requiring more steps to reach the rest. Moreover, nodes 634 and 846, with an eccentricity of 7, are the most distant nodes in the network, indicating they are relatively isolated. In contrast, nodes with lower eccentricity values (such as 4 or 5) are more central and efficiently connected.

Figure 10 further illustrates the distribution of nodes across different eccentricity levels. The majority of individuals have eccentricities of 4 or 5, meaning they can reach any other node within just a few email exchanges, suggesting a relatively compact communication structure. However, a smaller group of nodes has eccentricities of 6 or 7, meaning certain individuals are less accessible in the network, or we can say they are more distant from other individuals or researchers. This imbalance suggests that while most members can efficiently communicate, some areas of the network could benefit from improved connectivity to enhance information flow.

A graph of a number of nodes

AI-generated content may be incorrect.

Figure 10: Number of Nodes in Each Eccentricity Level

Overall, the analysis highlights that while the network is mostly efficient, the existence of peripheral nodes and disconnected components suggests opportunities for optimizing communication within the institution. Addressing these inefficiencies, such as encouraging more inter-departmental connections or ensuring better integration of isolated members, could improve overall collaboration and information dissemination.

**3. Identifying Key Players and Examining Group Centralities**

Our next step is to identify and examine the key actors in the network. For our purposes, the key actors are the most popular, or the most prestigious. The prestigious actors are those actors which receive extensive ties; i.e., the ones that receive the most emails. To find them, we count their in-degrees. From this, we were able to rank the ten nodes with the highest in-degree (Figure 11).

A graph of blue rectangular bars

AI-generated content may be incorrect.

Figure 11: Nodes with Highest In-degree

After this, we also determined how many emails each node received, and then ranked the distribution of in-degree for the nodes (Figure 12)–for example, how many nodes had an in-degree of 109, how many had an in-degree of 118, etc. From this, we also looked at the distribution of the departments. We were able to determine which departments had large numbers of nodes with an in-degree greater than 100 (Figure 13). This indicates which departments had large numbers of researchers or staff that were frequently in contact with other members of the institution and were therefore more prestigious.

A graph of a number of nodes

AI-generated content may be incorrect.

Figure 12: Distribution of Nodes with In-Degree >= 100

A graph of a number of individuals

AI-generated content may be incorrect.

Figure 13: Ranking of Departments by the Number of Influential Individuals

Our next method of identifying influential individuals started out by finding three different kinds of centrality–betweenness, closeness, and degree centrality. Degree centrality measures the number of direct connections a node has. Betweenness centrality measures how often a node appears on the shortest paths between other nodes. Closeness centrality measures how close a node is to all other nodes in the network, based on the shortest paths. All of these ways of determining centrality help tell us how well a node is positioned within the network.

After this, we also looked at group centralization. This allowed us to look at the centralization of entire departments at once, rather than at the individual members. From this we were able to determine that the groups mostly had a mixture of centralization and decentralization without a strict hierarchy. A few individuals may dominate communication, but others still contributed. Departments with high centralization (Figure 14) likely have a more structured communication pattern where certain individuals play key roles, whereas departments with lower centralization (Figure 15) likely have more collaborative natures.

A graph of green squares

AI-generated content may be incorrect.

Figure 14: Top Departments with Highest Group Centralization

A graph showing the number of groups of individuals

AI-generated content may be incorrect.

Figure 15: Bottom 5 Departments with Lowest Group Centralization

**5. Information Flow and Reciprocity**

**5.1 Reciprocity**

Reciprocity in the network, measured at 0.711, indicates a generally high level of mutual communication, suggesting that most individuals tend to reply to each other’s emails. Nevertheless, 28.88% of interactions remain one-sided, pointing to potential hierarchical or unbalanced communication patterns across certain departments or between specific individuals. These one-way exchanges may reflect top-down directives or mass emails that do not invite a response, potentially exacerbating uneven interactions in the network. While the overall reciprocity value signals a broadly collaborative environment, focusing on the remaining one-sided exchanges could mitigate hierarchical imbalances and encourage more inclusive, two-way communication.

**5.2 Transitivity**

Transitivity, measured at 0.22, reflects the extent to which individuals with a common contact also communicate directly. This low value indicates the network is fragmented: researchers may share contacts (e.g., through departments or projects) yet seldom engage in direct communication, leading to isolated subgroups or silos. Although mutual contacts exist, the lack of direct interaction curtails collaboration, highlighting limited cross-subgroup engagement. Consequently, this low transitivity underscores the need for strategies that encourage more direct exchanges between researchers, thereby reducing fragmentation and enhancing overall cohesion.

**5.3 Community Detection**

Community detection was carried out using the Girvan-Newman algorithm, which progressively removes edges with the highest betweenness centrality. The network, however, was highly connected, with 986 of 1005 nodes belonging to a single giant component, making it difficult to isolate distinct communities. Nodes on the network’s periphery were removed early in the process, but strong bridging edges maintained the core’s cohesion, preventing clear separations; recalculating betweenness centrality at each step was also computationally demanding.

In such a dense environment, Girvan-Newman’s limitations became clear: modest edge removals did not disrupt the extensive connections among most nodes. As a result, only a few significant clusters could be identified, suggesting that alternative methods—especially those leveraging departmental attributes or employing different detection algorithms—might yield more revealing insights into subgroups in a network where a handful of critical edges help maintain an otherwise unified structure.

**6. Node Removal and Network Robustness**

Node removal analysis examines how resilient the network is to the loss of specific, influential nodes. By removing a set number or percentage of nodes—based on different centrality measures or at random—we can observe changes in the network’s connected components, largest connected component size, and an overall network robustness score. Table 2 (below) summarizes these outcomes for four removal strategies, each highlighting different aspects of node importance.

Table 2: Summary of Results for Node Removal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Removal Strategy** | **Count of Nodes Removed** | **Connected Components** | **Largest Connected Component** | **Network Robustness Score** |
| **Degree Centrality** | 50 | 47 | 0.9045 | 0.0192 |
| **Closeness Centrality** | 50 | 40 | 0.9114 | 0.0228 |
| **Betweenness Centrality** | 50 | 55 | 0.8965 | 0.0163 |
| **Random Removal** | 5% of total nodes | 21 | 0.9303 | 0.0164 |

Removing 50 nodes with the highest degree centrality (those with the most direct connections) led to 47 connected components, leaving the largest connected component at 0.9045 of its original size. The overall network robustness score of 0.0192 indicates moderate fragmentation, although this strategy had less impact than betweenness-based removal.  
Targeting 50 nodes with the shortest average path to all others (closeness centrality) resulted in 40 connected components and a largest connected component fraction of 0.9114. While Betweenness-based removal focuses on 50 “bridge” nodes, generating 55 connected components and shrinking the largest connected component to 0.8965. With a robustness score of 0.0163, it causes the most severe fragmentation, underscoring these nodes’ critical role in linking otherwise distant subgroups.

Finally, removing 5% of nodes at random had the least disruptive effect, producing 21 connected components and a largest connected component fraction of 0.9303. Although the robustness score of 0.0164 is comparable to betweenness removal, the network remained more intact because there was no specific targeting of bridge nodes.

Key Takeaways:

* Betweenness-based removal inflicts the greatest damage, confirming that “bridge” nodes are indispensable for network cohesion.
* Degree and Closeness-based removals cause moderate fragmentation but isolate fewer components than betweenness removal.
* Random removal has minimal disruption, indicating the network can tolerate losing peripheral or non-bridge nodes without major breakdowns.
* Overall, these findings emphasize the network’s vulnerability to losing highly central connectors and highlight the need for diversifying communication pathways. With multiple routes available for information flow, the removal or departure of a few key individuals is far less likely to fragment the network.

**7. Hierarchical Nodes**

Since other methods for detecting hierarchical structures in communication networks are computationally expensive, we pursued an alternative approach that focuses on individuals who receive significantly more emails than they send. Specifically, our threshold flags nodes whose in-degree (emails received) is at least three times their out-degree (emails sent) and who have received at least ten emails in total. This approach identifies individuals whose communication patterns strongly suggest managerial or administrative roles.

The results reveal two types of hierarchical nodes. The first group consists of individuals who exclusively receive emails and never send any, such as Node 1 (50 received, 0 sent), Node 130 (35 received, 0 sent), and Node 365 (89 received, 1 sent). These nodes are likely high-ranking decision-makers, administrative personnel, or notification accounts used for institutional announcements. The second group includes individuals who primarily receive emails but occasionally send responses, such as Node 208 (51 received, 8 sent), Node 421 (47 received, 15 sent), and Node 950 (45 received, 9 sent). These individuals may hold senior researcher positions, team leadership roles, or advisory roles, where they oversee projects but are not heavily involved in daily back-and-forth communication. This method provides an efficient way to infer hierarchy in institutional email exchanges without requiring complex structural modeling.

**8. Structural Equivalence**

Since exploring the nearly equivalent structures was computationally expensive, and we were limited on time, we decided to focus on analyzing structural equivalence in the email network for subsets of individuals who share identical connection patterns. This means finding the same set of nodes that have identical rows and columns in the sociomatrix. We constructed a reduced network by grouping these structurally equivalent nodes, as visualized in Figure 16. This reduced representation provides insight into hierarchical structures and information flow across departments. A total of 18 groups were identified as structurally equivalent, and then we partitioned the socio matrix and put all the other individuals in one group (Group 19 or G19).

A diagram of a network

AI-generated content may be incorrect.

Figure 16: Reduced Network

A key observation is that Group 19 (G19) serves as the primary sender of emails to nearly all other groups rather than a receiver. This suggests that G19 represents individuals who are responsible for disseminating information institution-wide, which makes sense because the people grouped in G19 are the researchers and people working in the research institution. Evidently, we G19 send emails to all other groups. The other groups, such as G1, 2, 3, 4, etc., may hold administrative roles. The widespread outward connections from G19 indicate that these individuals are responsible for reporting to the peripheral groups rather than engaging in two-way discussions. Notably, Group 14 (G14) has a directed edge to G19 but there is no directed edge from G19 to G14, meaning that G19 members receive emails from G14 but do not send emails back. The department associated with G14 is 21, which could indicate that these individuals hold roles that require sending institutional announcements such as policy updates, event invitations, or institutional guidelines.

Another interesting finding is that Groups G7 and G8 have two-way connections with G19, suggesting reciprocal communication. This means that these individuals not only receive emails from the primary senders in G19 but also send messages back, potentially indicating a supportive or coordination role where they receive information or email from the institution members and send some response back. Since G7 consists of members from departments 29, 34, and 4, they may be playing a coordination role, such as project managers. Additionally, several groups (e.g., G5, G6, G9, G10, etc.) contain members from multiple departments, showing that structurally equivalent roles exist across departmental boundaries. These individuals may perform similar administrative roles or lead their own departments.

In conclusion, this analysis reveals a hierarchical and role-based email communication pattern, where G19 functions as the primary sender of information, broadcasting emails institution-wide. Some groups, like G7 and G8, maintain reciprocal communication, while others, such as G14, primarily send information. This suggests that certain individuals or units are central to institutional communication and responsible for distributing important updates, directives, and coordination messages across different departments. However, more analysis is required to identify patterns to help uncover the institution's formal and informal communication channels, revealing the flow of information and identifying key players in organizational interactions.

**Recommendations**

Addressing email overload among key individuals, such as department heads and senior administrators, is essential to enhancing the efficiency of email communication within the institution. One effective way to manage emails is to use systems that help with this task. These systems can filter out unimportant emails, highlight urgent issues, and assign tasks quickly and easily. Additionally, more information and metadata are needed to investigate the role of automated emails in the research institution. An unusually high number of outgoing emails may be a result of these automated systems.

To reduce the high number of email exchanges, people working in the research institution should use other communication tools like Slack or Microsoft Teams. These platforms let researchers work together directly, cutting down on back-and-forth emails. They can also use Google Drive to share files and documents, which helps avoid unnecessary emails asking for them. Additionally, to strengthen network resilience, organizations should rely less on important nodes, especially those that play a central role in connecting other nodes, since losing them can disrupt the network.

Improving communication within and between departments is very important. Departments that do not interact much should be encouraged to work on shared projects. This will help team members collaborate better. Departments that have little communication with others should also be motivated to take part in interdisciplinary research projects. This will strengthen the institution and promote the sharing of knowledge. Additionally, institutions should ensure that key individuals do not have too many responsibilities. By delegating communication tasks and managing workloads better, they can reduce communication problems and improve overall productivity.

**References**

1. Christidis, P., & Gomez Losada, Á. (2019). Email based institutional network analysis: applications and risks. *Social Sciences*, *8*(11), 306.
2. De Choudhury, M., Mason, W. A., Hofman, J. M., & Watts, D. J. (2010, April). Inferring relevant social networks from interpersonal communication. In *Proceedings of the 19th international conference on World wide web* (pp. 301-310).
3. Kazienko, P., Musiał, K., & Zgrzywa, A. (2009). Evaluation of node position based on email communication. Control and Cybernetics, 38(1), 67-86.
4. Kolli, N., & Narayanaswamy, B. (2013). Analysis of e-mail communication using a social network framework for crisis detection in an organization. *Procedia—Social and Behavioral Sciences*, *100*, 57-67.
5. Leskovec, J., & Krevl, A. (2014, June). SNAP Datasets: Stanford large network dataset collection. SNAP. http://snap.stanford.edu/data/email-Eu-core.html
6. Malinka, K., & Schäfer, J. (2009, May). Development of social networks in email communication. In *2009 Fourth International Conference on Internet Monitoring and Protection* (pp. 34-38). IEEE.
7. Michalski, R., Palus, S., & Kazienko, P. (2011). Matching organizational structure and social network extracted from email communication. In *Business Information Systems: 14th International Conference, BIS 2011, Poznań, Poland, June 15-17, 2011. Proceedings 14* (pp. 197-206). Springer Berlin Heidelberg.
8. Uddin, S., & Jacobson, M. J. (2013). Dynamics of email communications among university students throughout a semester. *Computers & Education*, *64*, 95-103.
9. Yee, J., Mills, R. F., Peterson, G. L., & Bartczack, S. E. (2005). Automatic generation of social network data from electronic-mail communications.
10. Weare, C., Loges, W. E., & Oztas, N. (2007). Email effects on the structure of local associations: A social network analysis. *Social Science Quarterly*, *88*(1), 222-243.