University of Tehran
ECE

Social Network

Instructor: Masoud Asadpour

Teacher Assistant:
Alireza Koohzad
Omid Mollaei

Homework 2
Fall 2025

Submission Guidelines and Policies

Submit your final file as a single **ZIP** file that includes the report in **PDF** format and all code files, and upload it to the **eLearn platform**. The name of the ZIP file should follow the pattern: `SN_HW#_StudentNumber`

- If you have any questions, contact the assignment TAs via **email**. Please avoid sending private messages on social media so that responses can remain organized and efficient.
- The length of the report is not a grading factor. For implementation questions, focus on providing clear explanations; clarity matters far more than word count.
- The procedure for submitting assignments is explained in detail separately in this file and in the Git workshop video. Before submitting your code, ensure that you have watched the entire workshop video.
- Every submission must include both the report and the corresponding code. Any code submitted without a report will receive **zero points**.
- In your assignment report, you must describe how you used these tools. Include details such as the tools you used, their specific applications, and any other relevant information.
- At the end of your report, you must include the link to the prompts used: (`The ChatGPT conversation link`).
- Assignments may be uploaded to eLearn for up to 7 days after the official deadline, but a **5% penalty per day** will be deducted from the grade for each late day. After 7 days, submissions will not be accepted.
- To verify your understanding of each assignment, there will be a brief 5–10 minute in-person or virtual review session. You will be selected for this session **once during the semester**. If there are discrepancies between your submitted report and your presentation, the chance of being chosen again will increase.
- **Plagiarism is strictly prohibited.** Any similarity in the report or code that indicates copying or if cheating occurs during exams, will result in a score of **0.25 for all students involved for the course.**

Contents

Theoretical Questions

Question 1: Structural Index

Consider an organization with n employees. Each employee, based on their assigned duties, interacts with a number of their colleagues. In the communication network of this organization:

- Each node represents an employee of the organization.
- If employee i has a working relationship with employee j, then there is an edge between them ($A_{ij} = 1$).
- This network is undirected, and all interactions are mutual.

Now consider the following functions defined on this organizational network:

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

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

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

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

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

Where:

- $d(u,v)$: Shortest path distance between node u and v.
- $d(v)$: Degree of node v.
- $N(i)$: Neighbors of node i.
- p_{ij}: Proportion of i's total investment (time/attention/interaction weight) devoted to neighbor j.
- w_{ij}: Absolute weight or strength of the connection between nodes j and i.
- $H(v)$: Hours worked online by node v.
- Difference between p_{ij} and w_{ij}: p_{ij} is a relative measure showing what fraction of a node's attention goes to neighbor j, while w_{ij} is an absolute measure representing the raw strength of the connection.

(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.

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

Question 2: Distances and Neighbors

(a) Consider the below network.

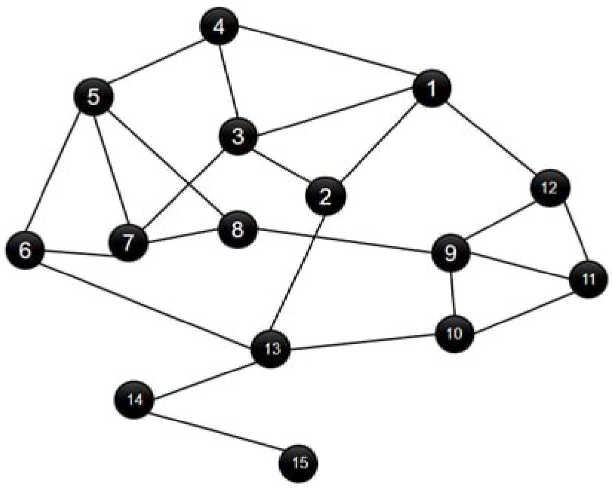

Figure 1: Network of 15 nodes

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

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

(b) Consider the below network.

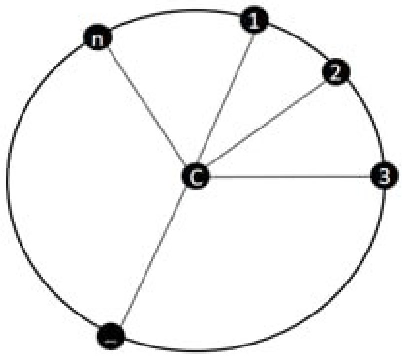

Figure 2: Ring network with a central hub

1. What is the closeness centrality for node C?
2. Derive the closeness centrality value for the nodes on the ring as a function of n.
3. Derive the betweenness centrality value for node C.

(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 n_1 and n_2 nodes, as sketched below:

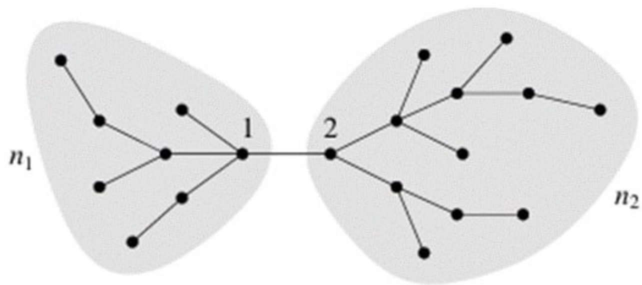

Figure 3: Tree with an edge separating two regions of sizes n_1 and n_2

- Show that the closeness centrality C_1 and C_2 of the two nodes are related by:

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

Hint: Use the formula for closeness centrality: $C_i = \frac{n}{\sum_j d_{ij}}$.

Question 3: Stress Centrality

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

$$C_s(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}(v), \qquad C_s(e) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(e)$$

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.
- 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.

Implementation Questions

Question 1: 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 from univariate statistical analysis to deep structural analysis to demonstrate how a politician's topological position determines their real-world role.

Data & Preprocessing:

- `politician_edges.csv`: The edge list representing connections.
- `politician_nodes.csv`: Node attributes (ID and real names).

(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.
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).
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.
     - *High Closeness:* The individual resides in the Geometric Heart of the network, acting independently.
     - *Low Closeness:* The individual is highly dependent on powerful neighbors (marginal attachment to the core).
   - **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)?

(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.
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.
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).

(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.
2. **Statistical Exploration:**
   Plot **Normalized Degree** (X-axis) vs. **Closeness Centrality** (Y-axis).
   - Analyze the distribution. Identify three politicians distinguished in the Top-Left quadrant (Closeness in Top 20, but Degree outside Top 100) and annotate their names on the plot.
3. **Structural Visualization (Ego Network):**
   To understand the network architecture surrounding these individuals, visualize the **Ego Network** of one selected politician.
   - **Requirements:**
     - Use force-directed algorithms (e.g., Spring Layout).
     - Adjust the spacing parameter (k) to prevent node overlap.
     - Scale node sizes proportional to their importance (Degree) and display the label only for the central node.
   - **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?

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).

(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:
   - **Neutral ($\beta \approx 0$):** Baseline regime for comparison
   - **Supportive ($\beta > 0$):** Positive dependence on neighboring nodes
   - **Suppressive ($\beta < 0$):** Negative dependence on neighboring nodes
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:
   - Construct a Slope Chart (or Bump Chart) that tracks the rank trajectories of key nodes across the three regimes to visualize the dynamic shifts in hierarchy.
   - Based on both quantitative data and visual trends, identify and categorize nodes into three roles: *Power Amplifiers* (substantial rise), *Power Inhibitors* (substantial drop), and *Stable Actors* (invariant).
   - For each group, describe the magnitude of rank shifts and provide a structural interpretation of why these nodes respond differently to variations in β. Relate your findings to the broader concept of power amplification or attenuation through network neighbors.

Question 2: Comparative Analysis of Ranking Algorithms in Directed Networks

(HITS vs. PageRank Dynamics)
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.
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.

(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:**
   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.
2. **Trajectory Interpretation:**
   Analyze the rank trajectories by addressing the following points:
   - Describe how the node's rank changes as the damping factor α increases (e.g., stable, improving, or declining).

- Interpret what this trend implies about the node's position in the network and whether its influence is primarily local or distributed across distant parts of the graph.
- Relate your observations to the behavior of a random surfer with different levels of persistence in following links.