Project Report

Temporal Email Communication Network Analysis using Graph Mining Techniques

(email-Eu-core-temporal dataset — SNAP)

DEPARTMENT OF COMPUTER SCIENCE

KESHAV MAHAVIDYALAYA

UNIVERSITY OF DELHI

Submitted by

NAME : Anubhav Bharti

ROLL NO: 22035570084

SUBJECT: Social Network Analysis

Submitted to

Ms. Astha Goyal

Department of Computer Science

Keshav Mahavidyalaya

1. Introduction

The dataset used in this project is the “email-Eu-core-temporal” network from the “Standford Network Analysis Project (SNAP)”. It contains records of internal email communication within a European research institution. Each entry in the dataset represents a directed, time stamped interaction, where a sender “u” sends an email to a receiver “v” at a time “t”.

Dataset Description

Dataset: email-Eu-core-temporal

Nodes	986
Temporal Edges	332334
Edges in static graph	24929
Time span	803 days

Objectives:

- Apply two graph mining techniques on the dataset.
- Understand structural and temporal patterns in social interactions.
- Interpret findings in the context of communication behaviour.

2. Graph Mining Techniques

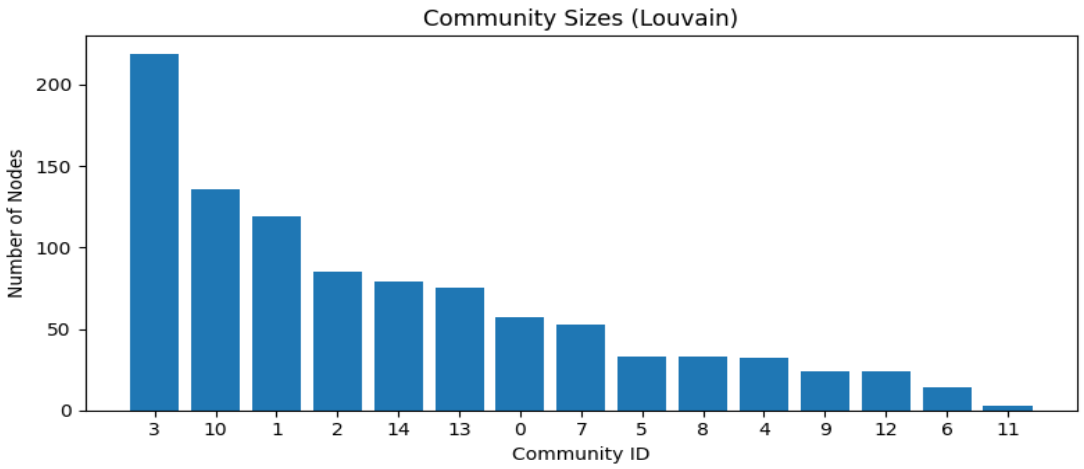
This project includes two major graph mining analyses:

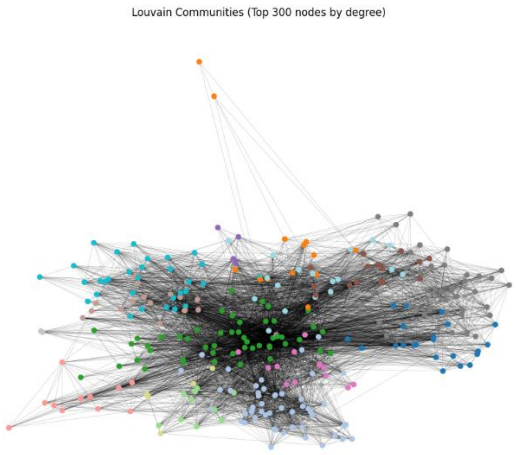
(A) Community Detection

(B) Centrality Analysis

2.1 Community Detection (Using Louvain)

Community detection helps identify groups of individuals who communicate more frequently among themselves than with others.






```
PS C:\Users\bhart> & C:/Users/bhart/AppData/Local/Programs/Python/Python313/python.exe "d:/SEM VII/SNA/Project/Project.py"
Loaded graph:
Nodes: 986
Edges: 16064
Modularity: 0.6852380502952331
Detected communities: 15

Top communities (by size):
 Community ID  Size   Top Nodes (by degree)
            3   219  746, 135, 214, 90, 742
           10   136 977, 506, 418, 107, 738
            1   119 168, 586, 356, 912, 416
            2    85 951, 534, 343, 832, 622
           14    79  718, 61, 423, 189, 595
           13    75 629, 891, 280, 288, 530
            0    57 987, 540, 231, 245, 955
            7    53 178, 502, 569, 605, 862
            5    33  684, 499, 558, 48, 844
            8    33  915, 316, 165, 428, 67
            4    32  927, 62, 542, 646, 854
            9    24  98, 401, 454, 415, 429
```

The Louvain method dectected 15 communities with a modularity score of 0.6852, indicating a strong and well-defined community structure.

The top five communities by size and their highest-degree nodes are:

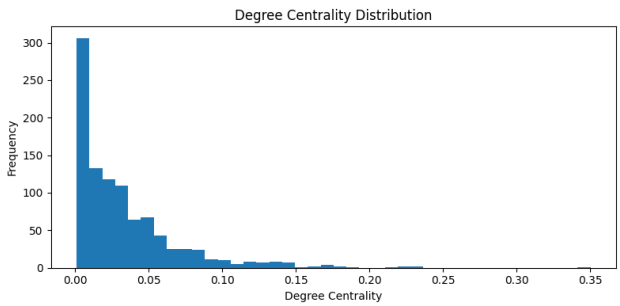
- Community 3 (219 nodes): top nodes — 746, 135, 214, 90, 742
- Community 10 (136 nodes): top nodes — 977, 506, 418, 107, 738
- Community 1 (119 nodes): top nodes — 168, 586, 356, 912, 416
- Community 2 (85 nodes): top nodes — 951, 534, 343, 832, 622
- Community 14 (79 nodes): top nodes — 718, 61, 423, 189, 595

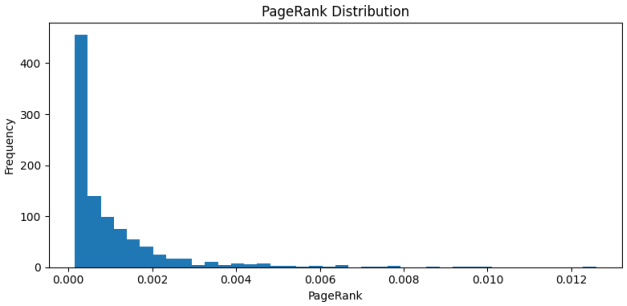
2.2 Centrality Analysis

Centrality measures identify influential individuals in the network.

Centrality Metrics Used:

- Degree Centrality – nodes with many connections
- Betweenness Centrality – nodes acting as bridges
- Eigenvector Centrality – nodes connected to other well-connected nodes
- PageRank – importance based on directed interactions






```
Nodes: 986
Edges: 16064

Top 10 Degree Centrality Nodes:
90 0.350253807106599
120 0.23553299492385787
772 0.23451776649746195
214 0.22233502538071068
951 0.21928934010152284
159 0.21725888324873097
607 0.18578680203045686
362 0.18071065989847718
890 0.17766497461928935
61 0.1736040609137056
```

```
Top 10 Betweenness Centrality Nodes:
744 0.06438068851373965
91 0.054372670806553114
700 0.048281440196174086
362 0.040416600645632336
2 0.037411725024999676
890 0.03725283696066453
772 0.03273796220708151
951 0.03251536193545348
454 0.023231306364861867
90 0.02283795500385722
```

```
Top 10 Eigenvector Centrality Nodes:
947 0.7070874844873659
774 0.7068630563271294
568 0.01425653392101515
450 0.00907991288697257
40 0.004900064636681124
586 0.004889255367650499
187 0.0028307488585266444
540 0.0023909578790113936
466 0.0020318706363259757
105 0.0018881447063819863
```

```
Top 10 PageRank Nodes:
987 0.01258559969118469
168 0.009881380414233405
586 0.009539299180858026
178 0.009470320922223604
629 0.008785363020981679
135 0.007906533329397539
356 0.00781714913007344
746 0.0075526389758943414
718 0.007191546766491148
915 0.006584327022967337
```

Four centrality measures (Degree, Betweenness, Eigenvector, and PageRank) computed to identify important individuals in the email network.

Degree Centrality identifies highly active communicators. Nodes 90, 120, and 772 have the highest number of direct connections.

Betweenness Centrality highlights nodes that act as bridges. Nodes 744, 91, and 700 play key roles in connecting different parts of the network.

Eigenvector Centrality emphasizes nodes connected to other influential nodes. Nodes 947 and 774 dominate this measure, indicating their strong influence within tightly connected groups.

PageRank measures importance based on incoming links. Nodes 987, 168, and 586 rank highest, suggesting they receive communication from many important individuals.

The centrality results shows that influence and communication load are unevenly distributed, with a small set of nodes playing major structural and communication roles.

3. Conclusion

The analysis of the temporal email network shows a clear structural patterns within the organization. Community detection using the Louvain method produced 15 well-defined communities with a modularity of 0.6852, showing a strong clustering aligned with natural communication groups.

Centrality analysis shows that communication is not evenly distributed.
A few nodes, such as **90**, **744**, **947**, and **987**, play critical roles as active communicators, bridges between communities, or highly influential members within tightly connected sub-groups. These individuals likely coordinate information flow, connect different teams, or hold important administrative or organizational positions.

Overall, the network highlights important nodes (people) who handle most of the communication.

Dataset Reference

Leskovec, J., & Krevl, A.
“SNAP Datasets: Stanford Large Network Dataset Collection”
email-Eu-core temporal network
Available at:
https://snap.stanford.edu/data/email-Eu-core-temporal.html

Google Drive Project Folder:

https://drive.google.com/drive/folders/1uBdvLQoKbQYd_Pbni4ySuu_oeZ4vnQD9