Network Analysis of the Hanoi Citation Graph

This project analyzes a citation network of research papers in Hanoi.

Objectives:
- Examine network structure using graph metrics.
- Identify key papers through centrality measures.
- Detect communities of related papers.
- Visualize the network and metric distributions.

3. C

Top 5 by In-Degree:

Node 8609853337.0: 0.000037
Node 4813567397.0: 0.000037
Node 7488855891.0: 0.000037
Node 1905643476.0: 0.000037
Node 8341138742.0: 0.000037

Top 5 by Out-Degree:

Node 8341138742.0: 0.000037
Node 8609853337.0: 0.000037
Node 4813567397.0: 0.000037
Node 7488855891.0: 0.000037
Node 1905643476.0: 0.000037

Top 5 by PageRank:

Node 7226570628.0: 0.000025
Node 445252475.0: 0.000018
Node 10183503218.0: 0.000017
Node 12116337047.0: 0.000017
Node 8332773638.0: 0.000016

Top 5 by Betweenness Centrality:

Node 10308025429.0: 0.285480
Node 10308015689.0: 0.284101
Node 282895148.0: 0.210321
Node 10246832783.0: 0.210314
Node 12623288819.0: 0.205905

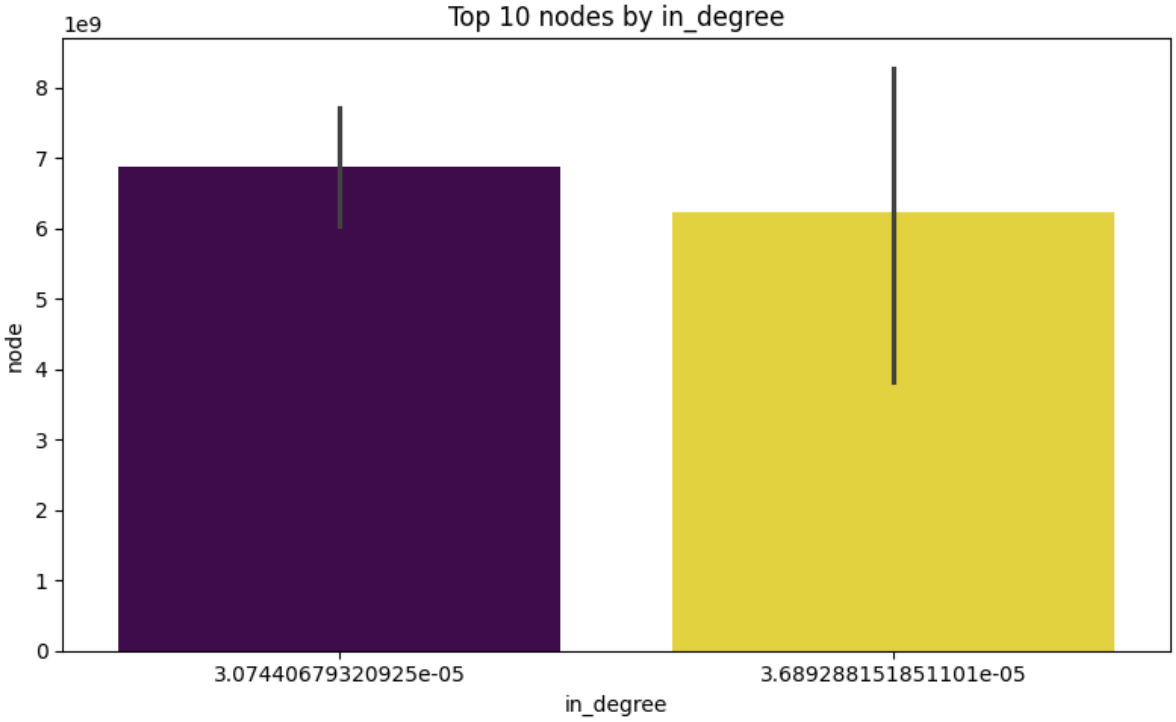
Top 5 by Hub Score:

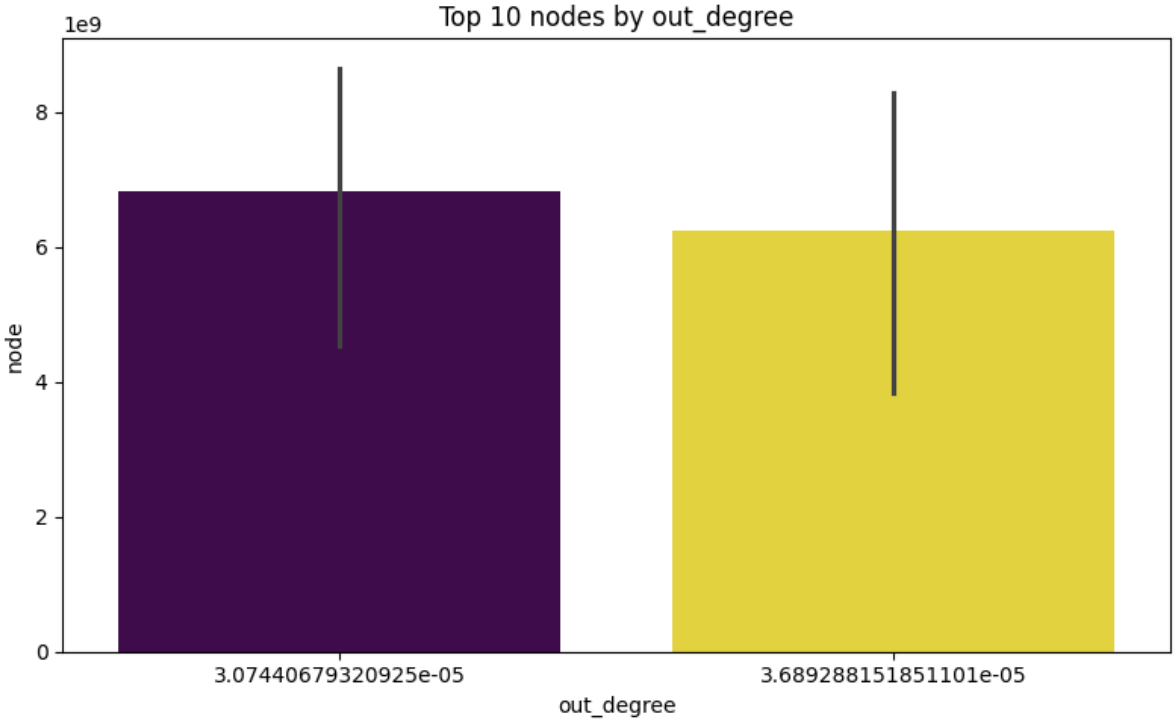
Node 4466994312.0: 0.084562
Node 3435245119.0: 0.083878
Node 3435239188.0: 0.083878
Node 4466994311.0: 0.082495
Node 3435245097.0: 0.082495

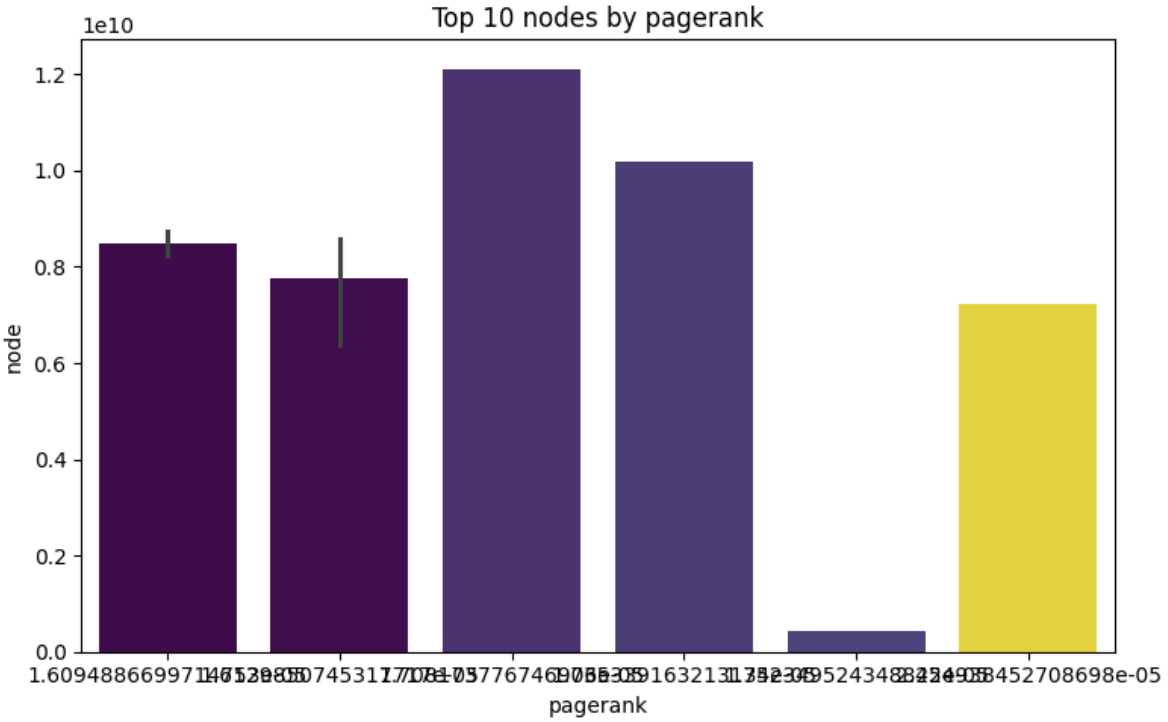
Top 5 by Authority Score:

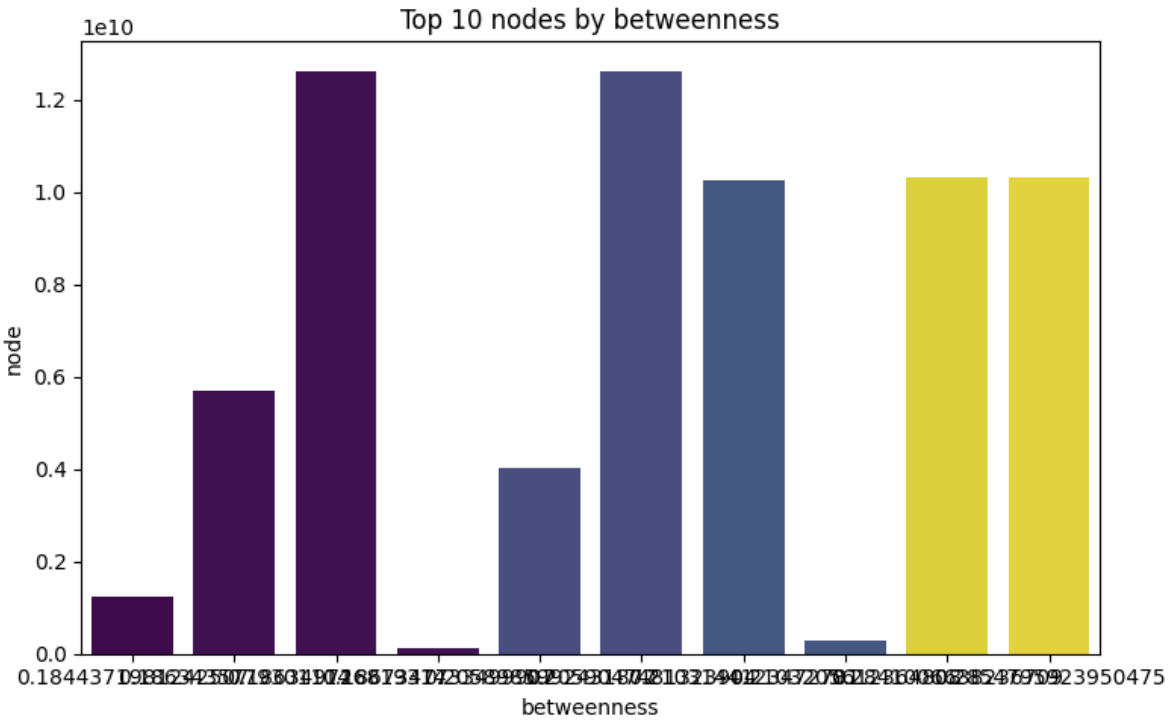
Node 4466994312.0: 0.084562
Node 3435239188.0: 0.083878
Node 3435245119.0: 0.083878
Node 3435245097.0: 0.082495
Node 4466994311.0: 0.082495

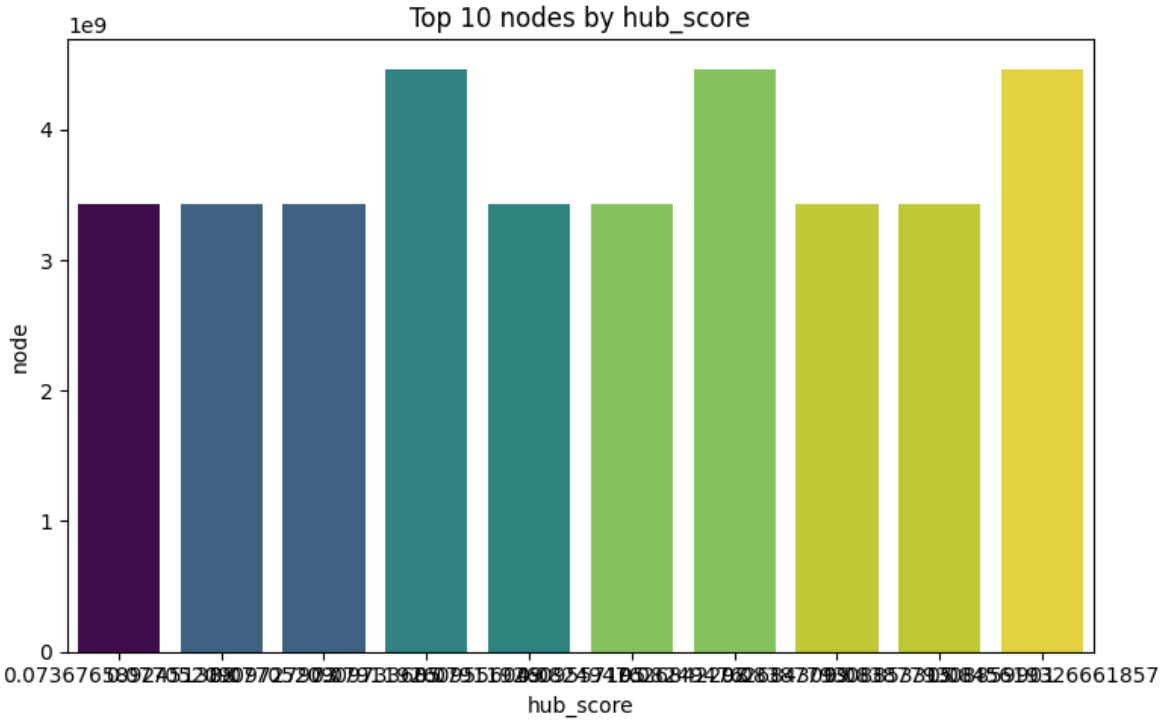
Centrality Visualizations

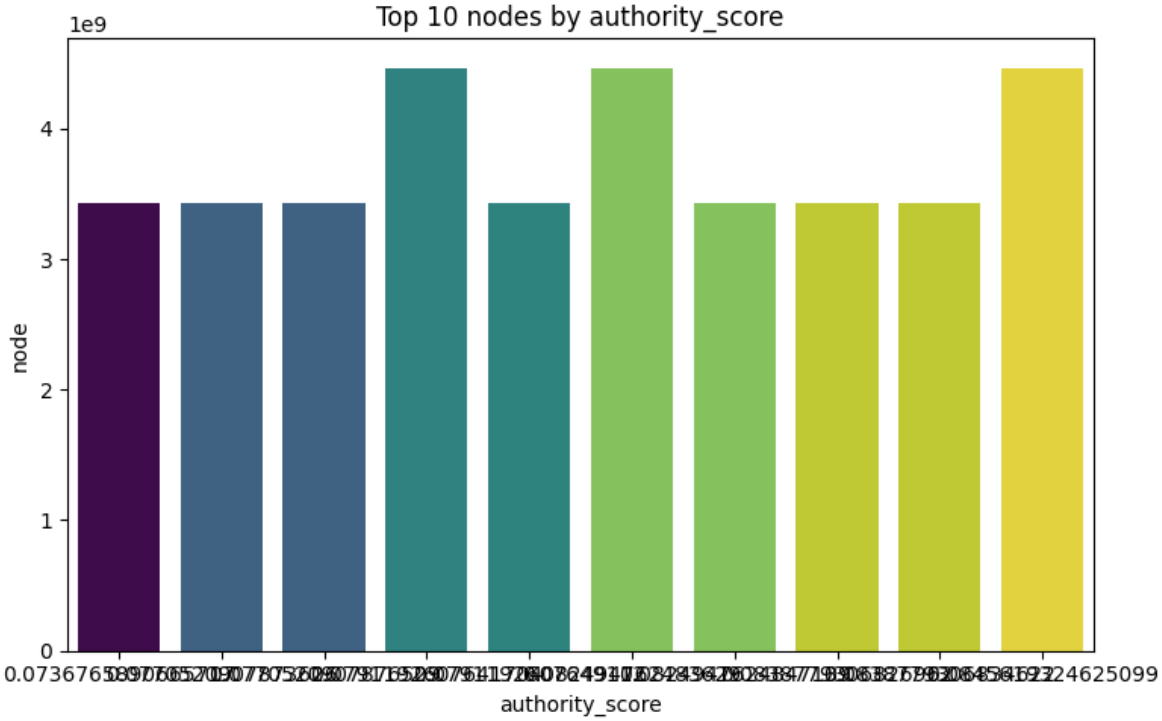


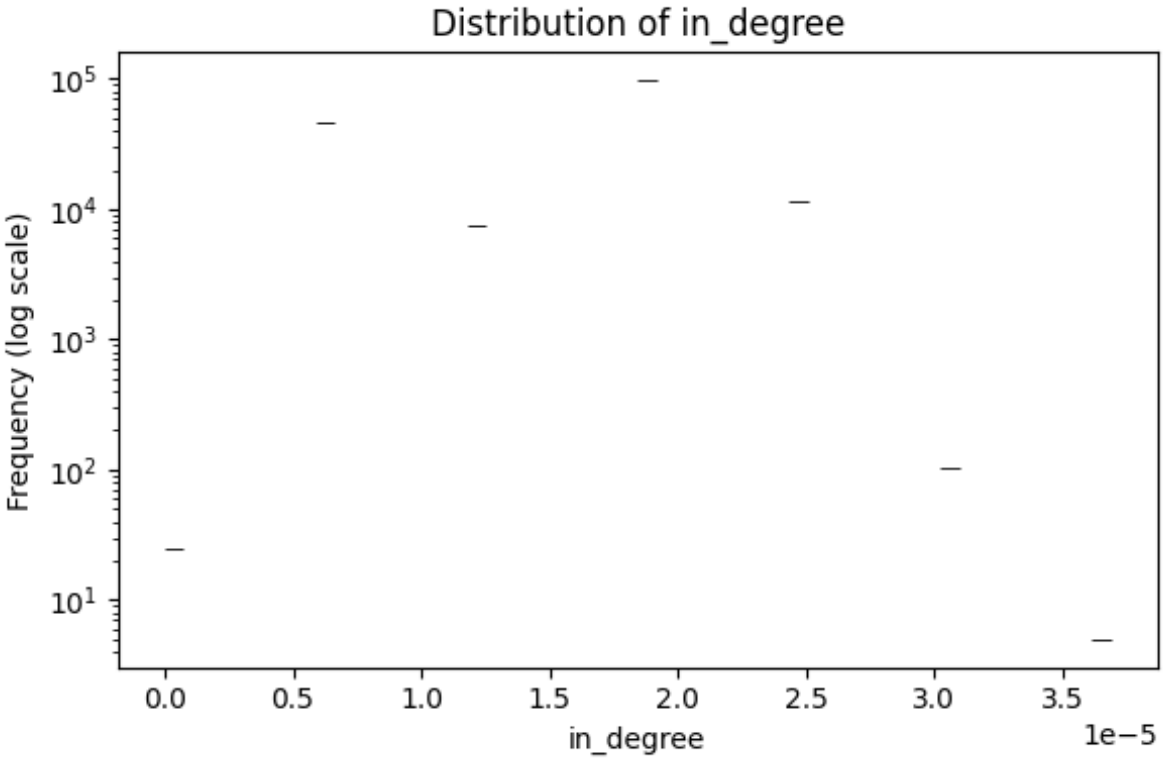


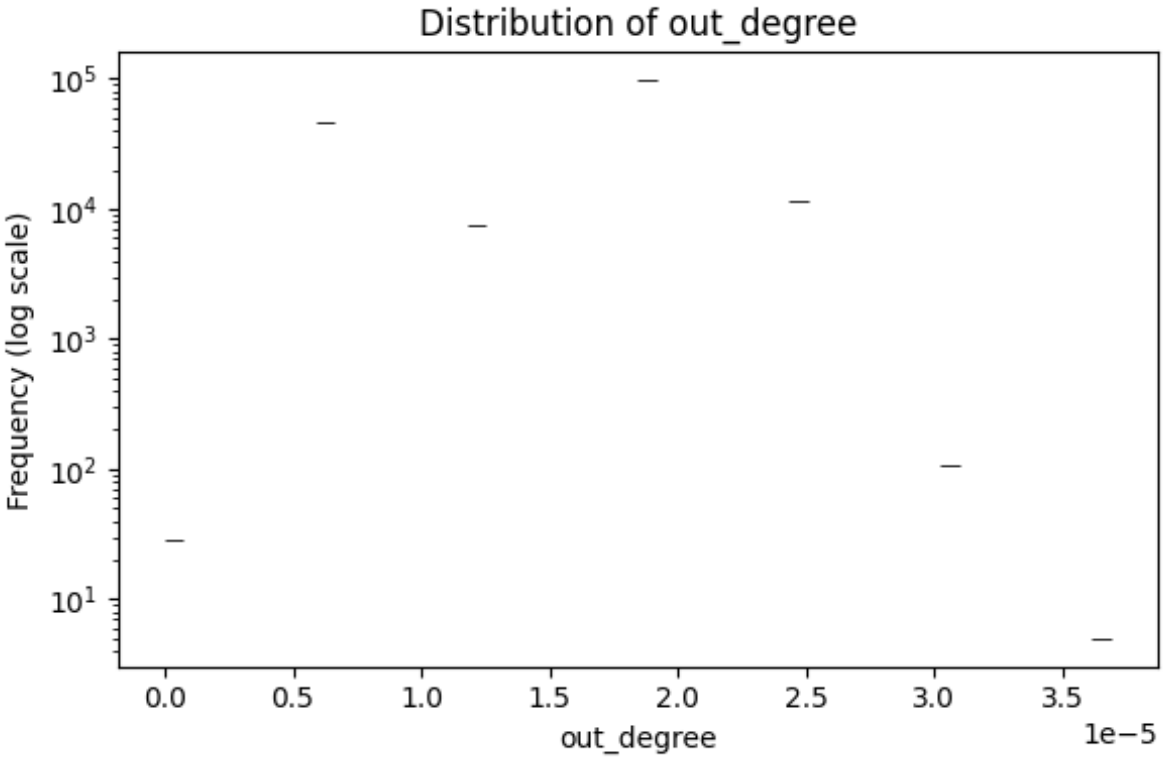

Top 10 nodes by pagerank
1e10
1.2
1.0
0.8
0.6
0.4
0.2
0.0
node
pagerank

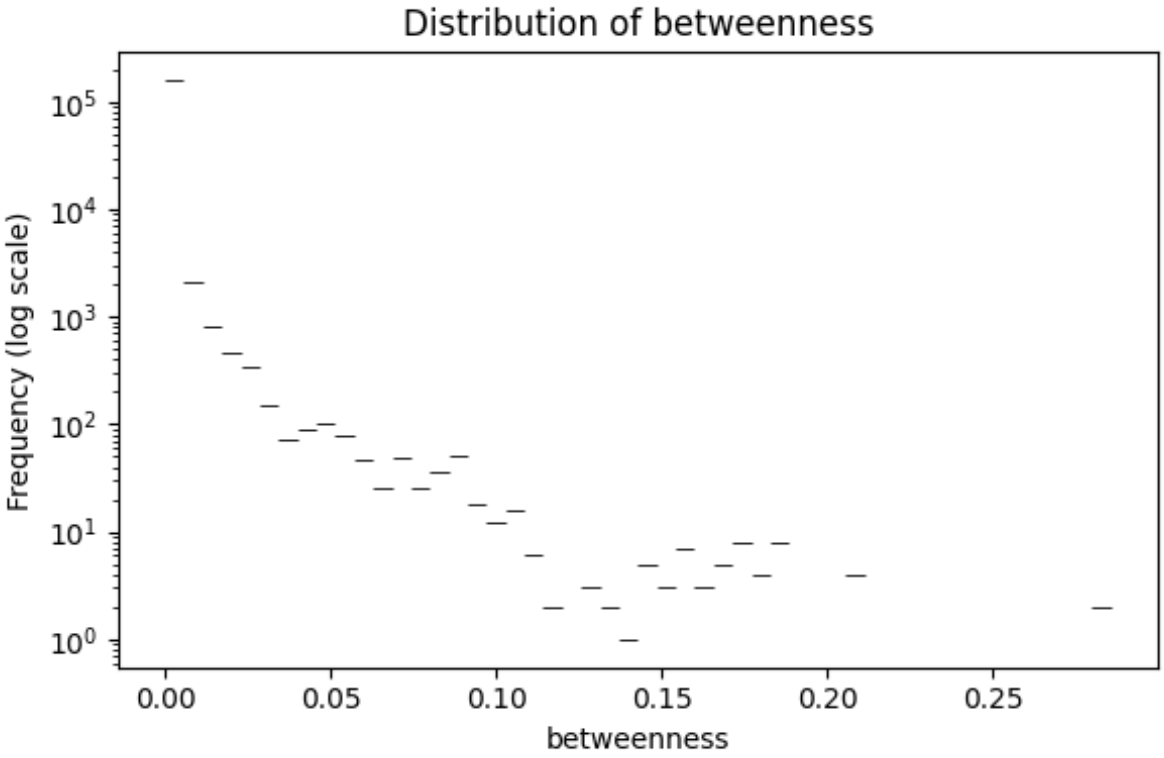

Top 10 nodes by betweenness
1e10
1.2
1.0
0.8
0.6
0.4
0.2
0.0
node
betweenness

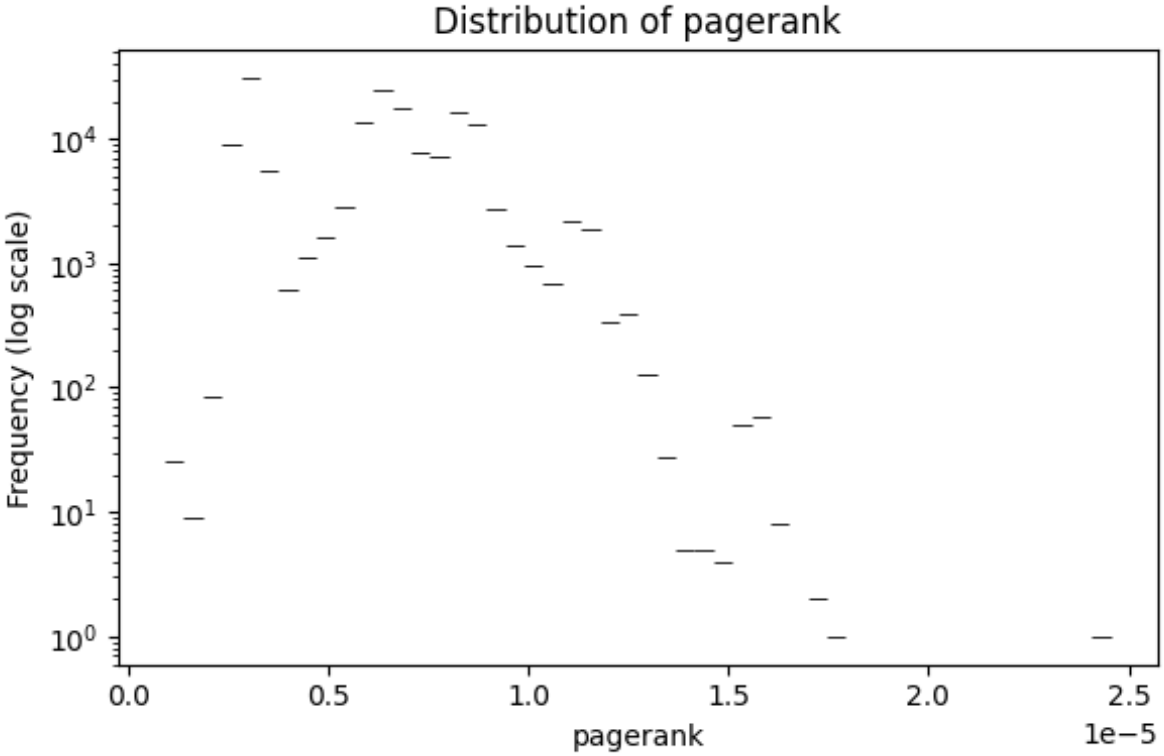

Top 10 nodes by hub_score
1e9
node
hub_score


Top 10 nodes by authority_score
1e9
node
authority_score

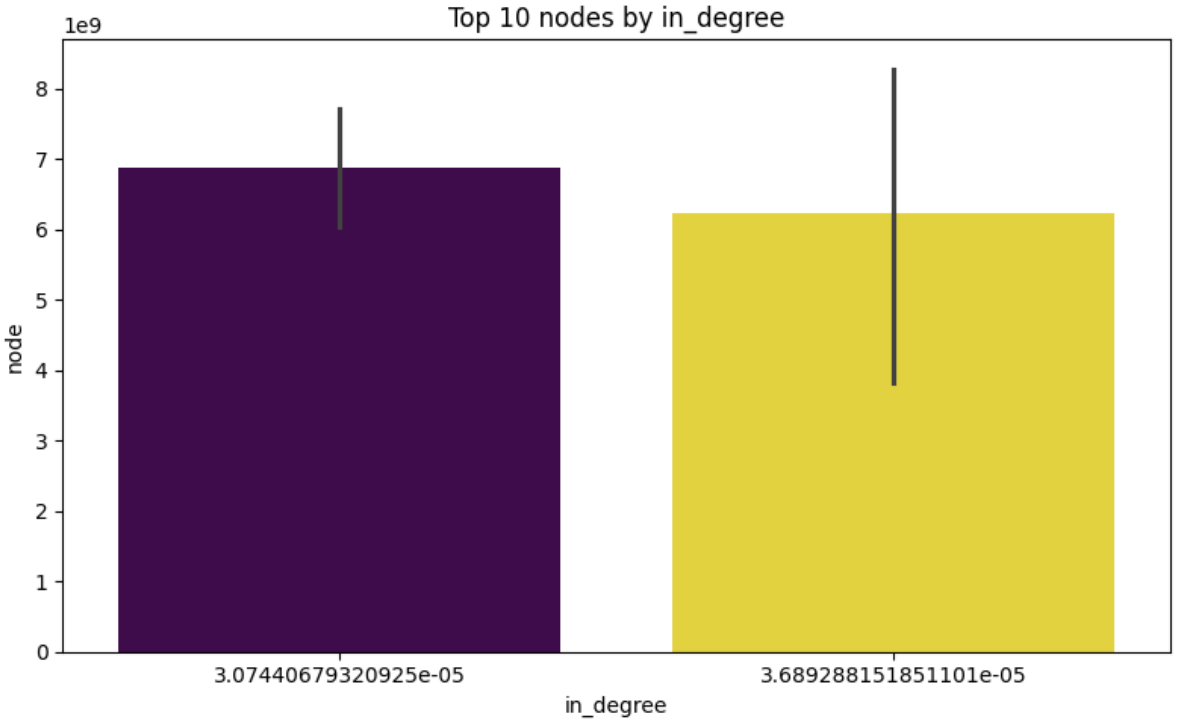

Distribution of in_degree
Frequency (log scale)
10^5
10^4
10^3
10^2
10^1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
in_degree
1e−5

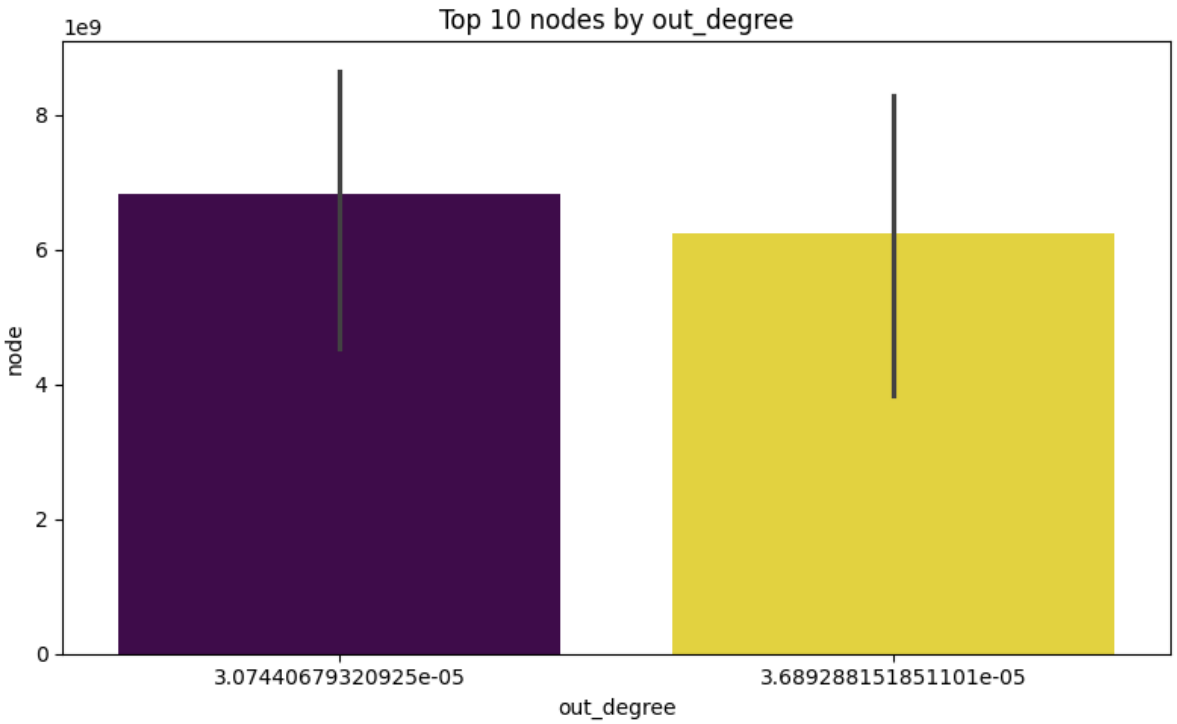

Distribution of out_degree
Frequency (log scale)
10^5
10^4
10^3
10^2
10^1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
out_degree
1e−5

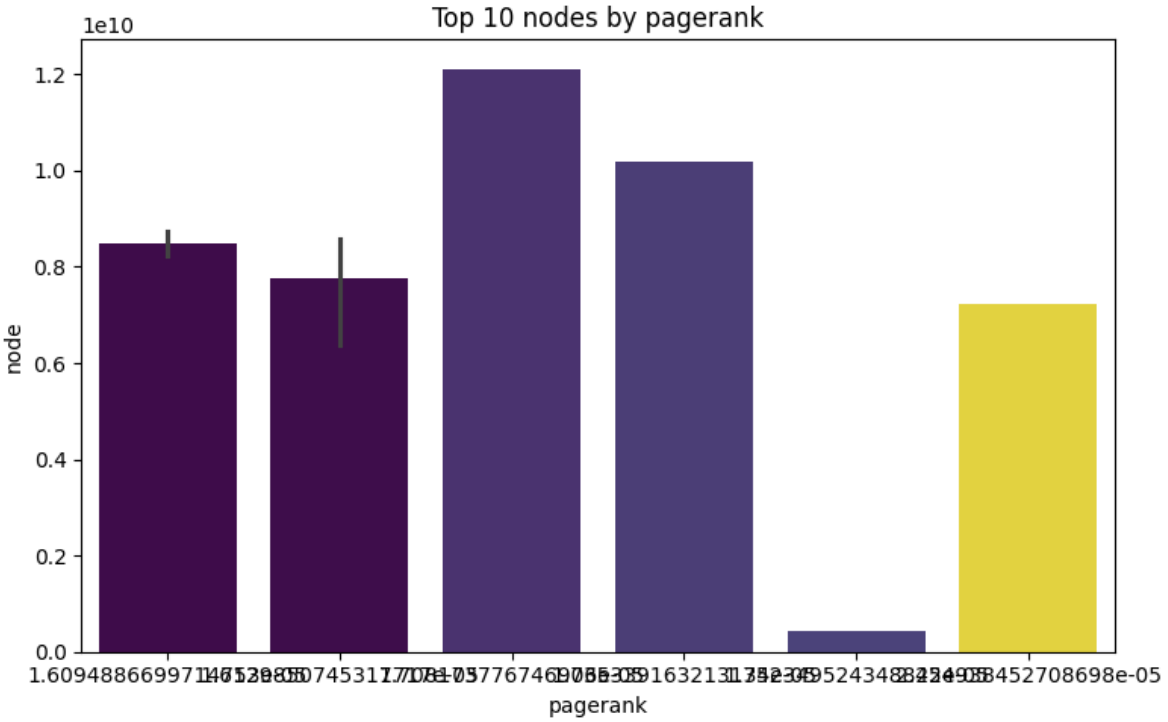


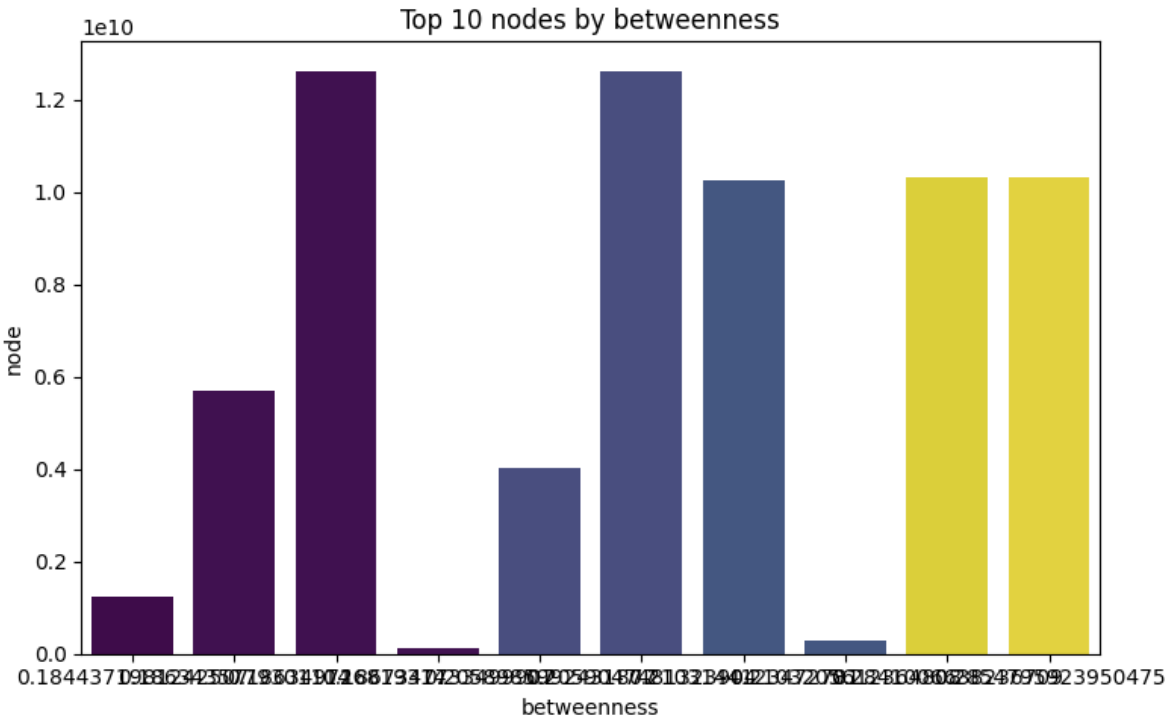


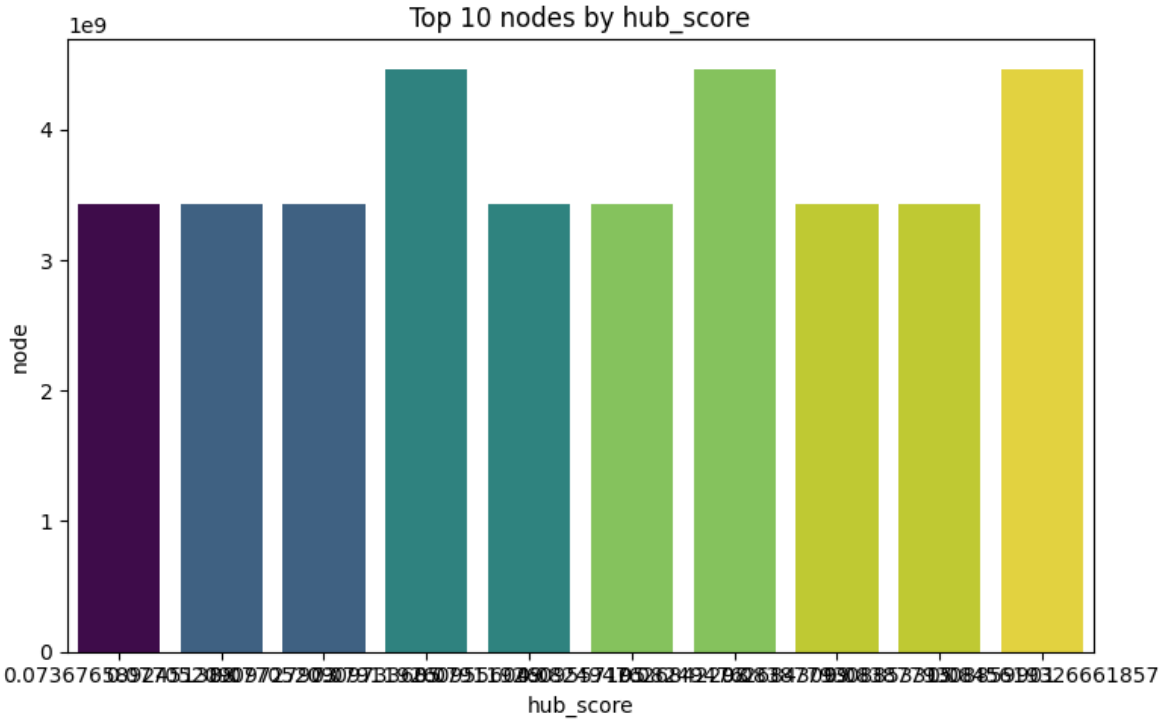
Centrality Visualizations

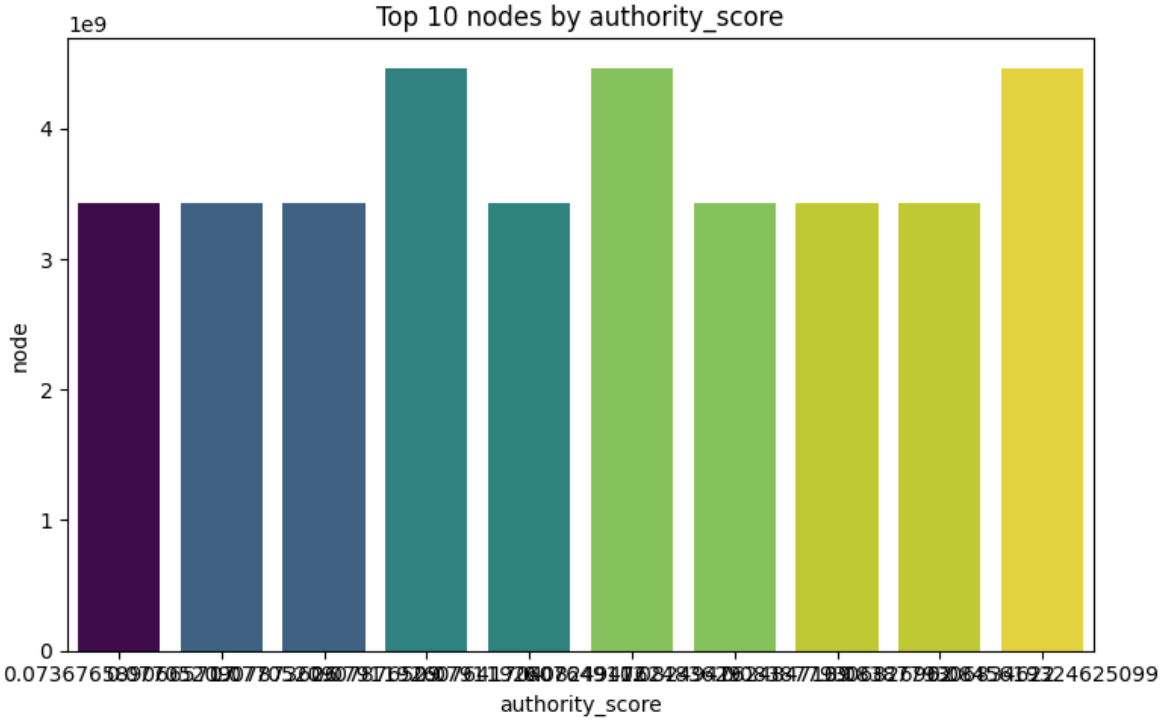

Top 10 nodes by in_degree
1e9
8
7
6
5
4
3
2
1
0
node
3.07440679320925e-05
3.689288151851101e-05
in_degree

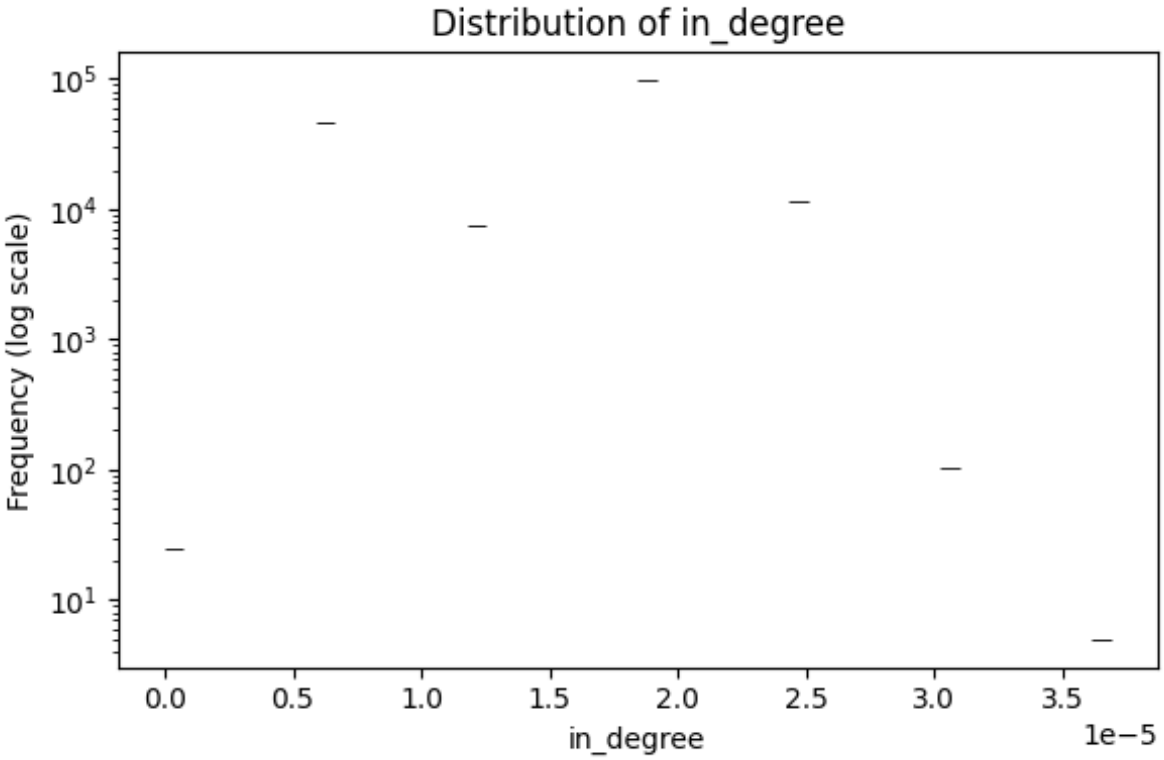

Top 10 nodes by out_degree
1e9
8
6
4
2
0
node
3.07440679320925e-05
3.689288151851101e-05
out_degree

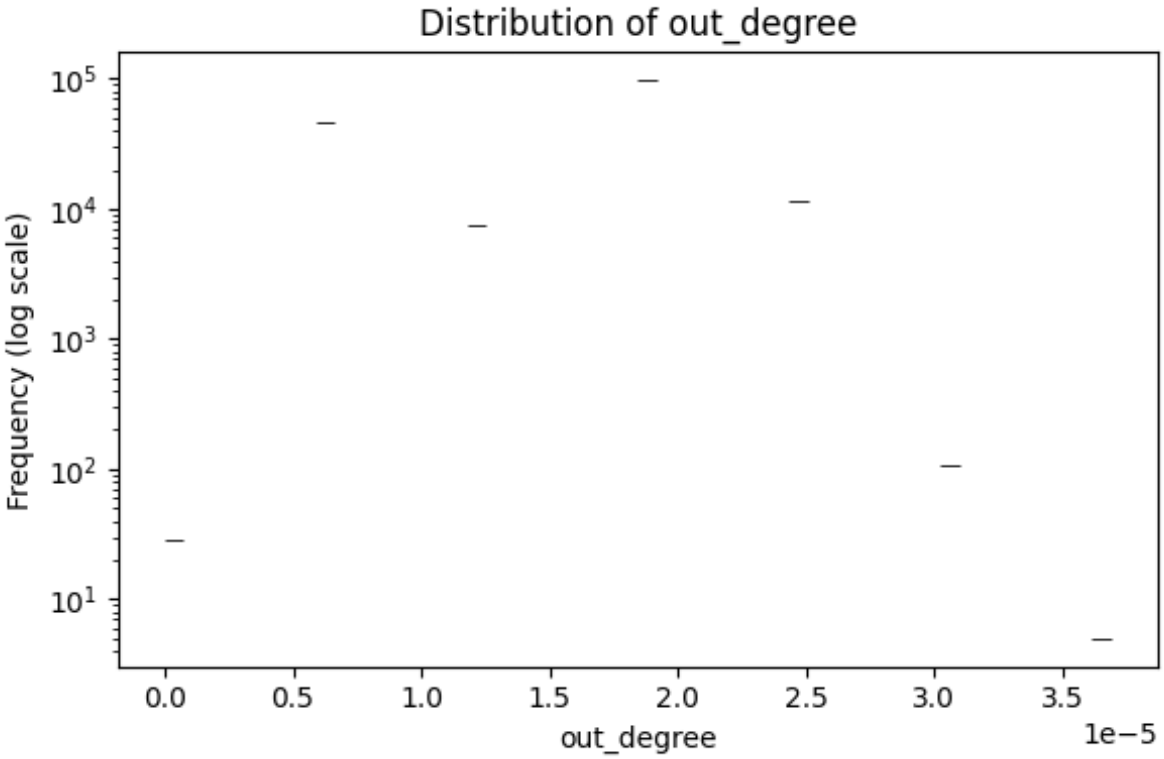

Top 10 nodes by pagerank
1e10
1.2
1.0
0.8
0.6
0.4
0.2
0.0
node
pagerank

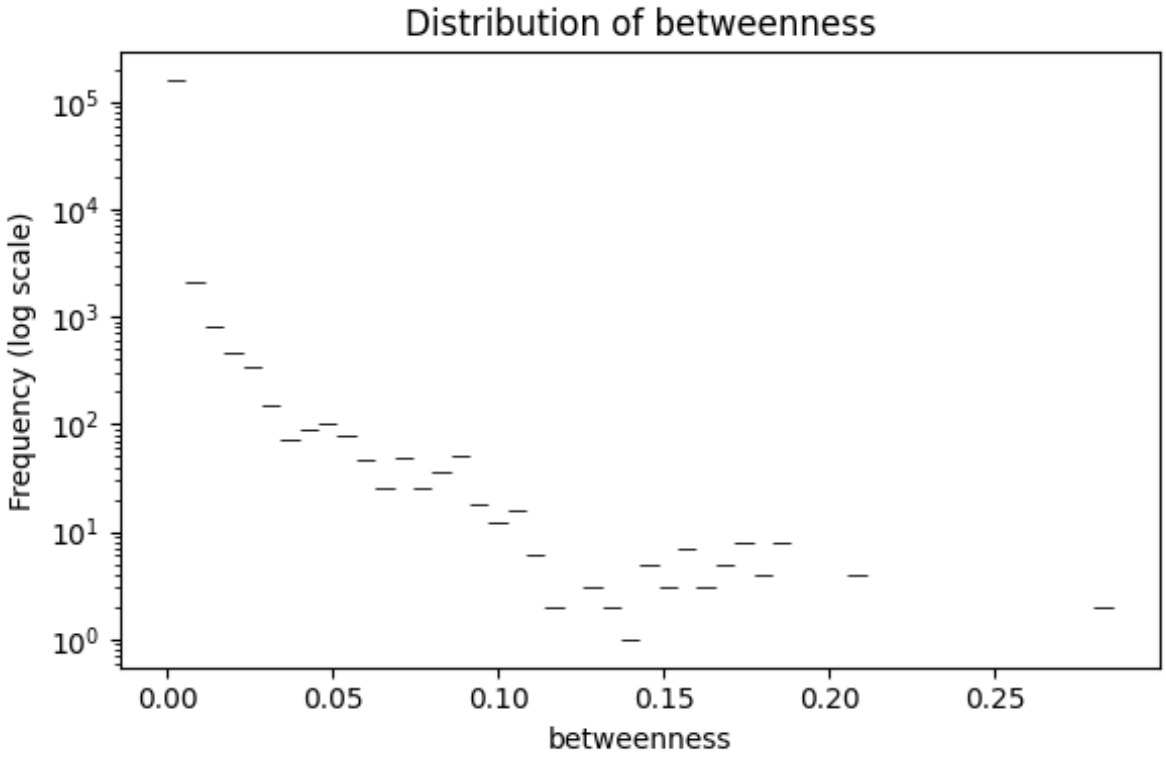

Top 10 nodes by betweenness
1e10
1.2
1.0
0.8
0.6
0.4
0.2
0.0
node
betweenness

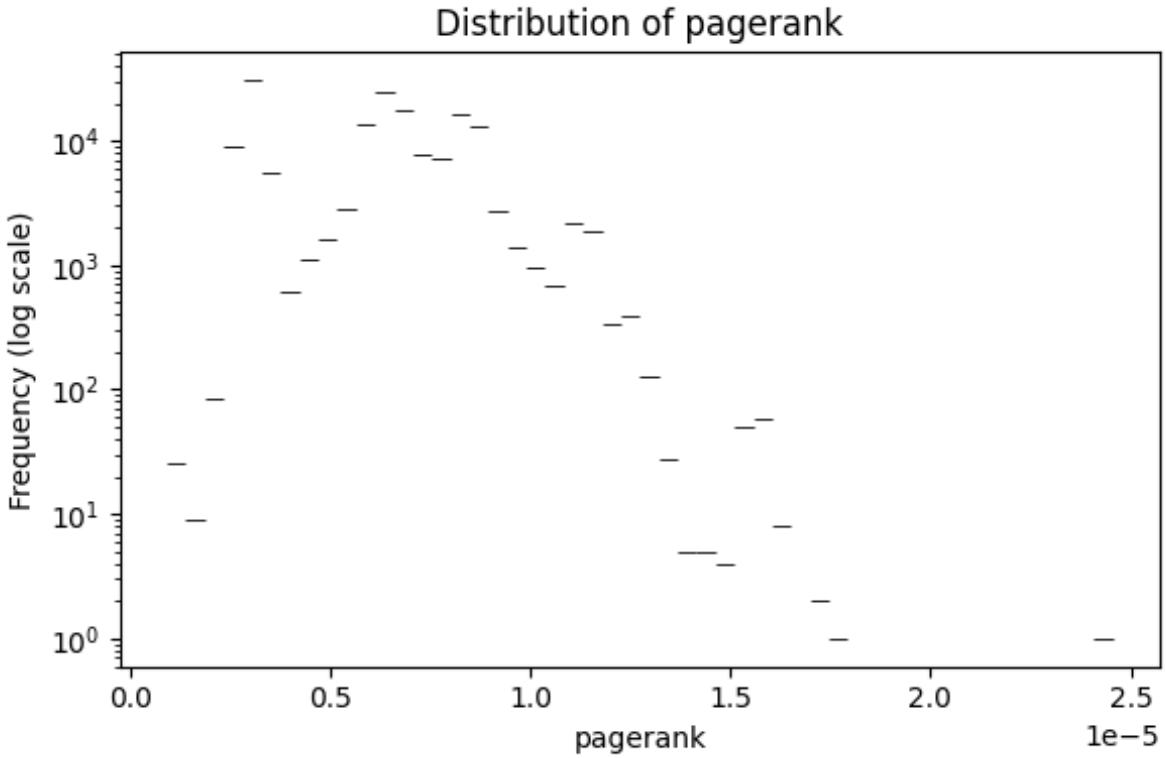

Top 10 nodes by hub_score
1e9
node
hub_score


Top 10 nodes by authority_score
1e9
node
authority_score


Distribution of in_degree
Frequency (log scale)
10^5
10^4
10^3
10^2
10^1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
in_degree
1e−5


Distribution of out_degree
Frequency (log scale)
10^5
10^4
10^3
10^2
10^1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
out_degree
1e−5





4. Community Detection

Communities were detected using the Louvain method applied to the undirected version of the graph.

Each community represents a cluster of papers that frequently cite each other.

CSV outputs:

- output/communities.csv
- output/community_analysis.csv

4. Community Detection

Communities were detected using the Louvain method applied to the undirected version of the graph.
Each community represents a cluster of papers that frequently cite each other.
CSV outputs:
- output/communities.csv
- output/community_analysis.csv

4. Community Detection

The Louvain algorithm was applied to detect communities within the citation network. Each community represents a group of papers that are more closely connected to one another based on citation relationships. The following visualizations and summaries show the main clusters and their interconnections.

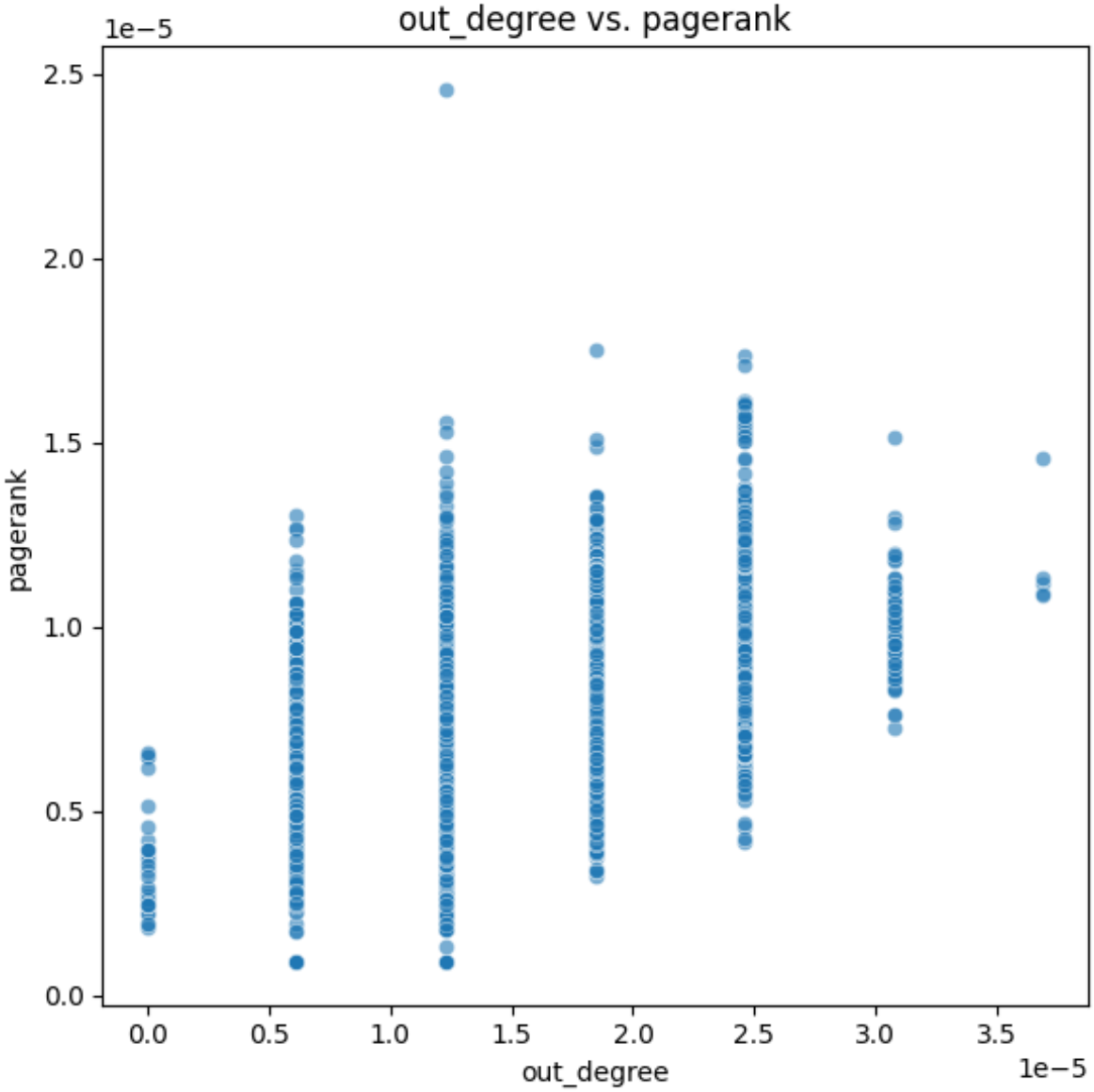
5. N

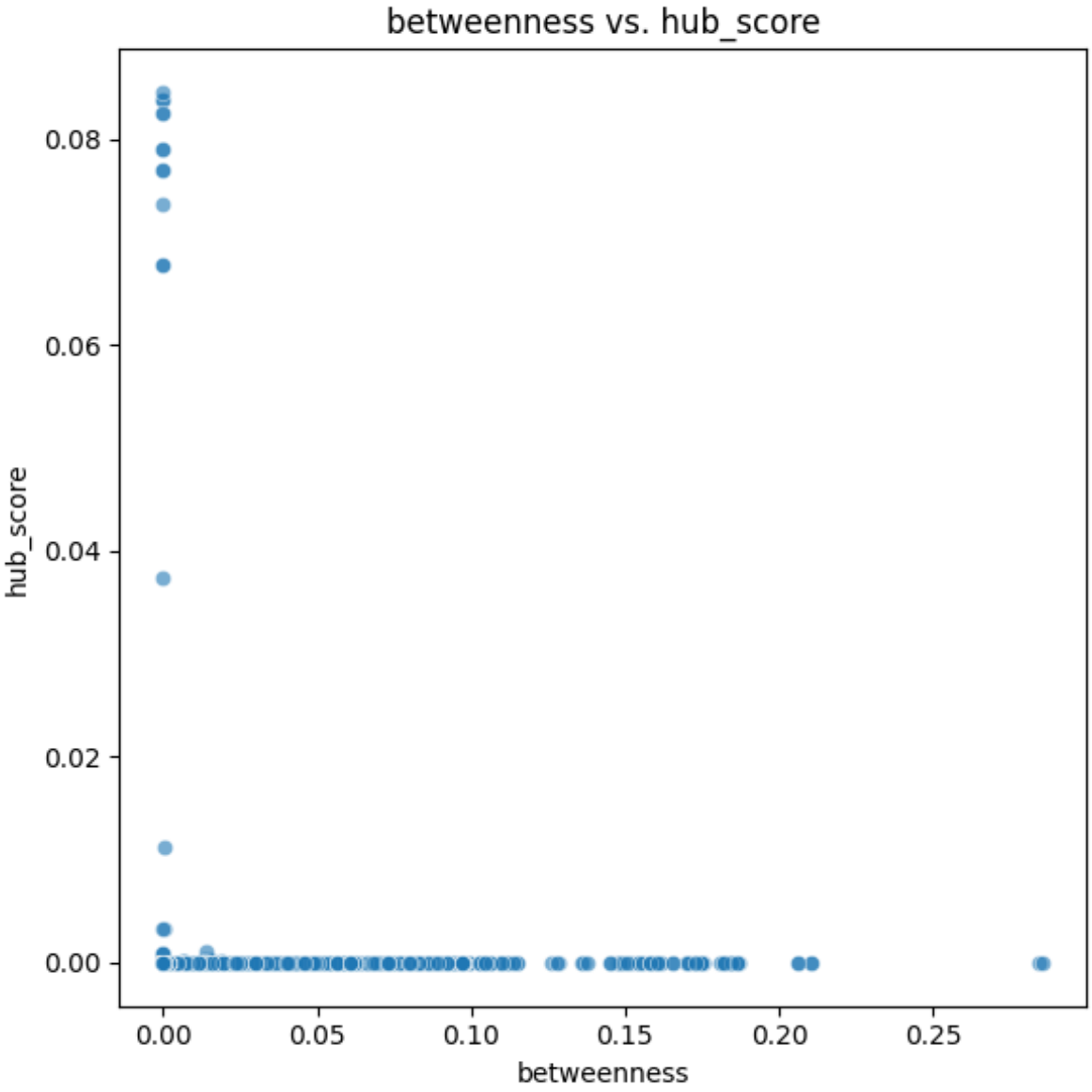
The largest weakly connected component of the graph was plotted using a spring layout. Nodes with higher centrality appear more central, and communities are visually identifiable.

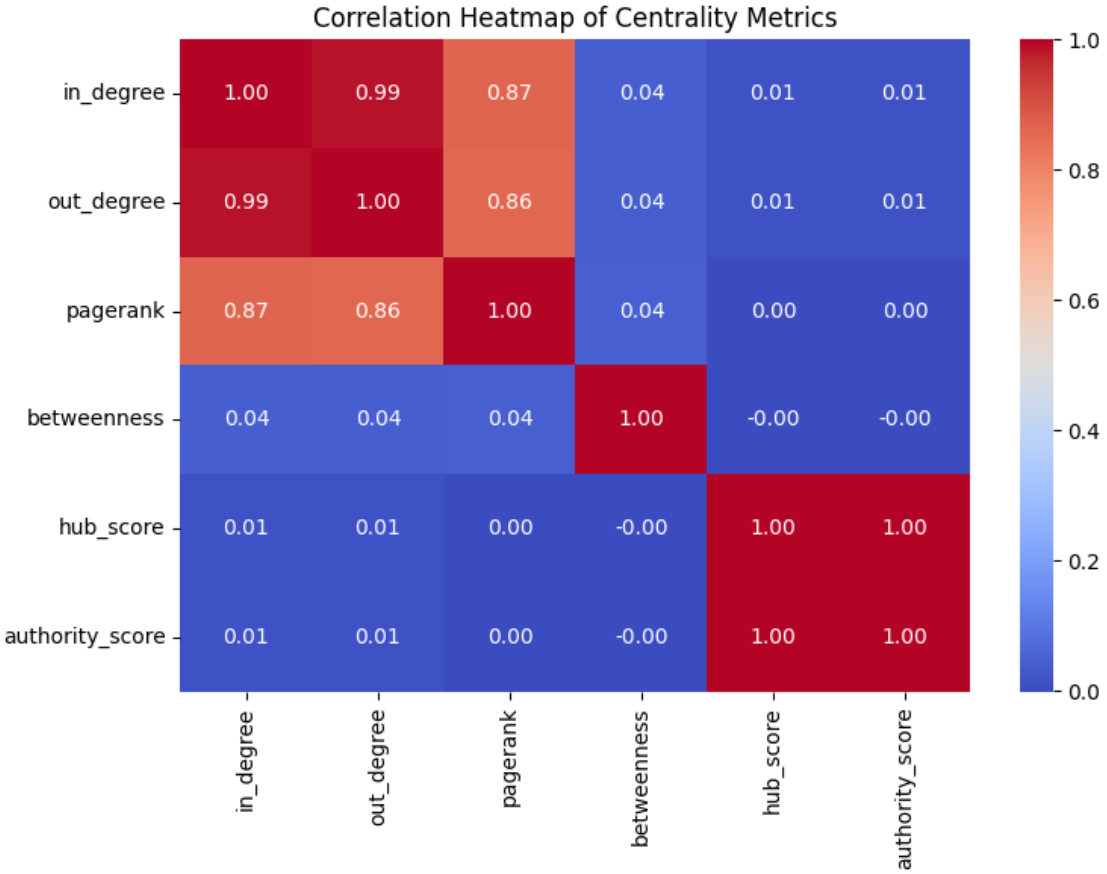
6. C

Scatterplots and a heatmap illustrate relationships between key centrality metrics. These visualizations help identify patterns and overlaps between different network measures.


pagerank


out_degree vs. pagerank
1e−5
2.5
2.0
1.5
1.0
0.5
0.0
pagerank
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
out_degree
1e−5


betweenness vs. hub_score
hub_score
0.08
0.06
0.04
0.02
0.00
0.00
0.05
0.10
0.15
0.20
0.25
betweenness



7. Summary & Insights

- Highly cited papers are not always hubs (high out-degree).
- Betweenness centrality highlights papers that bridge communities.
- Louvain communities reveal natural clusters of related research topics.
- PageRank and authority scores often overlap but capture slightly different notions of influence.

8. R

- NetworkX Documentation: https://networkx.org/documentation/stable/
- Louvain Method: Blondel, V.D. et al. (2008). Fast unfolding of communities in large networks.