Social Networks: Assignment #2

Due on Saturday, December 26, 2025

Dr. Masoud Asadpour 15:00am

Ali Dashtbozorg 810104302

Contents

List of Figures

Question 1

Structural Index

(a) Explain what each of these functions represents (what information they provide about the network). In doing so, describe what a high value versus a low value of each function indicates.

1. **Global Communication Efficiency (f_1)**

$$f_1(G) = \frac{1}{n(n-1)} \sum_{u \neq v} \frac{1}{d(u,v)} \tag{1}$$

This function measures the average speed and ease with which information spreads across the organization. A **high value** indicates a "flat" structure where employees can reach anyone quickly with few intermediaries, facilitating rapid decision-making. A **low value** suggests a siloed or highly hierarchical organization where information travels slowly through many bottlenecks.

2. **Collaborative Redundancy and Local Density (f_2)**

$$f_2(G) = \frac{1}{n} \sum_{i \in V} \sum_{j \in N(i)} \left(p_{ij} + \sum_{\substack{q \in N(i) \\ q \neq j}} p_{iq} p_{qj} \right)^2 \tag{2}$$

This metric represents the intensity of local team clusters and how much an employee's attention is reinforced by a tight-knit circle. A **high value** indicates strong, redundant team structures where social capital is high and the network is resilient to individual absences. A **low value** indicates sparse local connections, suggesting employees may be working in isolation without a strong peer support system.

3. **Structural Heterogeneity and Degree Entropy (f_3)**

$$f_3(G) = \frac{1}{n} \sum_{v \in V} \frac{d(v)}{\sum_{u \in V} d(u)} \log \left(\frac{d(v)}{\sum_{u \in V} d(u)} \right) \tag{3}$$

This formula measures the diversity of influence and connectivity within the workforce. A **high value** (in magnitude) indicates a diverse range of roles, featuring both specialized individual contributors and highly connected "hubs" or bridge-builders. A **low value** suggests a very uniform network where every employee has a similar number of contacts, which can result in a lack of clear organizational "connectors."

4. **Average Digital Presence (f_4)**

$$f_4(G) = \frac{1}{n} \sum_{v \in V} H(v) \tag{4}$$

This represents the average amount of time employees spend working online. A **high value** indicates a highly digitally-active or remote-first workforce, though if extreme, it may point toward meeting fatigue or burnout. A **low value** suggests an organization that operates primarily offline (such as manual labor or face-to-face services) or is currently in a low-activity phase.

5. **Total Organizational Connectivity (f_5)**

$$f_5(G) = \frac{1}{n} \sum_{v \in V} \sum_{u \in N(v)} w_{vu} \tag{5}$$

This sums and averages the raw weights of all interactions to measure the total volume of work flowing through the network. A **high value** indicates high-intensity collaboration and "heavy" workloads across the organization's links. A **low value** indicates "thin" relationships where, despite being connected, the actual strength and frequency of interaction between colleagues is minimal.

(b) Determine which of these functions is a Structural Index (SI) and which ones are not. Provide a brief justification for your answer

1. **Global Communication Efficiency (f_1)**: Yes, SI. Depends only on shortest path distances between nodes; measures how efficiently information can flow through the network.
2. **Collaborative Redundancy / Local Density (f_2)**: Yes, SI. If the proportions p_{ij} are derived from intrinsic edge weights w_{ij} of the weighted graph, f_2 depends purely on the weighted network topology and reflects relative connectivity patterns dictated by the network structure.
3. **Structural Heterogeneity / Degree Entropy (f_3)**: Yes, SI. Uses node degrees only; measures diversity in connectivity and roles within the network.
4. **Average Digital Presence (f_4)**: No, not SI. Depends on dynamic activity of nodes ($H(v)$), which is an external attribute not determined by network topology.
5. **Total Organizational Connectivity (f_5)**: Yes, SI. If the edge weights w_{ij} are intrinsic to the weighted graph structure, f_5 depends purely on the weighted topology and reflects the total connectivity encoded by the network structure.

Question 2

Distances and Neighbors

(a) Consider the below network.

For each of the following scenarios, indicate which node would be the best choice, giving reasons:

1. The mayor wants to install a radio broadcast station so that, in a crisis, a single nationwide message can reach all areas. The goal is that every node's distance to the station (independently of other nodes) is as small as possible —in other words, the maximum distance from any node to the station should be minimized.

Question 2 continued on next page...

To determine the optimal location for the radio station, we identify the **Graph Center** by computing the **Eccentricity** of each node. The eccentricity of a node is the distance to the farthest node in the network; minimizing this ensures that the maximum travel time for a broadcast signal is as small as possible.
The eccentricity vector ϵ for nodes 1 through 15 is:

$$\epsilon = \begin{bmatrix} 4.0 \\ 3.0 \\ 4.0 \\ 5.0 \\ 4.0 \\ 4.0 \\ 4.0 \\ 5.0 \\ 4.0 \\ 4.0 \\ 4.0 \\ 5.0 \\ 3.0 \\ 4.0 \\ 5.0 \end{bmatrix}$$

The minimum eccentricity is $\min(\epsilon) = \mathbf{3.0}$, which occurs at **Node 2** and **Node 13**. Therefore, placing the radio station at either of these nodes ensures that every other node is reachable within at most 3 hops, providing the most efficient coverage across the network.

2. The mayor wants to choose a node for a bookstore so that the sum of distances from all residents to that node is minimized.

To determine the optimal node for placing the bookstore, we use **closeness centrality**, which measures how close a node is on average to all other nodes in the graph. By choosing the node with the highest closeness centrality, we ensure that the nodes can reach the bookstore as quickly as possible on average.
The formula for closeness centrality of a node v is:

$$C(v) = \frac{n-1}{\sum_{u \in V} d(v,u)}$$

where $d(v,u)$ is the shortest-path distance between nodes v and u, and n is the total number of nodes.
For our 15-node graph, the closeness centrality values are:

$$\begin{bmatrix} 0.4667 & 0.4667 & 0.4667 & 0.4 & 0.4516 \\ 0.4516 & 0.4667 & 0.4242 & 0.4516 & 0.4375 \\ 0.4118 & 0.4 & 0.5185 & 0.3684 & 0.2745 \end{bmatrix}$$

Node $i = 13$ (value 0.5185) has the highest closeness centrality, so we choose it as the location for the bookstore to ensure efficient distance to other nodes.

Question 2 continued on next page. . .

3. Two stores have decided to open new branches in the city. Each person (node) buys from the nearest store. If a person is at equal distance from both stores, their purchases are split equally between them. First, select the best node to open store A, then determine the best location for store B given that choice.

From the closeness centrality analysis, we know that node 13 has the highest value, so it is the best choice for opening store A. Given store A at node 13, the nodes that would maximize the number of customers for store B are those with the highest customer counts. In this case, nodes 1, 3, 7, and 8 each attract 7.0 customers, making them all equally good candidates for store B.
The number of customers each node would get if store A is at node 13 is shown below, with the best candidates for store B highlighted in bold:

$$\begin{bmatrix} 1 & \mathbf{7.0} \\ 2 & 6.0 \\ 3 & \mathbf{7.0} \\ 4 & 6.5 \\ 5 & 6.5 \\ 6 & 5.5 \\ 7 & \mathbf{7.0} \\ 8 & \mathbf{7.0} \\ 9 & 6.5 \\ 10 & 5.0 \\ 11 & 5.5 \\ 12 & 6.5 \\ 14 & 2.0 \\ 15 & 1.5 \end{bmatrix}$$

Therefore, the recommended locations are: store A at node 13, and store B at any of nodes 1, 3, 7, or 8.

(b) Consider the below network.

1. What is the closeness centrality for node C ?

The closeness centrality(both standard and normalized) for node C are as follows:

$$C_C(c) = \frac{1}{\sum\limits_{j=1}^{n} d_{cj}} = \frac{1}{n-1}$$

$$\tilde{C}_C(c) = \frac{n-1}{\sum\limits_{j=1}^{n} d_{cj}} = \frac{n-1}{n-1} = 1$$

2. Derive the closeness centrality value for the nodes on the ring as a function of n

The closeness centrality(both standard and normalized) for nodes other than node

C are as follows:

$$C_C(n_i) = \frac{1}{\sum_{j=1}^{n} d_{ij}} = \frac{1}{1+1+1+(n-4)(2)} = \frac{1}{2n-5}$$
$$\tilde{C}_C(n_i) = \frac{n-1}{\sum_{j=1}^{n} d_{ij}} = \frac{n-1}{1+1+1+(n-4)(2)} = \frac{n-1}{2n-5}$$

for any node on the ring the 2 nodes beside it and the hub node have a distance of 1 and the rest $n-4$ nodes have a distance of 2 which goes through the hub

3. Derive the betweenness centrality value for node C .

we know that C is in every shortest path except for the neighbors of each node so 2 nodes are gone and we also know that for neighbors of our neighbors we have 2 paths which one goes through C so 2 nodes are gone and for the rest of the paths they all go through C so we would have $n-1$ nodes on the ring which have the following equation:

$$\delta_{uv}(n_i) = \frac{\sigma_{uv}(n_i)}{\sigma_{uv}}$$
$$C_B(n_i) = \sum_{u \neq n_i} \sum_{v \neq n_i} \delta_{uv}(n_i)$$
$$\delta_{uv}(C) = \begin{cases} 0 & d_{uv} = 1 \\ \frac{1}{2} & d_{uv} = 2 \\ 1 & d_{uv} > 2 \end{cases} \qquad where\ C \notin\ d_{uv}$$
$$C_B(C) = \frac{1}{2}(n-1)(0+0+0+\frac{1}{2}+\frac{1}{2}+(n-6)) = \frac{(n-1)(n-5)}{2}$$

(c) Consider an undirected tree of n nodes. A particular edge in the tree joins node 1 and 2 and divides the tree into two disjoint regions of n1 and n2 nodes, as sketched below:

Show that the closeness centrality C1 and C2 of the two nodes are related by:

$$\frac{1}{C_1} + \frac{n_1}{n} = \frac{1}{C_2} + \frac{n_2}{n}$$

The proof goes as follows:

$$
\begin{aligned}
C_C(1) &= \frac{n}{\sum\limits_{j=1}^{n} d_{1j}} = \frac{n}{\sum\limits_{j=1}^{n_1} d_{1j} + n_2 + \sum\limits_{j=n_1}^{n_1+n_2} d_{2j}} \\
C_C(2) &= \frac{n}{\sum\limits_{j=1}^{n} d_{2j}} = \frac{n}{\sum\limits_{j=n_1}^{n_1+n_2} d_{2j} + n_1 + \sum\limits_{j=1}^{n_1} d_{1j}} \\
\frac{n}{C_C(1)} &= n_2 + \sum_{j=1}^{n_1} d_{1j} + \sum_{j=n_1}^{n_1+n_2} d_{2j} \\
\frac{n}{C_C(2)} &= n_1 + \sum_{j=n_1}^{n_1+n_2} d_{2j} + \sum_{j=1}^{n_1} d_{1j} \\
\frac{n}{C_C(1)} - n_2 &= \frac{n}{C_C(2)} - n_1 \\
\frac{1}{C_C(1)} - \frac{n_2}{n} &= \frac{1}{C_C(2)} - \frac{n_1}{n} \\
\frac{1}{C_C(1)} + \frac{n_1}{n} &= \frac{1}{C_C(2)} + \frac{n_2}{n} \\
\frac{1}{C_1} + \frac{n_1}{n} &= \frac{1}{C_2} + \frac{n_2}{n}
\end{aligned}
$$

Question 3

Stress Centrality

(a) Given a directed graph G = (V, E) , the stress centrality of a node and an edge are defined as::

$$
\begin{aligned}
C_s(n_i) &= \sum_{u \neq n_i \in V} \sum_{v \neq n_i \in V} \sigma_{uv}(n_i) \\
C_s(e) &= \sum_{u \in V} \sum_{v \in V} \sigma_{uv}(e)
\end{aligned}
$$

where $\sigma_{st}(v)$ counts the number of shortest paths between s and t that pass through node v, and $\sigma_{st}(e)$ counts the number of shortest paths between s and t that pass through edge e.

(b) Explain the relationship between the stress centrality C_s of a node v and the stress centrality values of the edges incident to v (denoted by $\Gamma(v)$). Here, $\Gamma(v)$ represents the set of edges that have v as one of their endpoints.

To derive this relationship, we decompose the total stress centrality of all edges incident to v, denoted as $\Gamma(v)$. Every shortest path that utilizes an incident edge falls into one of two categories:

1. **Paths traversing through v:** For any shortest path where v is an internal node ($u \to \cdots \to v \to \cdots \to t$), the path must enter v through one incident edge and exit through another. Consequently, each such path contributes exactly 2 to the sum of incident

edge stresses. This component represents $2C_s(v)$.

2. **Paths starting or ending at v:** Shortest paths where v is either the source or the destination ($v \to \cdots \to t$ or $u \to \cdots \to v$) utilize only one edge in $\Gamma(v)$. Since the definition of node stress centrality $C_s(v)$ typically excludes paths where v is an endpoint, these must be summed separately.

The relationship is formally expressed as:

$$\sum_{e\in\Gamma(v)} C_s(e) = \underbrace{2\sum_{u\in V\setminus\{v\}}\sum_{t\in V\setminus\{u,v\}} \sigma_{ut}(v)}_{\text{paths passing through } v} + \underbrace{\sum_{u\in V\setminus\{v\}} (\sigma_{uv}+\sigma_{vu})}_{\text{paths starting or ending at } v}$$

Question 4

Structural Analysis of Political Power In this assignment, you will analyze the interaction network

(mutual likes) of 5,768 politicians worldwide on Facebook. This network is an Induced Subgraph containing exclusively political nodes. You must transition univariate statistical analysis to deep structural analysis to demonstrate how a politician's topological position determines their real-world role.

(a) Power Geometry (Quantity, Quality, and Access)

Power in networks manifests in three forms: the volume of connections (Quantity), the importance of connections (Quality), and the speed of access to the entire network (Access).

1. Centrality Calculations: Calculate Normalized Degree , Eigenvector Centrality , and Closeness Centrality for all. Extract and report the top 10 nodes for each metric.

We computed the *Normalized Degree Centrality*, *Eigenvector Centrality*, and *Closeness Centrality* for all nodes in the network. The top 10 nodes for each metric are reported below.

- **Top 10 by Normalized Degree Centrality:**
  (a) Manfred Weber
  (b) Joachim Herrmann
  (c) Katarina Barley
  (d) Arno Klare MdB
  (e) Katja Mast
  (f) Barack Obama
  (g) Angela Merkel
  (h) Niels Annen
  (i) Martin Schulz
  (j) Sir Peter Bottomley MP
- **Top 10 by Eigenvector Centrality:**

Question 4 continued on next page. . .

(a) Katarina Barley
(b) Arno Klare MdB
(c) Katja Mast
(d) Christian Petry
(e) Heike Baehrens
(f) Klaus Mindrup
(g) Michelle Müntefering
(h) Niels Annen
(i) Johannes Schraps
(j) Sigmar Gabriel

- **Top 10 by Closeness Centrality:**

(a) Barack Obama
(b) Michael Roth
(c) Niels Annen
(d) Tanja Fajon
(e) Malcolm Turnbull
(f) Mariano Rajoy Brey
(g) Achim Post
(h) Peter Tauber
(i) Dietmar Nietan
(j) Hillary Clinton

2. Gap Analysis: Generate a Scatter Plot with Degree on the X-axis and Eigenvector on the Y-axis. Identify nodes that deviate significantly positively from the correlation line (Low Degree but High Eigenvector).

We performed a gap analysis by generating a scatter plot shown in Figure 1 with *Normalized Degree Centrality* on the X-axis and *Eigenvector Centrality* on the Y-axis. A linear regression line was fitted to capture the overall correlation between quantity of connections (degree) and quality of connections (eigenvector).
Nodes that lie significantly *above* the regression line exhibit **high eigenvector centrality despite relatively low degree**, indicating disproportionate structural influence or *power efficiency*. These nodes are identified via large positive residuals from the regression.
The top 10 positive deviants (low degree / high eigenvector) are:

(a) Christian Petry

(b) Heike Baehrens

(c) Katarina Barley

(d) Klaus Mindrup

(e) Arno Klare MdB

(f) Michelle Müntefering

(g) Katja Mast

(h) Sigmar Gabriel

(i) Johannes Schraps

(j) Wolfgang Hellmich

These actors occupy strategically influential positions by being strongly connected to highly central nodes rather than maintaining a large number of connections. This highlights a distinction between *connection volume* and *connection quality*, which is not captured by degree centrality alone.

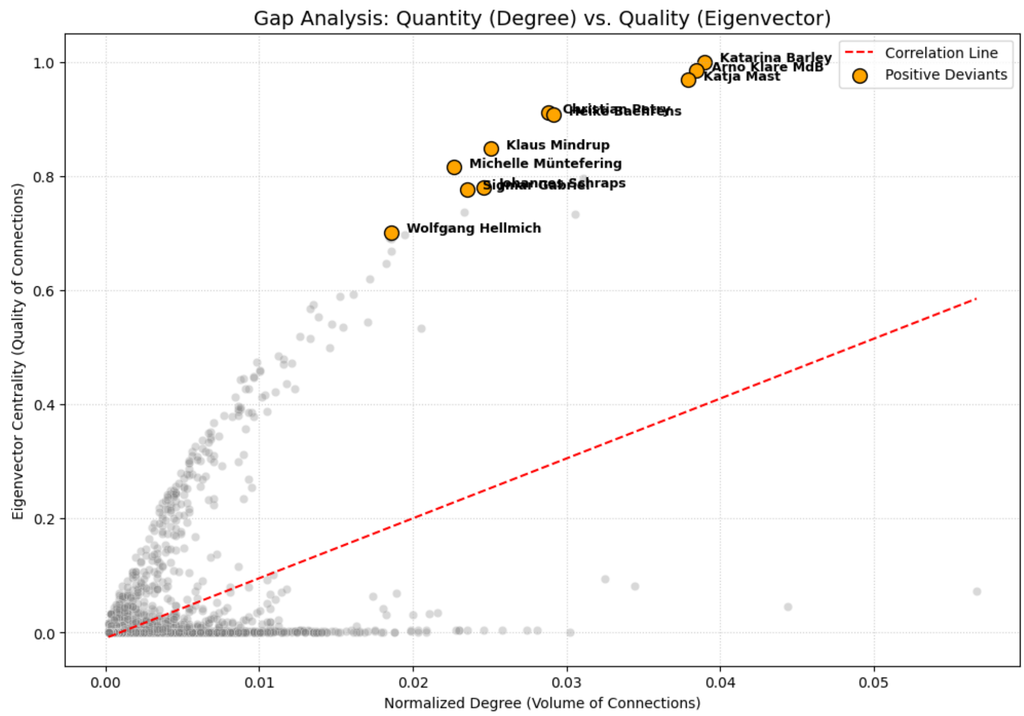

Figure 1: scatter plot *Normalized Degree Centrality* vs *Eigenvector Centrality* showing positive deviants.

3. Three-Way Case Study: Select three politicians who exhibit Low Degree (ranked outside the top 100) but High Eigenvector (ranked within the top 50). Closeness Analysis: Examine the Closeness rank of these three individuals Role Analysis: Investigate the real-world names and positions of these selected individuals. Does the mathematical analysis corroborate their actual roles (e.g., Chief of Staff, Executive Secretary, or Senior Advisor)?

To connect structural metrics with real-world roles, we selected three politicians who exhibit *Low Degree Centrality* (ranked outside the top 100) but *High Eigenvector Cen-*

Question 4 continued on next page...

trality (ranked within the top 50). This combination highlights actors who are not broadly connected, yet are embedded within highly influential neighborhoods of the network.

Interpretive Framework (Closeness Centrality):

- **High Closeness:** The individual resides in the *geometric heart* of the network, acting with relative independence and direct access to many parts of the graph.
- **Low Closeness:** The individual is *structurally dependent* on powerful neighbors, indicating a marginal attachment to the core and influence mediated through elite ties.

(a) **Carsten Schneider** (Degree Rank $= 228.5$, Eigenvector Rank $= 43$, Closeness Rank $= 29$) Schneider's high closeness places him near the geometric center of the network, suggesting operational independence despite a limited number of direct connections. This aligns with his real-world role as a senior parliamentary coordinator in the German Bundestag, where influence stems from central positioning rather than high interaction volume.

(b) **Bernd Lange** (Degree Rank $= 118$, Eigenvector Rank $= 44$, Closeness Rank $= 14$) Lange exhibits both high eigenvector and very high closeness, indicating strong central embeddedness and autonomy. This corroborates his role as Chair of the European Parliament's Committee on International Trade, a position that naturally situates him at the network's core with direct access to multiple influential actors.

(c) **Barbara Hendricks** (Degree Rank $= 130.5$, Eigenvector Rank $= 33$, Closeness Rank $= 283.5$) In contrast, Hendricks shows low closeness despite high eigenvector centrality. This pattern indicates reliance on a small set of powerful neighbors rather than broad network accessibility. Such a structure is consistent with her status as a former Federal Minister, where residual influence is maintained through elite, high-impact ties rather than central operational presence.

Conclusion: Incorporating closeness centrality sharpens the interpretation of influence. High eigenvector centrality alone captures *who one is connected to*, while closeness reveals *how independently* that influence is exercised. The three cases demonstrate two distinct modes of power: centrally embedded autonomous actors versus peripheral elites whose influence is mediated through powerful neighbors—both of which are accurately reflected by the network metrics.

(b) Information Bottlenecks

Identify actors who are not necessarily the most famous, but who control the vital arteries of connection within the network.

1. Calculations & Ranking: Calculate Betweenness Centrality for the entire network. Extract the top 10 nodes and report their corresponding Degree Rank alongside their names.

Question 4 continued on next page...

We computed the *Betweenness Centrality* for all nodes in the network, capturing the extent to which a node lies on the shortest paths between others and thus acts as a broker or bridge. The top 10 nodes based on betweenness centrality, along with their corresponding degree ranks, are reported below.

Real Name	**Betweenness Rank**	**Degree Rank**
Barack Obama	1	6
Hillary Clinton	2	145.5
Angela Merkel	3	7
Justin Trudeau	4	27
Malcolm Turnbull	5	35.5
Manfred Weber	6	1
Peter Tauber	7	38
Betinho Gomes	8	122
Niels Annen	9	8
Boris Johnson	10	679.5

2. Rank Gap Analysis: Examine the table for individuals with Top-Tier Betweenness (Top 10) but Lower Degree. These individuals are mathematical Bridges. Explain the structural difference between their position and that of Hubs.

We examined nodes with *Top-Tier Betweenness Centrality* (top 10) but comparatively lower *Degree Centrality*. These individuals act as **mathematical bridges** in the network—nodes that connect otherwise distant parts of the graph without necessarily having many direct connections themselves.
Observation from the Top 10 Betweenness Table:

- Nodes such as **Hillary Clinton** (Degree Rank $= 145.5$) and **Boris Johnson** (Degree Rank $= 679.5$) exemplify this phenomenon. Despite having moderate or low degree, they appear in the top 10 for betweenness because they lie on many shortest paths that connect different clusters.
- In contrast, traditional **hubs** like **Manfred Weber** (Degree Rank $= 1$) or **Angela Merkel** (Degree Rank $= 7$) are highly connected nodes that accumulate influence through sheer volume of ties rather than strategic positioning.

Structural Difference:

- **Bridges:** Moderate or low degree, strategically positioned on key paths, critical for network connectivity between clusters, often controlling information or influence flow.
- **Hubs:** High degree, central in local neighborhoods, influence derived from extensive direct connections rather than bridging disparate regions.

Conclusion: Betweenness highlights nodes that maintain *structural control* over the network, whereas degree identifies nodes with *volumetric influence*. Bridges and hubs represent complementary forms of power: one by connecting clusters, the other by commanding a dense local neighborhood.

3. Contextual Role Analysis: Select three mediators from the list above. Using reliable sources, identify their real-world positions. Explain which countries, parties, or international organizations they bridge. (Reference the diversity of their Facebook friends' nationalities as evidence).

We selected three individuals from the betweenness centrality ranking—**Malcolm Turnbull**, **Hillary Clinton**, and **Betinho Gomes**—to analyze whether their roles as mathematical bridges correspond to real-world bridging positions across countries, parties, or international networks.

(a) **Malcolm Turnbull Real-World Position:** Former Prime Minister of Australia (2015–2018), leader of the Liberal Party of Australia, and long-standing Member of Parliament.:contentReference[oaicite:0]index=0 **Bridging Role:** As Prime Minister, Turnbull acted as a key interlocutor between Australia and major global partners including the United States, United Kingdom, and Asia-Pacific nations. His high betweenness likely reflects cross-regional links—politically between differing factions within the Commonwealth and economically between advanced and emerging markets. Turnbull's public profile and leadership in major trilateral and multilateral forums suggest social media ties that span diverse nationalities and political milieus, consistent with structural bridging.

(b) **Hillary Clinton Real-World Position:** Former United States Secretary of State (2009–2013), U.S. Senator, and presidential candidate. Clinton has extensive diplomatic experience across Europe, Asia, the Middle East, and Africa.:contentReference[oaicite:1]index=1 **Bridging Role:** Clinton's role as Secretary of State placed her at the center of U.S. foreign policy and major international negotiations (e.g., United Nations engagements, NATO diplomacy). Her network position as a bridge likely captures ties not only to U.S. political elites but also to foreign governments, international NGOs, and global civil society — reflecting heterogeneous nationalities in her extended professional and public networks. High betweenness despite moderate degree suggests influence through strategic elite connections rather than broad grassroots ties.

(c) **Betinho Gomes Real-World Position:** Brazilian politician; former member of the Chamber of Deputies representing Pernambuco, affiliated with the Brazilian Social Democracy Party. He served in both state and federal legislatures (2003–2006, 2011–2019).:contentReference[oaicite:2]index=2 **Bridging Role:** Gomes occupies a different scale of bridging: within Brazilian domestic politics, he connects regional actors (Pernambuco political circles) to the national legislature and party coalitions. His betweenness suggests that, although not a global political figure like Clinton or Turnbull, he functions as a connector across factions or legislative blocs within Brazil's party system and between urban/rural constituencies. In network terms this reflects bridging across sub-communities rather than international political blocs.

for facebook connection only the following section was analyzed and it goes like:

(a) Malcolm Turnbull's Facebook connections demonstrate his structural role as a bridge across diverse media, political, and institutional networks. His friends and pages include national media outlets (e.g., ABC News, Media Watch, The

Chaser, The Age), political figures and organizations (Liberal Party of Australia, Young Liberal Movement, Tony Abbott, Scott Morrison, Josh Frydenberg), international leaders (Barack Obama, David Cameron, Narendra Modi, Christine Lagarde), and civic and cultural institutions (CSIRO, Australian Red Cross Lifeblood, Woollahra Council, Penrith Panthers). This diversity reflects both domestic and international reach, spanning party lines, governmental agencies, media, and global networks, which supports his mathematical position as a high-betweenness node: Turnbull connects otherwise separate clusters, mediating information and influence across multiple communities.

(b) Hillary Clinton's Facebook connections highlight her bridging role across political, educational, and advocacy networks. Her connections include key political figures (President Bill Clinton, Kamala Harris, Joe Biden), family and legacy links (Chelsea Clinton), and institutions promoting education and women's rights (Wellesley College, Women's Suffrage National Monument Foundation). This network reflects both domestic and international influence, spanning U.S. federal politics, gender equality advocacy, and educational leadership, reinforcing her high-betweenness position: she connects otherwise distinct clusters of political, familial, and institutional networks, mediating influence across multiple spheres.

(c) Betinho Gomes' Facebook connections illustrate his role as a political and regional bridge within Brazil. His network includes local and state-level actors (e.g., Regivaldo Antonio, Raul Henry, Armando Monteiro, Fernando Bezerra Coelho), government institutions and programs (Ministério da Integração e do Desenvolvimento Regional, Programa Cidades Sustentáveis, Governo de Pernambuco), party affiliations (PSDB, Roberto Freire, Luciana Santos), and regional media and cultural organizations (Diario de Pernambuco, Cultura PE, Porta dos Fundos). Additionally, he connects to national political figures and media outlets (Lula, Dilma Rousseff, Aécio Neves, Época, VEJA). This diversity reflects his structural position as a high-betweenness node: despite a moderate number of direct connections, he links regional, state, and national political and civic clusters, effectively mediating influence across multiple communities within Brazil.

(c) Power in Local Structures (Efficiency & Visualization)

Closeness Centrality serves as an index for access speed and independence. The goal is to identify Efficient Monitors: politicians who achieve optimal geometric positioning with minimal communication cost

1. Calculations: Calculate Closeness Centrality and Normalized Degree for all nodes. List the Top 10 Closeness nodes.

Closeness centrality and normalized degree were computed for all nodes in the network. Table below reports the top 10 politicians ranked by closeness centrality, representing the most central actors in terms of average shortest-path distance to all other nodes (the geometric heart of the network).

Question 4 continued on next page...

Real Name	Closeness	Normalized Degree
Barack Obama	0.3522	0.0344
Michael Roth	0.3093	0.0205
Niels Annen	0.3074	0.0311
Tanja Fajon	0.3020	0.0079
Malcolm Turnbull	0.2996	0.0198
Mariano Rajoy Brey	0.2993	0.0112
Achim Post	0.2990	0.0221
Peter Tauber	0.2968	0.0190
Dietmar Nietan	0.2939	0.0234
Hillary Clinton	0.2938	0.0107

These nodes form the geometric core of the network, as they can reach all other nodes with the smallest average number of steps. Notably, high closeness does not necessarily coincide with a high normalized degree, indicating that strategic positioning in the network is not solely determined by direct connectivity.

2. Statistical Exploration: Plot Normalized Degree (X-axis) vs. Closeness Centrality (Y-axis).

We plotted *Normalized Degree Centrality* against *Closeness Centrality* for all nodes in the network to examine how local connectivity relates to global accessibility. The resulting scatter plot is shown in Figure 2.
Distribution Analysis: The distribution exhibits a dense clustering in the lower-left corner, transitioning into a sparse "efficiency frontier" along the upper boundary.

- **The Core Cluster:** The vast majority of nodes reside at low values for both metrics, indicating a network dominated by peripheral actors with limited reach.
- **Diminishing Returns:** We observe a logarithmic-like trend where Closeness Centrality increases sharply with initial gains in Degree, but levels off as Normalized Degree exceeds 0.02. This suggests that after a certain threshold of popularity, adding more connections provides negligible improvements to an actor's "closeness" to the rest of the network.
- **The Efficiency Gap:** The most notable feature is the vertical spread at low degree values. Nodes with identical degree counts show vastly different closeness scores, highlighting that *who* one is connected to is mathematically more significant for global access than *how many* connections one possesses.

Structural Liaisons: The **top-left quadrant** identifies actors who circumvent the typical correlation between quantity and access. As seen in the table below, these individuals occupy structurally advantageous positions, serving as efficient access points rather than highly connected hubs.

Real Name	Closeness Rank	Degree Rank
Hillary Clinton	10.0	145.5
Bernd Lange	14.0	118.0
Gianni Pittella	18.0	1236.5

The placement of **Gianni Pittella** is particularly significant; his high Closeness rank relative to a very low Degree rank suggests he acts as a "bridge" to high-degree hubs, allowing him to reach the network core with minimal direct social ties.

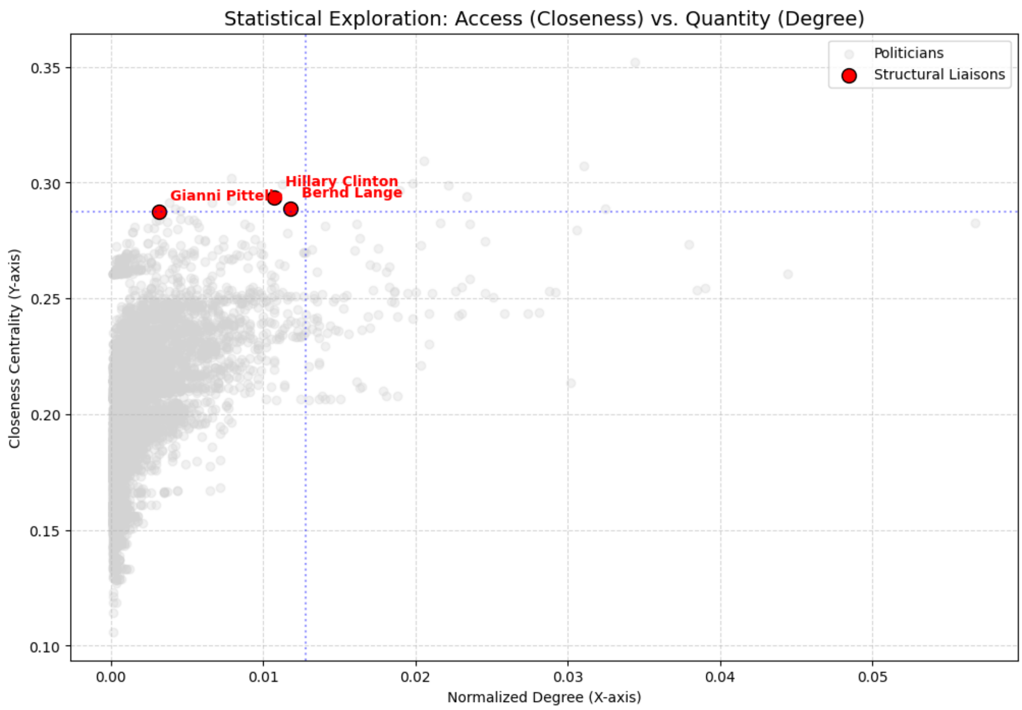

Figure 2: Scatter plot of *Normalized Degree Centrality* vs. *Closeness Centrality*. The distribution highlights a dense peripheral core and a sparse set of efficient structural liaisons in the upper-left quadrant.

3. To understand the network architecture surrounding these individuals, visualize the Ego Network of one selected politician. Morphological Analysis: Is the central individual surrounded by a dense cluster, or are their neighbors dispersed across separate branches? How does this visual structure justify the high Closeness score?

We selected **Hillary Clinton** as the representative politician from the top-left quadrant of the Normalized Degree vs. Closeness plot. The ego network shown in Figure 3 exhibits a hub-and-spoke morphology, where the central individual connects directly to a large number of otherwise weakly interconnected neighbors. Rather than forming a dense local cluster, the neighborhood is dispersed across multiple shallow branches. This structure results in very short average path lengths from the central node to the rest of the network, visually and structurally justifying her high closeness centrality score.

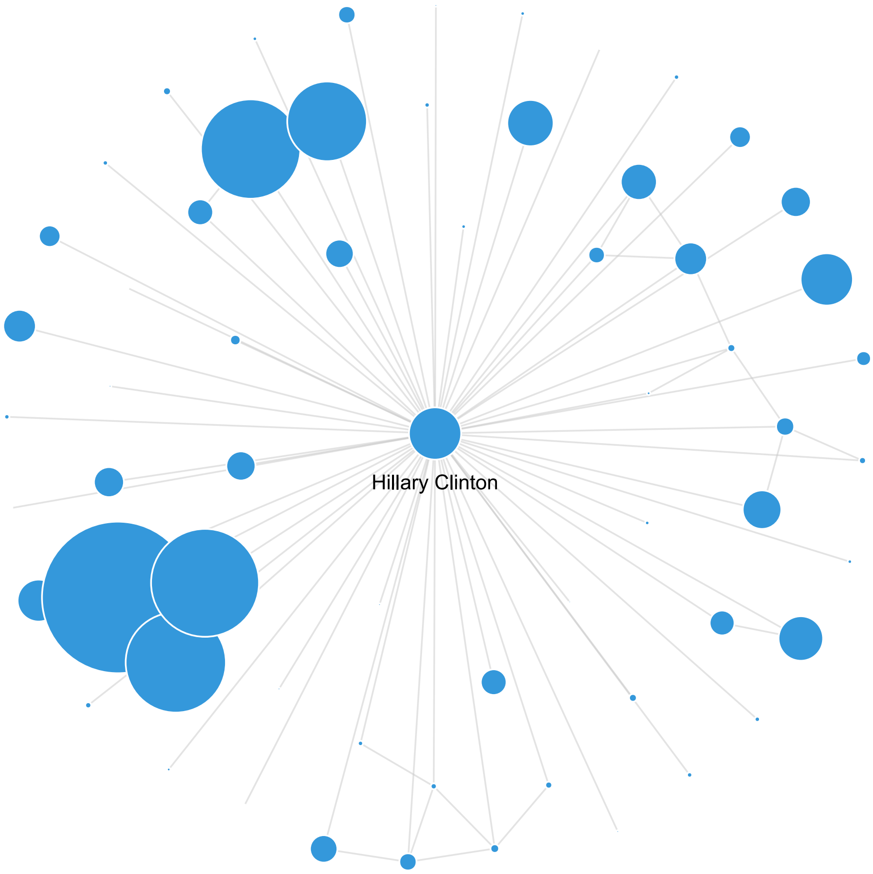

Figure 3: Ego network of **Hillary Clinton**. The visualization shows a hub-and-spoke structure with dispersed neighbors rather than a dense cluster, explaining her high Closeness Centrality despite moderate degree.

4. Contextual Analysis: Based on their real-world titles, explain why their job description requires them to be at the geometric center of the graph. Contrast their structural position with Political Hubs (like Barack Obama, who possesses both High Degree and High Closeness).

The politicians in the top-left quadrant—**Hillary Clinton**, **Bernd Lange**, and **Gianni Pittella**—occupy positions of high closeness but relatively low degree. Their real-world roles, such as senior legislators, committee chairs, or influential policymakers, require them to efficiently access and influence a wide range of other actors without maintaining numerous direct connections. Being at the geometric center allows them to reach most nodes in the network quickly, facilitating coordination, negotiation, and information flow across political factions.
For example:

- **Hillary Clinton** (Closeness Rank 10, Degree Rank 145.5) – Former Secretary of State and Senator, able to access multiple political spheres efficiently.
- **Bernd Lange** (Closeness Rank 14, Degree Rank 118) – European Parliament member involved in international trade policy, requiring rapid access to diverse committees.
- **Gianni Pittella** (Closeness Rank 18, Degree Rank 1236.5) – Senior MEP coordinating across numerous legislative areas, benefiting from efficient network positioning despite lower local connectivity.

In contrast, political hubs like **Barack Obama** combine high closeness with high degree, maintaining many direct connections while also having short paths to the rest of the network. While top-left quadrant politicians excel in efficiency of access and strategic positioning, hubs excel in both visibility and connectivity, making them highly influential across multiple network layers.

(d) Bonacich Power Dynamics

In this section, the role of Bonacich power in the political network is examined. First, compute the largest eigenvalue of the adjacency matrix to ensure convergence. Then,calculate power for three scenarios:

1. Spectral Calculation and Power Regimes: Calculate the centrality scores under the following conditions:

The largest eigenvalue of the adjacency matrix was computed to ensure convergence of the Bonacich power calculation:

$$\lambda_{\text{max}} = 60.6432, \quad \text{Convergence Bound } \frac{1}{\lambda_{\text{max}}} = 0.0165$$

We define $\kappa = 1/\lambda_{\text{max}}$ and calculate the Bonacich power centrality under three scenarios:

- **Neutral Regime** ($\beta = 0$): Centrality depends purely on direct connections without reinforcement or suppression from neighbors.
- **Supportive Regime** ($\beta = 0.5 \cdot \kappa \approx 0.00825$): Positive influence from neighbors increases the centrality of nodes connected to already powerful nodes.
- **Suppressive Regime** ($\beta = -0.5 \cdot \kappa \approx -0.00825$): Connections to powerful neighbors reduce a node's centrality, emphasizing nodes that are less connected to influential actors.

The Bonacich power centrality is then calculated using:

$$\mathbf{c}(\alpha, \beta) = (\mathbf{I} - \beta\mathbf{A})^{-1}\mathbf{1} \cdot \alpha$$

where α is a scaling factor (commonly set to 1), $\mathbf{A}$ is the adjacency matrix, and β is set according to each regime. Each scenario highlights different structural advantages within the political network.

2. Structural Classification and Visual Analysis: Compare node rankings obtained under the different Bonacich power parameter settings (neutral, supportive, and suppressive). Perform the analysis in the following steps:

To analyze how nodes respond to changes in the Bonacich power parameter β, we tracked the rank trajectories of key nodes across the three regimes (neutral, supportive, suppressive) using a slope chart (Figure 4).
Based on both the quantitative rank data and visual trends, nodes were categorized into three structural roles:

Structural Role	Count of Nodes
Stable Actor	3112
Power Inhibitor	1389
Power Amplifier	1195

Interpretation:

- **Stable Actors**: These nodes exhibit minimal change in rank across the three regimes. Their centrality is largely determined by local connections, and they are neither strongly reinforced nor diminished by neighbors.
- **Power Amplifiers**: Nodes in this category experience substantial rank increases under the supportive regime ($\beta > 0$). Their structural position—typically connected to other influential nodes—allows them to benefit from positive feedback effects, amplifying their overall network power.
- **Power Inhibitors**: These nodes experience significant rank drops under the suppressive regime ($\beta < 0$). They are often linked to highly central neighbors, which in this regime reduces their relative power. This reflects the attenuation effect of being connected to already powerful nodes when influence is inversely weighted.

Overall, the slope chart and classification reveal how network structure modulates power: nodes may gain or lose influence not only due to their direct connections but also through the amplification or inhibition transmitted by neighbors, illustrating the dynamic nature of Bonacich power in complex political networks.

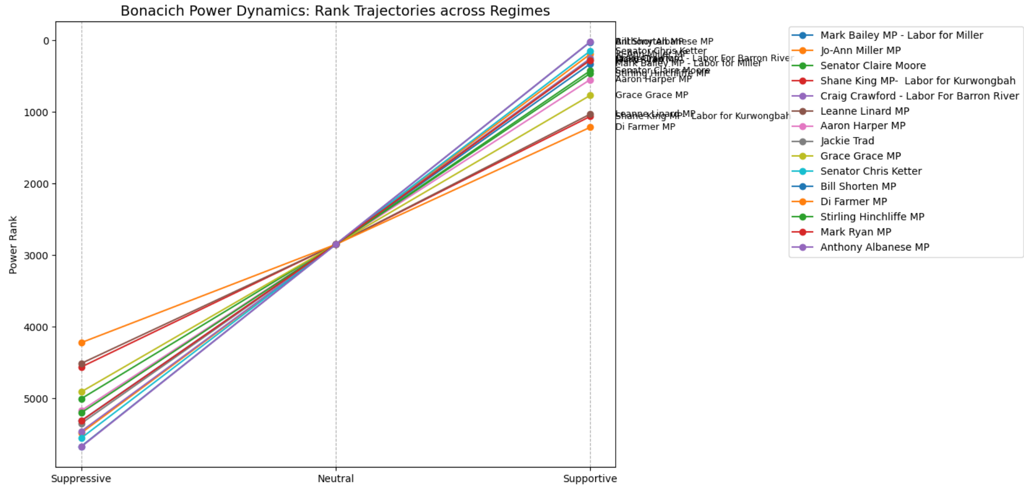

Figure 4: Slope Chart tracking the rank trajectories of key nodes across the neutral, supportive, and suppressive Bonacich power regimes.

Question 5

Comparative Analysis of Ranking Algorithms in Directed Networks In directed networks such as elections or scientific citations, the direction of edges signifies the flow of reputation. The objective of this assignment is to empirically observe the structural differences between two fundamental ranking methodologies: the HITS algorithm (which relies on endorsement by active peers or Hubs) and the PageRank algorithm (which operates on the principle of weighted voting). The analysis will be performed on the Wiki-Vote.txt dataset, which represents the voting network of Wikipedia users for administrator elections.

(a) Ranking Comparison (HITS vs. PageRank)

In this section, you will investigate whether the individuals identified as competent administrators (Authorities) by HITS correspond to those selected by PageRank.

1. Calculation and Mapping: Execute both algorithms—HITS to extract Authority Scores and PageRank with the standard damping factor of $\alpha = 0.85$. To facilitate a meaningful comparison, convert the raw scores into Ranks for each node (where Rank 1 represents the highest score). Visualize the divergence between these two metrics by generating a Scatter Plot on a Log-Log scale , plotting the Authority Rank on the horizontal axis and the PageRank Rank on the vertical axis.

    - **Methodology and Ranking Logic:** The implementation utilizes `networkx` to calculate HITS Authority scores and PageRank ($\alpha = 0.85$). To transform these into a comparable format, the raw scores s are converted to ordinal ranks using `rankdata(-s, method='ordinal')`. This ensures that the node with the highest score receives Rank 1, and every node is assigned a unique rank based on its relative standing.
    - **Overall Correlation:** The Log-Log scatter plot shown in Figure 5 reveals a strong positive correlation between Authority Rank and PageRank Rank, as ev-

Question 5 continued on next page. . .

idenced by the dense clustering of data points along the $y = x$ dashed red line. This suggests that in the Wiki-Vote network, nodes that are considered "authoritative" (pointed to by high-quality hubs) are generally the same nodes that accumulate high PageRank (weighted votes).

- **High-Rank Consistency (Top Nodes):** At the top of the ranking (near 10^0 and 10^1), there is extremely high agreement between the two algorithms. In the context of Wikipedia administrator elections, this indicates that the most "obvious" or prominent candidates are identified consistently regardless of whether the algorithm prioritizes hub-based reinforcement (HITS) or the random-walk/weighted voting model (PageRank).
- **Mid-to-Low Rank Dispersion:** As we move toward the lower-ranked nodes (10^2 to 10^3), the variance increases. The wider "cloud" of points in this region indicates that for average or less-active users, the two algorithms diverge:
  - **HITS Authority** is highly sensitive to the presence of "Hubs" (nodes that vote for many authorities). If a node is voted for primarily by "ordinary" users who aren't active voters (low Hub score), HITS may rank it lower than PageRank would.
  - **PageRank** is influenced by the global link structure and the damping factor $(1-\alpha)$, which provides a "baseline" rank to all nodes, potentially leading to more stable rankings for nodes in sparse regions of the graph.
- **Conclusion:** The empirical results demonstrate that HITS and PageRank are structurally synergistic for social voting networks. While they rely on different mathematical foundations—eigenvector-based mutual reinforcement vs. stationary distribution of a Markov chain—they converge on the same "elite" set of nodes in the Wiki-Vote dataset.

Question 5 continued on next page. . .

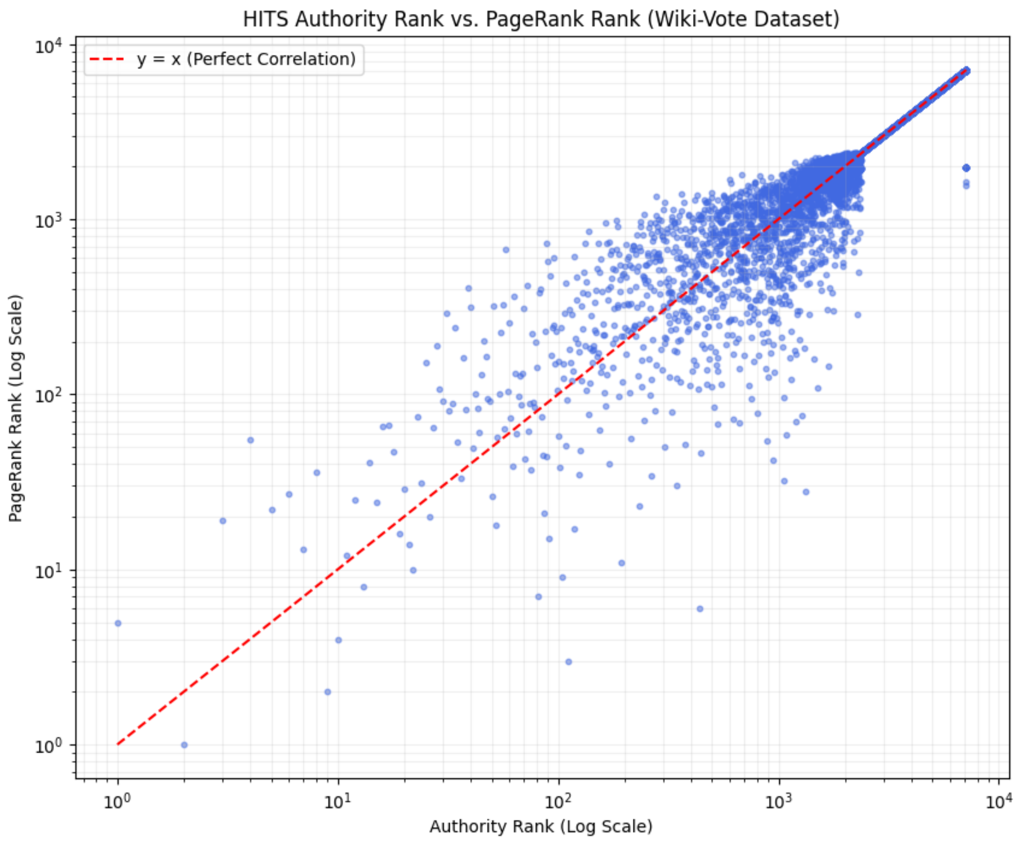

Figure 5: Comparative analysis of node rankings in the Wiki-Vote network using HITS Authority and PageRank ($\alpha = 0.85$). The scatter plot, visualized on a Log-Log scale, illustrates a high degree of correlation between the two metrics; the red dashed line signifies perfect rank-order correspondence ($y = x$).

2. Divergence Analysis: Focus on nodes that deviate significantly from the diagonal line (y = x) in the rank comparison plot. Select representative nodes from different regions of the plot and analyze the structural reasons behind their divergent rankings. In your discussion, you should examine the local and global patterns of incoming links, consider the activity level and connectivity of the nodes endorsing them, and explain how these structural characteristics may lead to different evaluations by HITS and PageRank.

To analyze the divergence between HITS Authority and PageRank in the Wiki-Vote dataset, we examined nodes that deviate significantly from the $y = x$ diagonal. The structural reasons for these discrepancies are categorized below:

Category 1: Authority Rank $\gg$ PageRank (e.g., Nodes 5132, 5637)

Nodes in this region are evaluated as high-quality "Authorities" by HITS but are penalized by PageRank.

- **Structural Observation:** These nodes possess relatively high in-degrees but are pointed to by "Professional Hubs" (nodes with an average out-degree of ≈ 200).

- **Reason for Divergence:**
  - **HITS:** The Authority score is the sum of the Hub scores of its predecessors. Since these nodes are endorsed by major hubs in the Wiki community, their Authority rank is high.
  - **PageRank:** PageRank distributes a node's influence equally among its out-links. Because the predecessors here point to hundreds of other nodes, the "rank juice" passed to any single node is heavily diluted (the $1/L(u)$ factor), resulting in a much lower PageRank.

Category 2: PageRank Rank $\gg$ Authority (e.g., Nodes 7467, 8076)

Nodes in this region appear in the "long tail" of the dataset, where PageRank values them significantly higher than HITS does.

- **Structural Observation:** These nodes have very low in-degrees (1 or 2) and are pointed to by "Exclusive Voters" (predecessors with an out-degree of exactly 1).
- **Reason for Divergence:**
  - **PageRank:** Since the voter points to *only* this node, the node receives 100% of the voter's transferred rank without dilution. This makes the node a "stronger" destination in a random walk.
  - **HITS:** HITS requires a node to be pointed to by a good "Hub" to gain Authority. A voter who only points to a single person has a Hub score of nearly zero, as they do not provide a "directory" of multiple authorities. Consequently, HITS overlooks these nodes.

(b) Rank Stability Analysis

The PageRank algorithm utilizes a parameter α (damping factor), which determines the patience of the random surfer in following links. This section examines how the hierarchy of power shifts as this parameter changes.

1. Simulation & Trajectory Interpretation: Perform a sensitivity analysis by executing PageRank across a spectrum of α values ranging from 0.50 to 0.99 . Construct a Line Chart to visualize the rank trajectories of the top 10 nodes (as well as the specific anomalies identified in the previous section), with the vertical axis representing the rank.

The sensitivity analysis conducted across the spectrum $\alpha \in [0.50, 0.99]$, shown in Figure 6, reveals a clear separation between structurally robust nodes and rank artifacts within the Wiki-Vote network.

- **High-Altitude Stability (Top 10 Nodes):** A key observation is the remarkable stability of the ranking among the **Top 10 nodes**. Across the entire range of α, these nodes exhibit minimal rank fluctuations and almost no rank crossings. This stability indicates that their importance is not dependent on a particular balance between teleportation and link-following, but instead arises from a combination of strong local connectivity and deep integration into the global structure of the network. Consequently, PageRank consistently identifies these nodes as the dominant power-holders regardless of the random surfer's level

Question 5 continued on next page...

of persistence.

- **Category 1 Nodes (Structural Ascent):** Nodes such as 5132 and 5637 exhibit a clear **improvement in rank** as α increases.
  - *Interpretation:* These nodes are strongly favored by HITS because they are pointed to by well-established hubs. Although the individual incoming links may originate from high–out-degree nodes, increasing α allows PageRank to incorporate longer random walks and multi-hop paths. As teleportation diminishes, probability mass accumulates through upstream structural reinforcement, enabling these nodes to rise toward their HITS-perceived authority.
  - *Influence scope:* Their influence is therefore **distributed across distant regions of the graph**, rather than confined to a small local neighborhood.
- **Category 2 Nodes (Structural Descent):** Nodes such as 7033 and 3245 show a pronounced **decline in rank** as α increases.
  - *Interpretation:* These nodes rely on a small number of exclusive, low-degree predecessors and lack meaningful multi-hop reinforcement. At lower α, the high teleportation rate acts as a smoothing mechanism that artificially boosts their rank. As $\alpha \to 0.99$, this smoothing effect disappears and global structural positioning dominates, causing these nodes to be overtaken by more deeply embedded vertices.
  - *Influence scope:* Their importance is primarily **local and fragile**, supported by shallow connectivity rather than by global reachability.

Random Surfer Interpretation: These trends can be directly explained through the behavior of the random surfer. At low α, the surfer frequently teleports, increasing the visibility of peripheral or weakly connected nodes. As α increases, the surfer becomes more persistent in following links, favoring nodes that lie along many long paths and serve as convergence points for probability flow. Consequently, nodes with distributed, multi-hop influence improve in rank, while nodes dependent on teleportation experience systematic rank decay. This demonstrates how increasing α shifts PageRank from a locally smoothed measure toward a globally structural one.

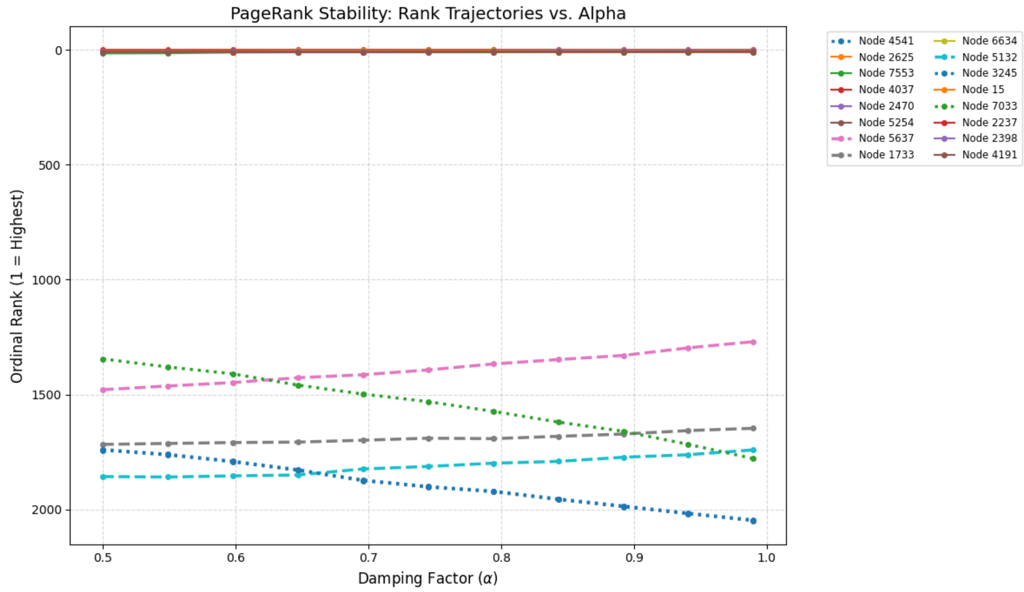

Figure 6: PageRank Stability Analysis: Rank trajectories of top-tier nodes and anomalous outliers across a range of damping factors ($\alpha \in [0.5, 0.99]$). The Y-axis is inverted so that Rank 1 appears at the top. The plot demonstrates the high stability of the Top 10 nodes (reliable core) while highlighting the volatility of Category 1 and Category 2 nodes as the algorithm shifts from local to global structural prioritization.

References

[1] ChatGPT, *Chat session used for assignment guidance*, `https://chatgpt.com/share/69518d24-9800-800d-8fe8-185a102b5f10`, accessed December 28, 2025.