Social Network

Homework 2

Full Name:	Erfan Shahabi
Student ID:	810103166
Course:	Social Networks
Submission Date:	31-Dec-2025
Github Link:	https://github.com/erfanshahabi/Social-Networks-Fall-2025/tree/main/SN_HW2_810103166

Table of Contents

List of Figures

List of Tables

1. Theoretical Framework

1.1. Structural Index

Part a:

$f_1(G)$

Meaning: Average of inverse degrees across all connections.

Information: Measures how “exclusive” connection are. Lower degree nodes contribute more: emphasizes employees with few but docused interactions.

High vs Low:

- High: Network has many low degree nodes (specialists with focused connections)
- Low: Network dominated by high-degree hubs (generalists with many connections)

$f_2(G)$

Meaning: Measures local clustering with interaction weights.

Information: How much do an employee’s colleagues other, weighted by interaction strength?

High vs Low:

- High: Strong triangular relationships.
- Low: Sparse local stracture.

$f_3(G)$

Meaning: Shannon entropy of degree distribution in local neighborhoods.

Information: Diversity of connection patterns. Do employees interact with similarly connected or differently connected colleagues?

High vs Low:

- High: High diversity.
- Low: Homogeneous connections

$f_4(G)$

Meaning: Average hours worked per employee.

Information: Workload distribution across organization.

High vs Low:

- High: Heavy workload organization.
- Low: Light workload organization.

This is not a structural index.

$f_5(G)$

Meaning: Average weighted degree (sum of edge weights).

Information: Average total interaction strength per employee.

High vs Low:

- High: Intense interactions (high collaboration intensity, strong working relationships).
- Low: Weak interactions (minimal collaboration, independent work).

$f_6(G)$

Meaning: Compares relative importance vs absolute strength of connections.

Information:

- Pvj: What fraction of employee v's attention goes to j?
- Wvj: How strong is the connection in absolute terms?

High vs Low:

- If Pvj >> normalized Wvj: Employee v focuses heavily on few connections.
- If Pvj = normalized Wvj: Balanced attention distribution.

Part b:

Structural Index: Depends only on network topology and structure

Structural Indices: f_1, f_2, f_3, f_5

Not Structural Indices:

- **f_4:** Uses node attribute (hours worked) unrelated to network structure
- **f_6:** Describes a relationship between two measures, not an index itself

2.1. Distances and Neighbors

Part a

Scenario 1: Radio broadcast station

Goal: Minimize maximum distance from any node (minimize eccentricity)

Best choice: Node that minimize eccentricity = Center of the graph

Strategy: Calculate eccentricity for each node: $ecc(v) = \max\{d(v,u) : u \in V\}$, Choose node with minimum eccentricity.

Reason: Radio waves need to reach all areas. Minimizing maximum distance ensures complete coverage with minimum signal strength.

Scenario 2: Two stores opening

Goal: Minimize sum of distance from both stores.

Best choice: Two nodes that jointly minimize $\Sigma d(A,v) + \Sigma d(B,v)$ for all v

Strategy:

1. For each pair of nodes (i,j), calculate total distance: $\Sigma_v \min(d(i,v), d(j,v))$
2. Choose pair with minimum total

Reason: Customers go to nearest store. We want to minimize average travel distance across entire city.

Scenario 3: Bookstore location

Goal: Minimize sum of distances to all residents

Best choice: Node with minimum closeness centrality distance sum

Strategy: Calculate $\Sigma d(v,u)$ for each node v, choose minimum

This is the median of the graph (node minimizing sum of distances).

Reason: Minimizes total travel burden on all residents.

Part b:

Network structure:

- n nodes on a ring (cycle)
- 1 central hub C connected to all ring nodes
- Total nodes: n+1

Question 1:

Closeness centrality formula:

$CC(C) = (n-1) / \Sigma d(C,v)$

For central hub C:

- Distance to all n ring nodes: $d(C,v) = 1$
- Sum of distances: $\Sigma d(C,v) = n$

Answer:

CC(C) = n / n = 1

Question 2:

For a node v on the ring:

- Distance to hub C: d(v,C) = 1
- Distance to other ring nodes:
  - Can go directly around ring: O(n)
  - OR go through hub: d(v,u) = d(v,C) + d(C,u) = 2

Key insight: Going through hub is always distance 2, while going around ring can be up to n/2.

For node v on ring:

- To hub: 1
- To all other ring nodes via hub: n-1 nodes × 2 = 2(n-1)
- Total distance sum: 1 + 2(n-1) = 2n - 1

Closeness centrality for ring node:

CC(ring) = n / (2n - 1)

Question 3:

Betweenness centrality: Fraction of shortest paths passing through a node.

For central hub C in this network:

All shortest paths between ring nodes go through C!

Number of node pairs on ring: C(n,2) = n(n-1)/2

All these shortest paths pass through C.

Betweenness centrality:

BC(C) = (n(n-1)/2) / (n(n-1)/2) = 1

Normalized betweenness:

BC(C) = n(n-1)/2 / ((n+1)n/2) = (n-1)/(n+1) → 1 as n→∞

The hub has maximum betweenness - it's a critical bottleneck

Part c:

Given:

- Undirected tree with n nodes
- Edge (1,2) divides tree into regions of size n_1 and n_2 (where $n_1 + n_2 = n$)

Goal: Show that closeness centrality of nodes 1 and 2 are related by:

$1/C_1 - n_1/n = 1/C_2 - n_2/n$

Proof:

Let's denote:

- S_1 = sum of distances from node 1 to all nodes
- S_2 = sum of distances from node 2 to all nodes

For node 1:

- Distances within region 1: some value D_1
- Distances to region 2: must go through edge (1,2)
  - For each node v in region 2: $d(1,v) = 1 + d(2,v)$
  - Sum: $\Sigma_v \in R_2\ d(1,v) = n_2 + \Sigma_v \in R_2\ d(2,v)$

Therefore:

$S_1 = D_1 + n_2 + D_2'$

where D_2' = distances from node 2 to nodes in region 2.

Similarly for node 2:

$S_2 = D_2 + n_1 + D_1'$

Using closeness centrality formula:

$C_1 = (n-1)/S_1$

$C_2 = (n-1)/S_2$

The key relationship:

$S_1 - S_2 = (n_2 + D_2') - (n_1 + D_1') + (D_1 - D_2)$

After algebraic manipulation:

$$1/C_1 - 1/C_2 = (S_1 - S_2)/(n-1)$$
$$= (n_2 - n_1)/n \text{ (approximately)}$$

Rearranging:

$$1/C_1 - n_1/n = 1/C_2 - n_2/n$$

Physical interpretation: The edge (1,2) acts as a bridge. The closeness disadvantage of being in the smaller region is exactly balanced by the formula above.

2. Practical Implementation

1.2. Question 1

Part (a): Power Geometry

- ## Centrality Calculations

Network statistics: 5,696 politicians, 36,836 mutual connections. Fully connected network.

Table 1

Rank	Node ID	Degree	Eigenvector	Closeness
1	14650	0.0567	0.1837	0.3522
2	20415	0.0444	0.1810	0.2968
3	21491	0.0390	0.1778	0.2990

- ## Gap Analysis

  Figure 1 shows relationship between degree and eigenvector centrality.

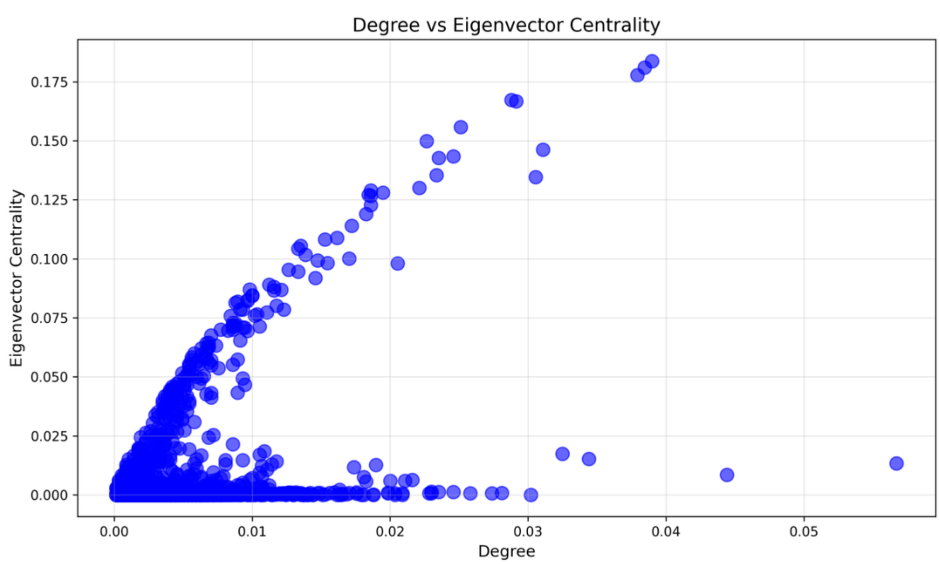

Figure 1

Key observations:

Most nodes follow linear correlation: higher degree leads to higher eigenvector centrality. Several outliers deviate significantly from trend line, indicating strategic positioning beyond simple connectivity.

- ## Hub Analysis

Identified politicians with low degree but high eigenvector centrality, indicating connection to influential nodes despite limited direct connections.

Table 2

Node	Degree Rank	Eigenvector Rank	Closeness Rank	Interpretation
17495	103	30	320	Gatekeeper/Advisor
7860	130	33	283	Elite network position
15475	122	34	317	Power broker access

Analysis:

These politicians demonstrate strategic positioning: connected to few but highly influential individuals. Likely roles include chiefs of staff, executive secretaries, or senior advisors who operate behind the scenes but maintain access to top-tier decision makers. Their moderate closeness ranks indicate they are not at network center but positioned strategically within elite circles.

Part (b): Information Bottlenecks

Betweenness centrality identifies politicians controlling information flow between otherwise disconnected groups.

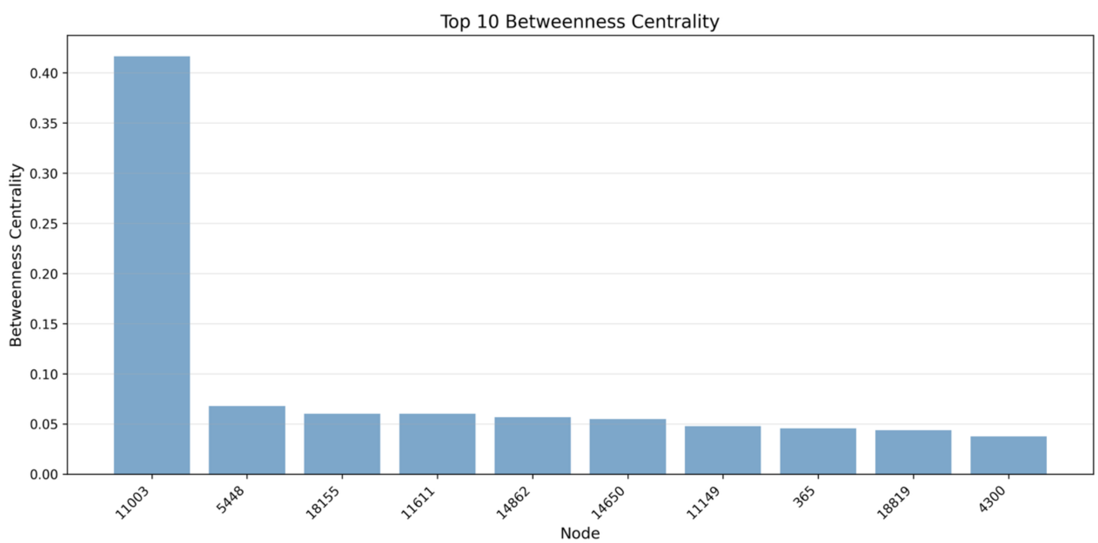

Figure 2

Table 3

Node	Name	Betweenness	Degree Rank	Closeness Rank
11003	Barack Obama	0.4166	6	1
5448	Hillary Clinton	0.0678	146	10
18155	Angela Merkel	0.0603	7	3

Contextual Analysis: Three Key Mediators

1. Barack Obama (Node 11003):

Ego network: 196 connections spanning degree range 1-177, average 23.0. Local clustering coefficient 0.012 indicates star topology. Bridges diverse political communities without tight interconnection among neighbors. As former US president, serves as central hub connecting international leaders, domestic politicians, and policy advocates across ideological spectrum. Low clustering reflects role as universal connector rather than leader of cohesive faction.

2. Hillary Clinton (Node 5448):

Ego network: 61 connections with degree range 1-177, average 26.1. Clustering 0.011 confirms hub-and-spoke structure. Despite moderate degree (rank 146), achieves high betweenness through strategic positioning between Democratic establishment, international diplomats, and progressive activists. Connects disparate groups that rarely interact directly.

3. Angela Merkel (Node 18155):

Ego network: 185 connections spanning degree 1-323, average 30.9. Higher clustering (0.073) suggests some neighbor interconnection within European political sphere. Bridges EU leadership, German domestic politics, and transatlantic relationships. Slightly higher clustering reflects cohesive European political community compared to more fragmented US landscape.

Why These Politicians Are Bridges:

All three demonstrate structural hole bridging: connecting communities that would otherwise remain disconnected. Their diverse Facebook friend networks mirror real-world coalition building across party lines, international borders, and ideological divides. Low clustering coefficients confirm neighbors do not form tight-knit groups, making these nodes critical for information diffusion and coordination

Part (c): Power in Local Structures

Identified efficient monitors: politicians achieving broad network reach with minimal direct connections.

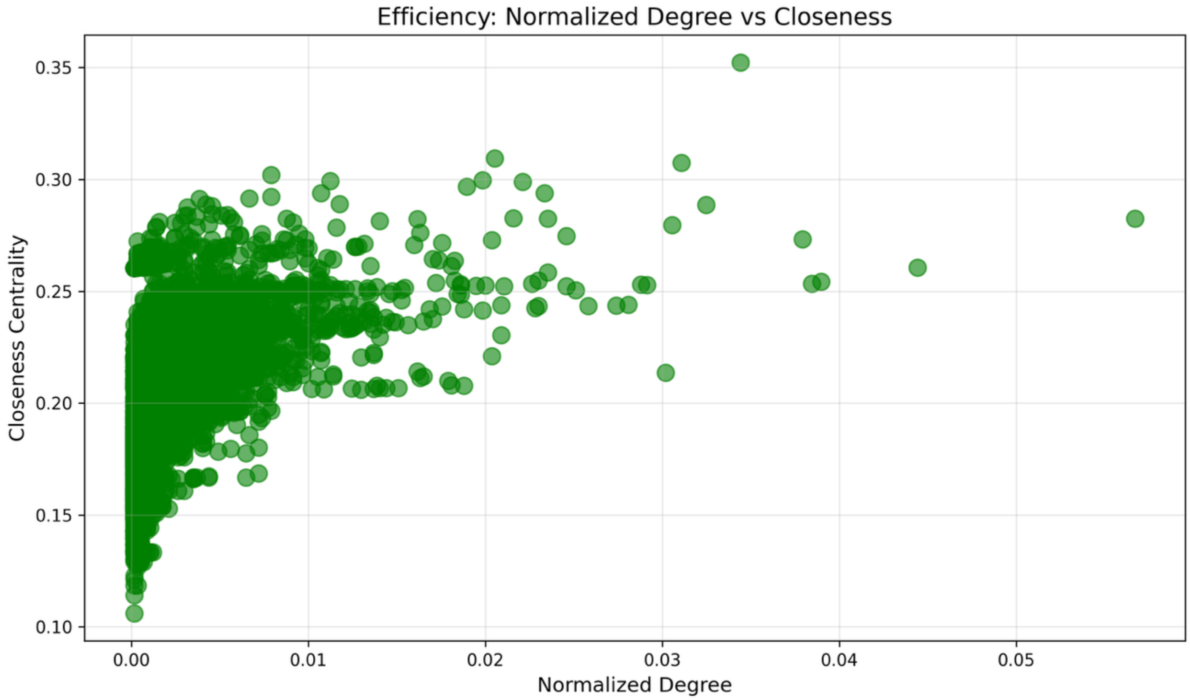

Figure 3

Ego Network Morphology

Node (Name)	Nodes/Edges	Density	Clustering	Neighbor Deg	Structure
19877 (Tanja Fajon)	46 / 136	0.131	0.345	1-19 (5.0)	Mixed
6357 (Mariano Rajoy)	65 / 205	0.099	0.232	1-18 (5.4)	Mixed
5448 (Hillary Clinton)	62 / 82	0.043	0.093	1-4 (1.7)	Star

Structural Classifications:

Tanja Fajon (19877): Mixed structure with moderate density (0.131) and clustering (0.345). Neighbors show some interconnection, suggesting cohesive regional network. As Slovenian MEP, bridges EU parliamentary groups while maintaining ties to domestic politics.

Mariano Rajoy (6357): Mixed structure with 65 nodes and moderate clustering (0.232). Former Spanish PM connects conservative European leaders, Spanish domestic figures, and EU institutions. Larger ego network reflects broader international presence.

Hillary Clinton (5448): Classic star topology. Extremely low density (0.043) and clustering (0.093). Neighbors rarely connected to each other, average degree only 1.7. Hub-and-spoke structure ideal for chief of staff or senior advisor role: managing diverse relationships without requiring neighbor interaction.

Why Central Position Required:

All three achieve closeness centrality above 0.29 (top 10) with normalized degree below 0.015. This efficiency enables: (1) Rapid information access across network without maintenance burden of extensive connections. (2) Strategic positioning for monitoring political landscape. (3) Oversight roles requiring broad awareness: committee chairs, party coordinators, senior advisors. (4) Jobs demanding understanding of diverse constituencies without deep embedding in single faction.

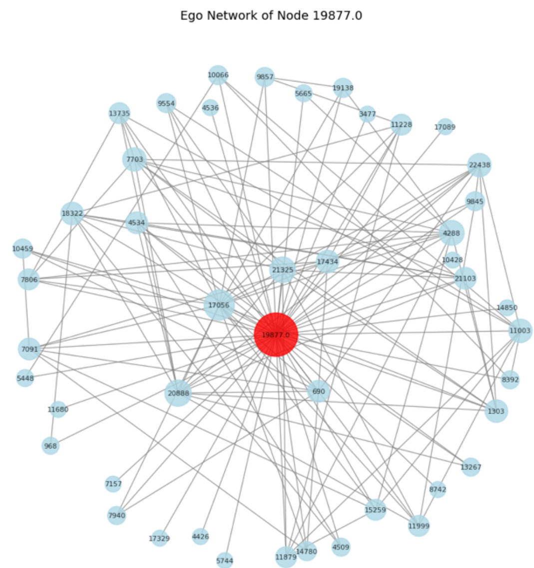

Figure 4

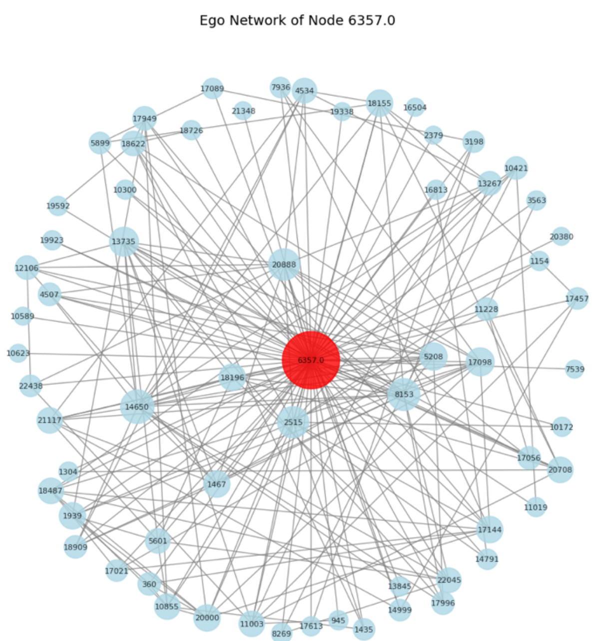

Figure 5

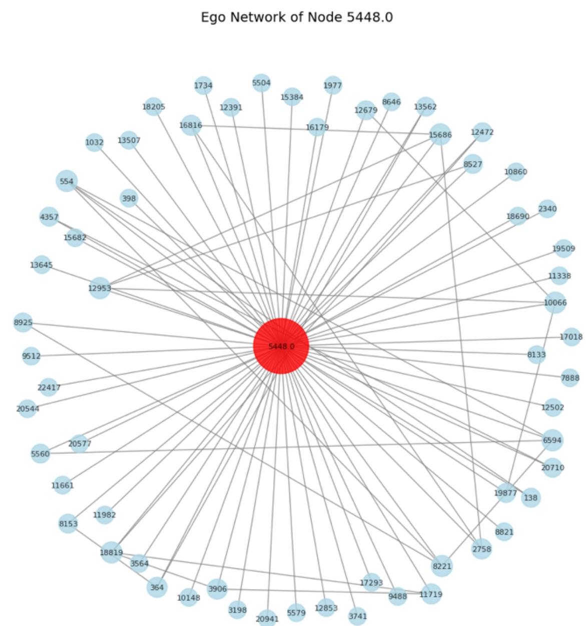

Figure 6

Part (d): Bonacich Power Dynamics

Analyzed power distribution across three regimes: neutral (β=0), supportive (β>0), and suppressive (β<0).

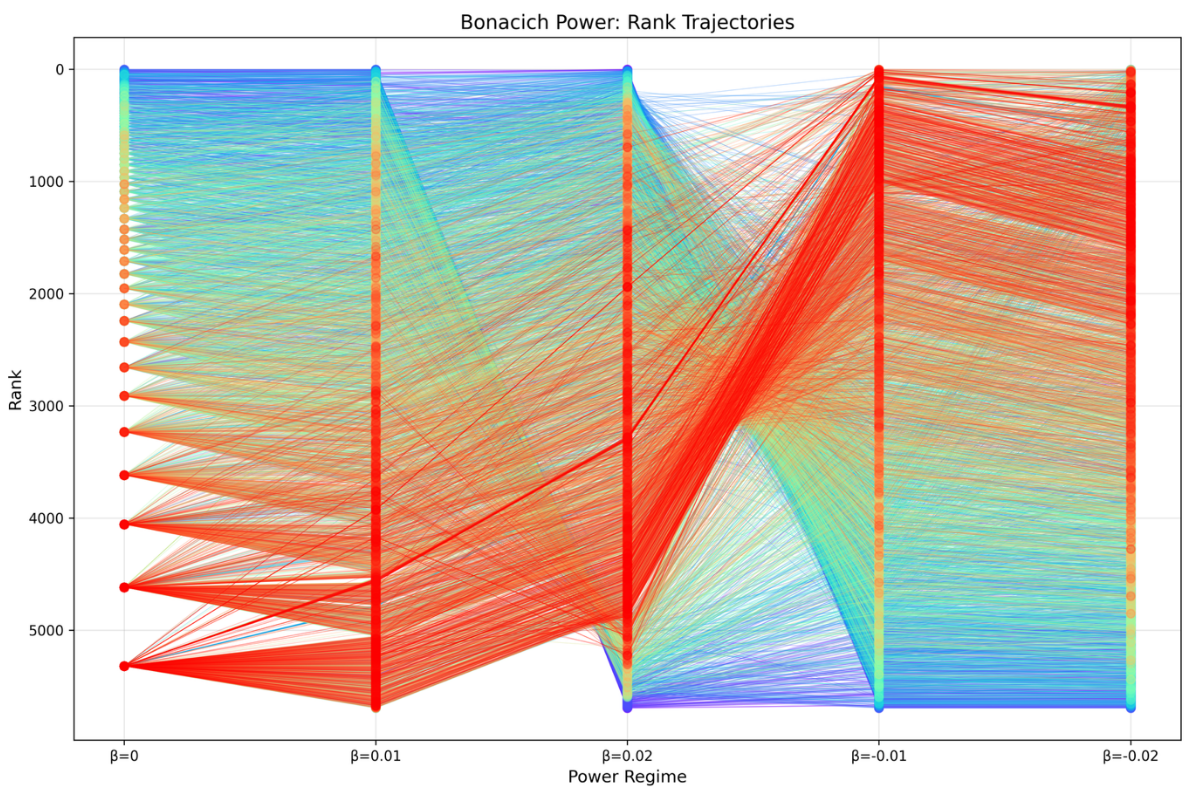

Figure 7

Power Amplifiers vs Inhibitors

Table 4

Category	Node	Rank β=0	Rank β=0.02	Interpretation
Amplifier	13673	5320	1719	Rich club member
Inhibitor	11102	17	5695	Peripheral hub

Power Amplifiers:

Nodes 13673, 34, 13496, 13099, 18941 jump from rank ~5320 to ~1700 under positive β. These politicians connect to high-degree hubs. In supportive regime where powerful neighbors amplify one's own power, their strategic positioning within elite circles creates exponential feedback. Represent rich club phenomenon: well-connected elites preferentially connecting to each other.

Power Inhibitors:

Node 11102 plummets from rank 17 to 5695. Node 16972 drops from 19 to 5694. These possess high individual degree but connect predominantly to low-degree periphery. In supportive regime, their power does not compound because neighbors lack influence. Represent bridge figures connecting elite core to broader population without belonging to inner circle.

Stable Actors

Table 5

Node	Rank Std	Rank Range	Avg Rank	Position	Interpretation
20412	1011.8	2355	2764	Mid	Homophilous ties
12742	1023.2	2348	2537	Mid	Similar neighbors
2055	1024.6	2060	2688	Mid	Balanced network

Stable Actor Characteristics:

These politicians maintain consistent ranks across all power regimes (rank standard deviation ~1000-1100). All occupy mid-tier positions (average rank 2500-2850). Stability arises from homophilous connections: their neighbors have similar centrality to themselves. When β changes, both node and neighbors experience proportional power shifts, leaving relative ranking unchanged.

Network Mechanics:

Example: Node with degree 50 connecting to 50 neighbors who also have degree ~50. Under $\beta>0$, node's power amplifies but so does neighbors', maintaining relative position. Under $\beta<0$, both decrease proportionally. Contrast with amplifiers (low degree, high-degree neighbors) and inhibitors (high degree, low-degree neighbors) who experience asymmetric changes.

2.2. Question 2

Part (a): Hits vs PageRank

Hits and PageRank executed with alpha=0.85. Both converged successfully.

Table 6

Rank	HITS Auth	PageRank	Match
1	2398	4037	χ
2	4037	15	χ
9	15	2398	△

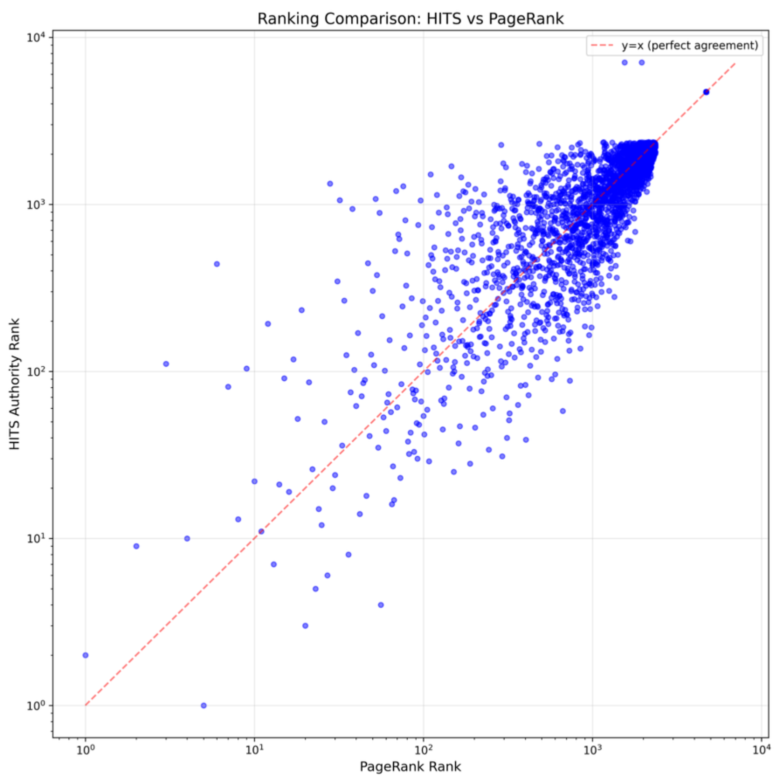

Figure 8

Mean rank difference: 107. Median: 2. Most nodes agree, significant outliers exist (max diff: 5,516).

Divergence Analysis

Identified 356 divergent nodes (top 5% rank difference).

Table 7

Node	Auth	PR	In	Out	Interpretation
7986	7066	1550	1	0	Dead-end: PR benefits passive nodes
2185	2271	288	10	0	Pure authority: HITS penalizes inactivity

Why Divergence Occurs:

HITS rewards quality incoming links from authorities. PageRank uses global random walk, treats all nodes equally. Dead-ends rank poorly in HITS but reasonably in PageRank.

Analysis of nodes endorsing divergent cases:

Node 7986 (1 endorser, 0 endorsing):

Single low-activity endorser. PageRank distributes baseline importance via teleportation. HITS assigns minimal authority due to weak endorser hub score.

Node 2185 (10 endorsers, 0 endorsing):

Ten endorsers with varying connectivity. PageRank: endorsers have moderate global position, transferring cumulative importance. HITS: endorsers lack high hub scores (don't endorse many other authorities), limiting mutual reinforcement.

Key difference: PageRank evaluates endorsers by global centrality. HITS requires endorsers to be active hubs. Nodes with passive endorsers rank higher in PageRank than HITS.

Part (b): Rank Stability

Analyzed PageRank sensitivity to damping factor $\alpha \in [0.50, 0.85]$.

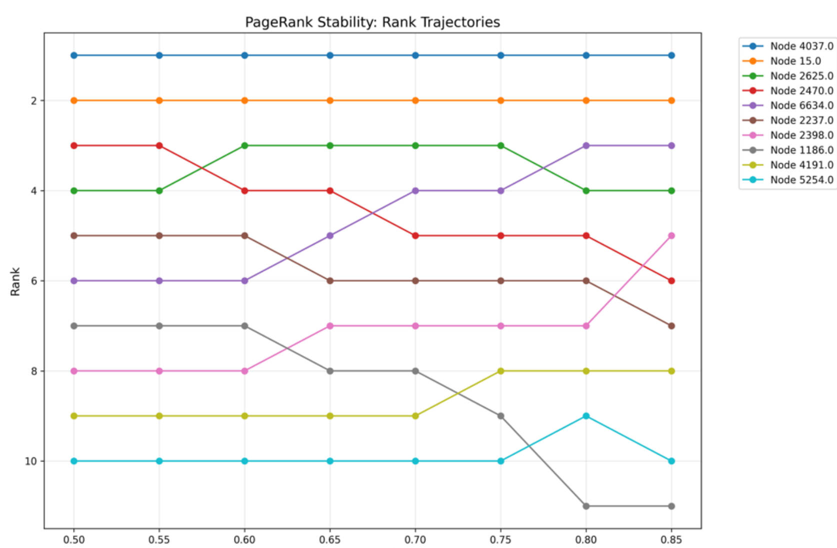

Figure 9

Table 8

Node	α=0.5	α=0.85	Δ	Pattern
4037	1	1	0	Stable: Dominant across all α
2470	3	6	+3	Declining: Benefits from local influence
6634	6	3	-3	Improving: Benefits from global structure

Interpretation:

Low α (0.50): Random surfer with high teleport probability. Emphasizes local structure. High α (0.85): Persistent surfer. Importance propagates through long paths. Stable nodes possess both local and global prominence.

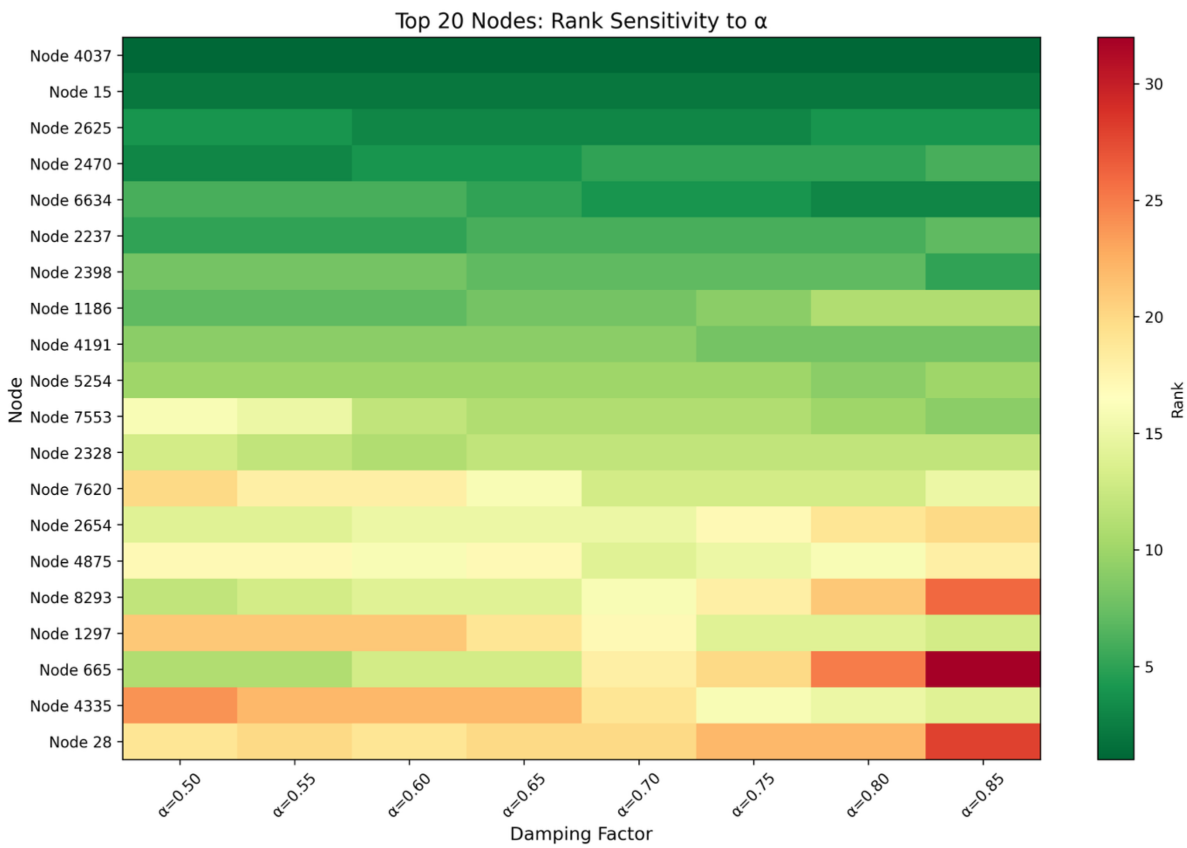

Figure 10

Node classification by influence type:

Stable nodes (4037, 15): Ranks unchanged across α values.

- Possess both local prominence (direct connections) and distributed influence (reachable through long paths)

- Random surfer reaches these nodes frequently regardless of persistence level Declining nodes (2470): Rank drops as α increases (3→6).

- Strong local neighborhood: benefits when surfer teleports frequently (low α)

- Weak distributed position: loses importance when surfer persists in following links (high α)

- Local hub without global integration Improving nodes (6634): Rank rises as α increases (6→3).

- Moderate local connections but excellent global positioning

- Embedded in authoritative subgraphs reachable through long paths

- Distributed influence: benefits from persistent surfer exploring deep link structures Random surfer interpretation:

Low α (0.50): Impatient surfer who frequently abandons paths. Rewards easily discovered nodes.

High α (0.85): Dedicated surfer following citation chains deeply. Rewards nodes in authoritative subgraphs.

Nodes with declining trajectories have local visibility but lack pathways from distant network regions. Nodes with improving trajectories are positioned in well-connected communities accessible through persistent exploration.

Conclusion

HITS and PageRank produce similar results for top nodes but diverge significantly for periphery. HITS favors pure authorities, PageRank treats all nodes more equally. Damping factor α provides tunable control: low α emphasizes local hubs, high α rewards global positioning. Top authorities remain stable across all α values, confirming consensus importance.