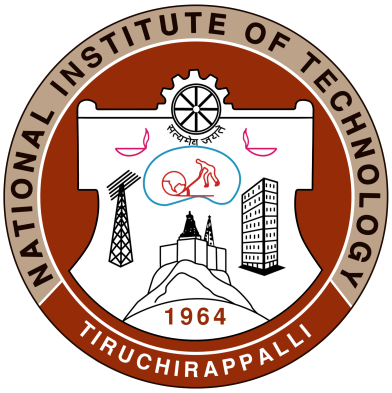
Social Network Analysis

Project Report



National Institute of Technology, Tiruchirappalli

Student	Priyansh Kumar Paswan
Roll Number	205124071
Professor	Dr. S.R. Balasundaram
Department	Computer Applications
Institute	National Institute of Technology, Tiruchirappalli

Tools used: Python (NetworkX, pandas, scikit■learn, matplotlib), Gephi.

1. Dataset Overview

Email-EU-core undirected graph. The analysis below summarizes structure and degree characteristics.

Metric	Value
Nodes	1005
Edges	16706
Density	0.0331
Avg clustering	0.3994
Connected	False

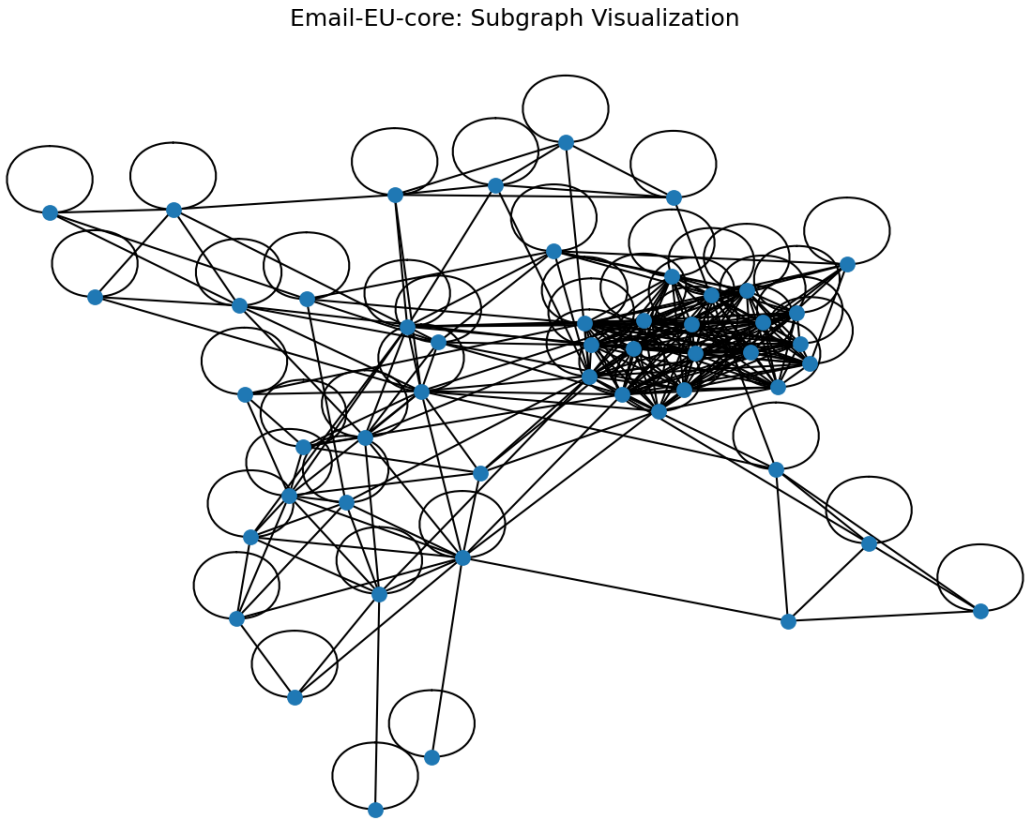


Figure 1: Subgraph visualization

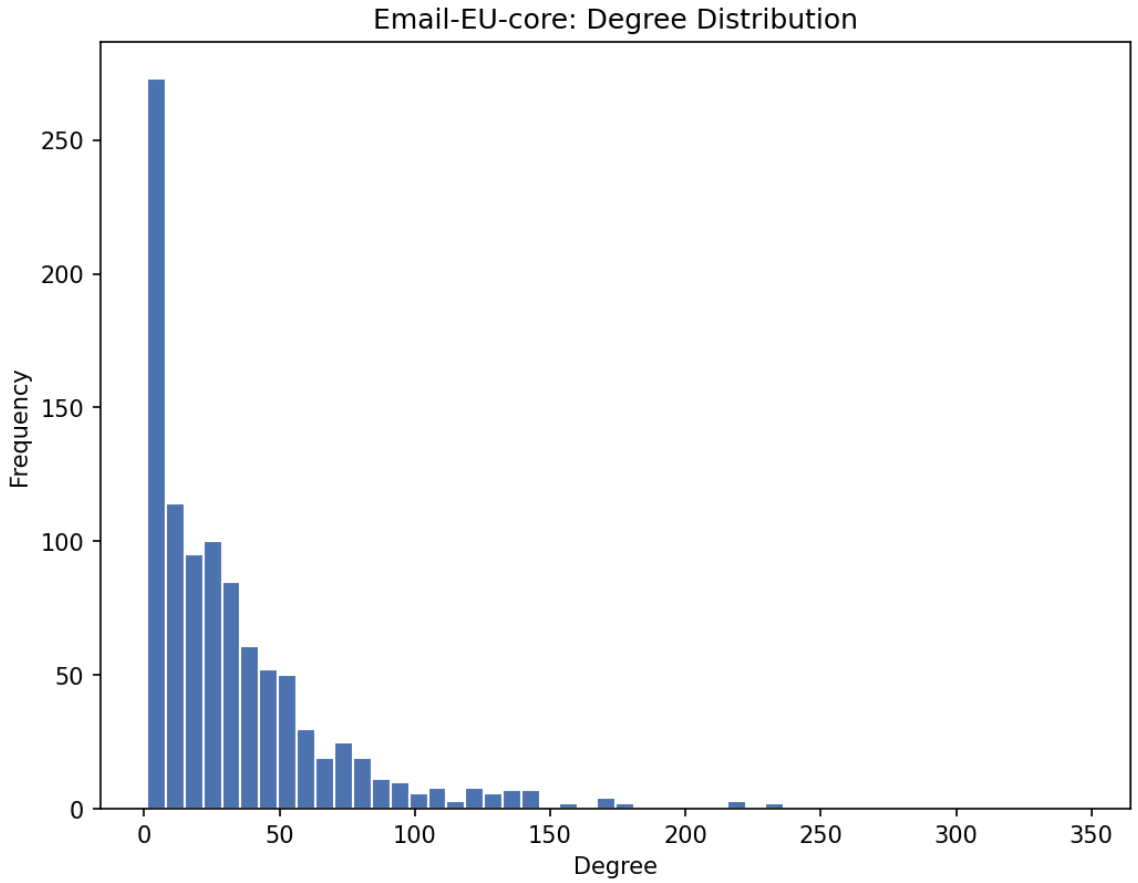


Figure 2: Degree distribution histogram

**** Optional Gephi screenshot: Global layout (ForceAtlas2), color by community (from node_classification), size by pagerank ****

2. Link Analysis (PageRank & Eigenvector)

Methods: PageRank for importance via random walks; Eigenvector centrality for importance via influential neighbors.

Top-10 by PageRank

node	pagerank	eigenvector	degree
160.0	0.0090709484950265	0.1658461052562613	347.0
121.0	0.0060687854256318	0.1484213057175369	234.0
82.0	0.0060307053368408	0.145251809177294	233.0
107.0	0.0058380960432493	0.139876476623585	221.0
86.0	0.0057215196123223	0.1122173025767131	218.0
62.0	0.0054316159798508	0.1314982021050847	216.0
5.0	0.0049141637345112	0.0794635644309267	171.0
13.0	0.004589938693459	0.0856933490779659	180.0
166.0	0.0045516651162006	0.1103349118134628	177.0
434.0	0.0045327987830447	0.1253049110776443	185.0

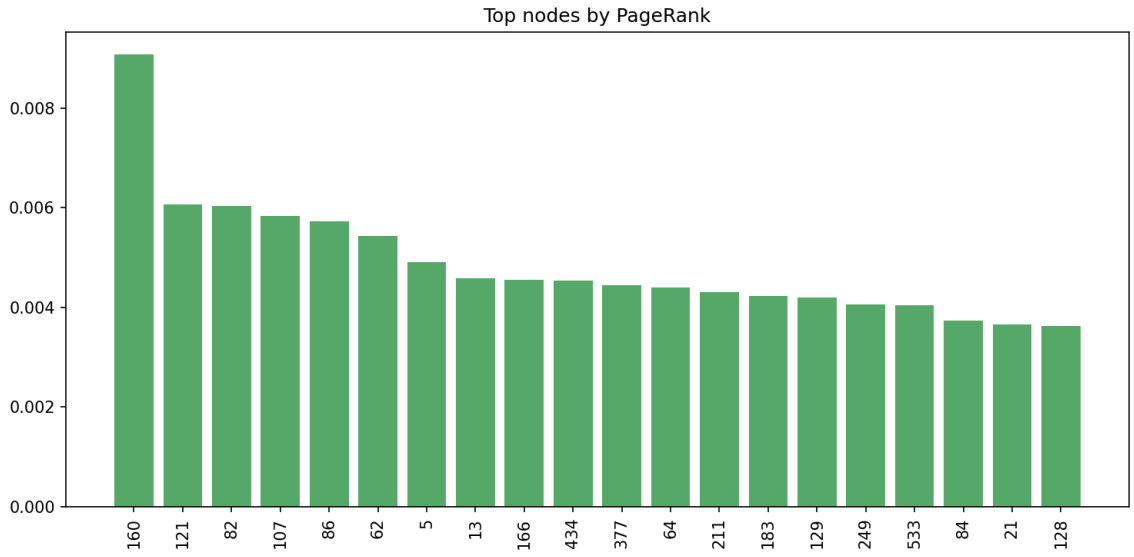


Figure 3: Top nodes by PageRank (bar chart)

**** Optional Gephi screenshot: Influence view — size by pagerank, color gradient by eigenvector ****

3. Node Classification (Label Propagation)

Method: Asynchronous label propagation to discover communities without ground truth labels.

Communities by size (Top 15)

community	size
0	969
19	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1

**** ADD GEPHI SCREENSHOT HERE: Community visualization — color by community (categorical), size by pagerank; layout: ForceAtlas2 ****

4. Influence Analysis (PageRank & Betweenness)

Methods: PageRank for global influence; Betweenness for brokerage across communities.

Top-10 influencers

node	pagerank	betweenness	degree
160.0	0.0090709484950265	0.0874147349363879	347.0
121.0	0.0060687854256318	0.0278415388258006	234.0
82.0	0.0060307053368408	0.0278807411351142	233.0
107.0	0.0058380960432493	0.0243403121826939	221.0
86.0	0.0057215196123223	0.0377885326911519	218.0
62.0	0.0054316159798508	0.0225098451925391	216.0
5.0	0.0049141637345112	0.0309946865452777	171.0
13.0	0.004589938693459	0.0235649895706901	180.0
166.0	0.0045516651162006	0.0176393735896517	177.0
434.0	0.0045327987830447	0.0154127096447369	185.0

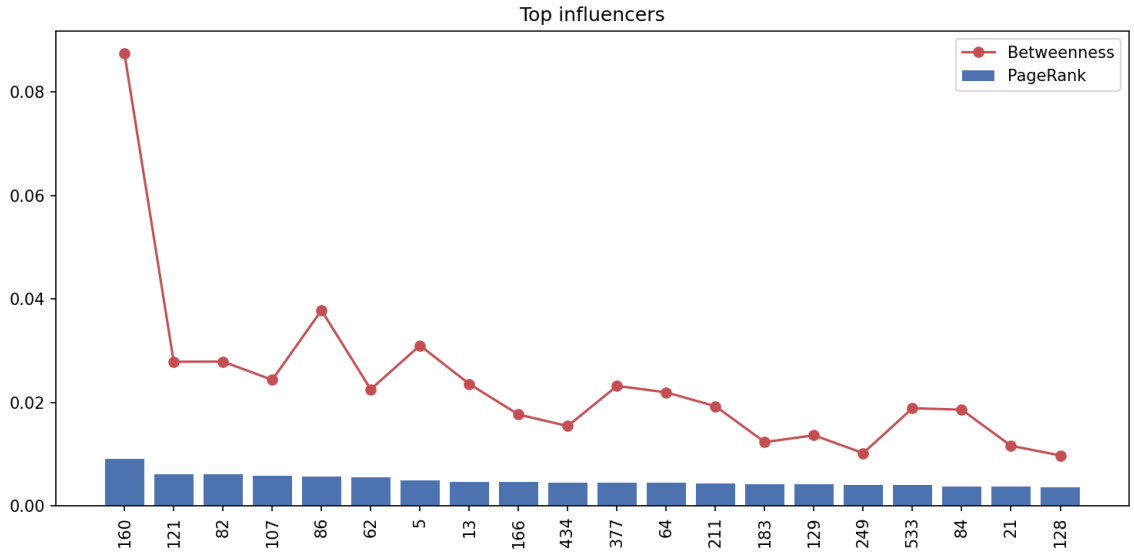


Figure 4: Top influencers (PageRank + Betweenness)

**** ADD GEPHI SCREENSHOT HERE: Influencers — size by betweenness or pagerank; optionally label top nodes ****

5. Link Prediction (Adamic-Adar, Jaccard, Preferential Attachment)

Procedure: Hold out 10% edges; score with three heuristics on the train graph; evaluate ROC/AUC.

AUC by metric

metric	auc
adamic_adar	0.9468674701384132
jaccard	0.935174248697092
pref_attach	0.8597603122394544

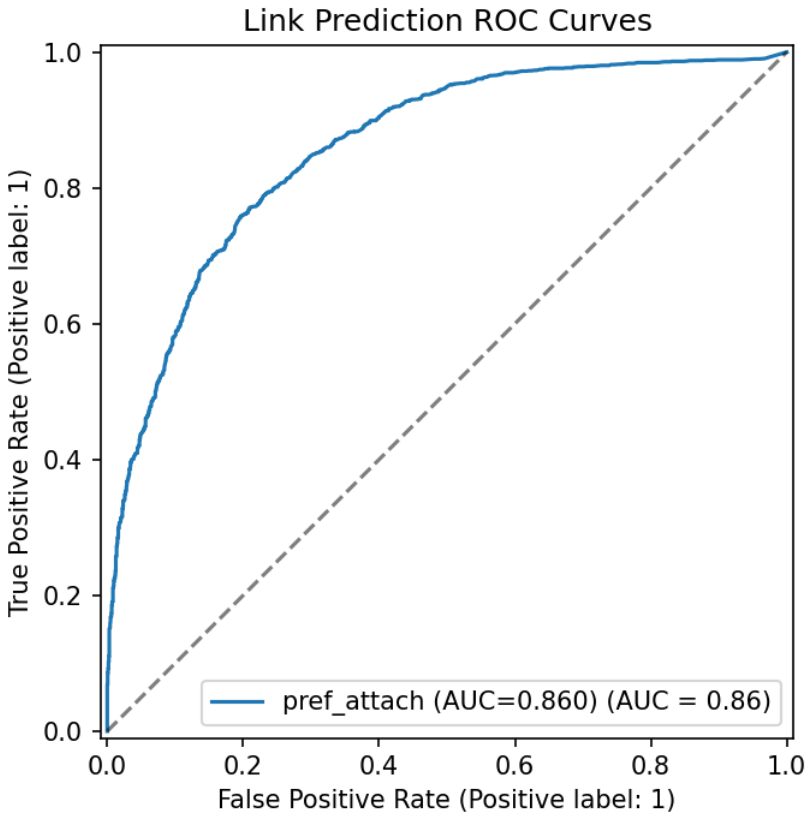


Figure 5: ROC curves for link prediction metrics

6. Anomaly Detection (IsolationForest on egonet features)

Features: degree, clustering, average neighbor degree, egonet edges; Model: IsolationForest (1% contamination).

Top-10 anomalous nodes

node	degree	clustering	avg_neighbor_degree	ego_edges	decision_function
160.0	347.0	0.0935119649477586	56.73198847262248	6177.0	-0.1259716934781769
121.0	234.0	0.1728989401403194	70.74358974358974	5056.0	-0.0815577059253059
82.0	233.0	0.1660831921701487	69.78111587982832	4828.0	-0.0814038916550311
107.0	221.0	0.1700389594068116	70.77828054298642	4462.0	-0.0651324610742785
62.0	216.0	0.1520336975121758	68.41203703703704	3856.0	-0.0565494375542041
86.0	218.0	0.1205857019810508	61.821100917431195	3194.0	-0.052512605438006
434.0	185.0	0.2011049060229388	76.41621621621621	3685.0	-0.0213113652596759
13.0	180.0	0.110899511204215	58.111111111111114	2082.0	-0.0128058498699874
5.0	171.0	0.1070019723865877	57.046783625730995	1840.0	-0.0072600496357142
882.0	2.0	1.0	262.0	5.0	-0.0025955469404698

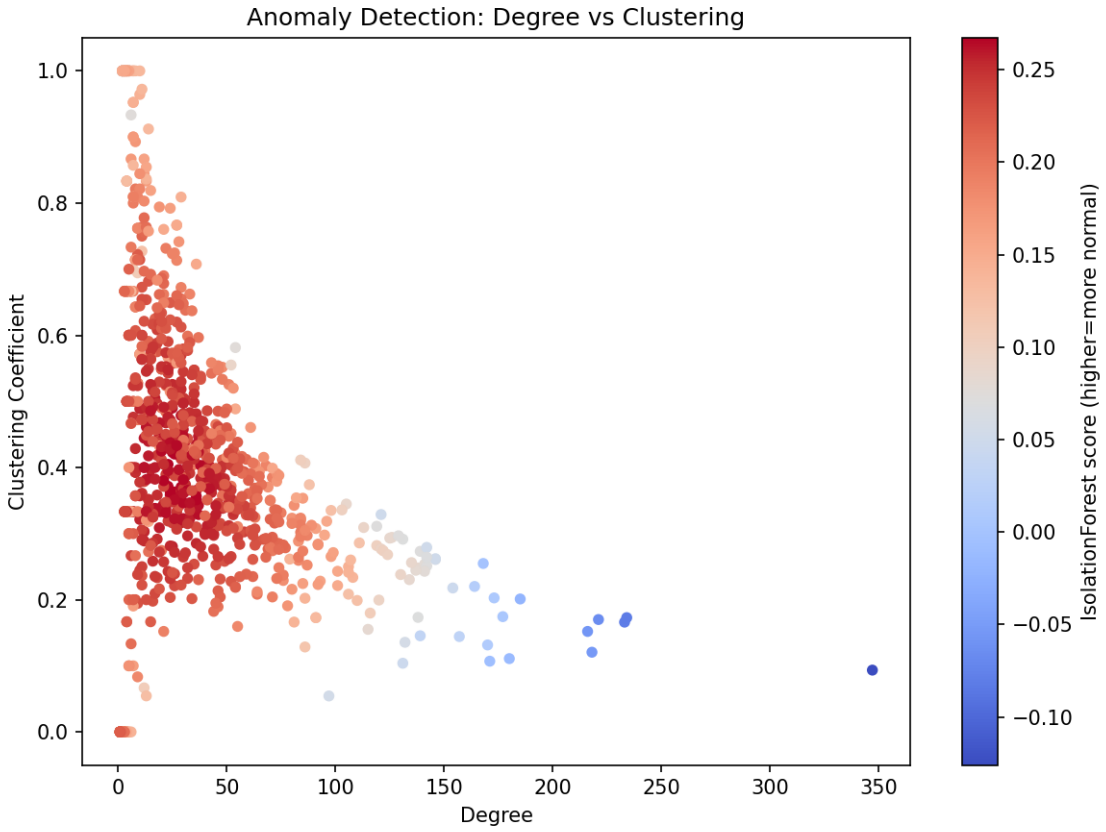


Figure 6: Anomaly scatter (Degree vs Clustering, colored by score)

**** ADD GEPHI SCREENSHOT HERE: Anomaly view — color by anomaly_score (continuous), highlight top outliers ****

7. Gephi Visualizations (What to include)

Open outputs/graph_attributes.gexf in Gephi and capture these views:

• Community view: Color by 'community' (categorical), size by 'pagerank'.

• Influence view: Size by 'betweenness' or 'pagerank'; optional color gradient by betweenness.

• Anomaly view: Color by 'anomaly_score' (continuous gradient).