Social Networks: Assignment #1

Due on Monday, December 15, 2025

Dr. Masoud Asadpour 15:00am

Ali Dashtbozorg 810104302

Contents

List of Figures

Question 1

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 1 reveals a strong positive correlation between Authority Rank and PageRank Rank, as evidenced 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.

Question 1 continued on next page...

  - **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.

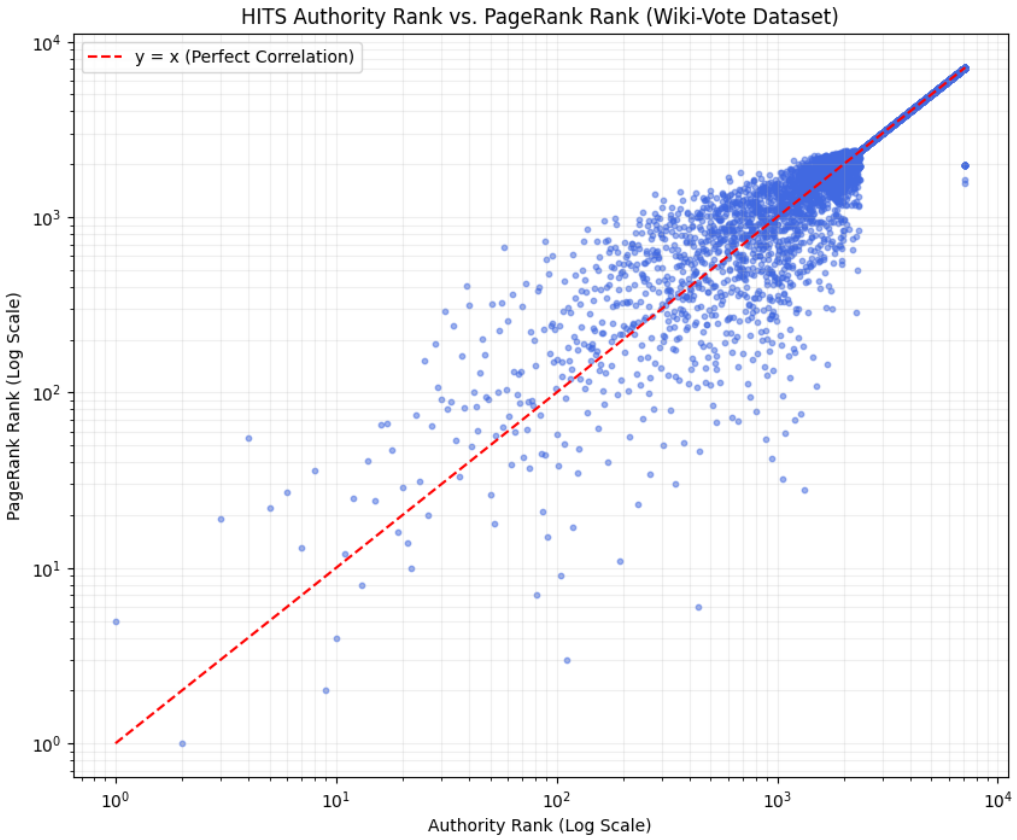

Figure 1: Comparative analysis of node rankings in the Wiki-Vote network using HITS Authority and PageRank ($\alpha = 0.85$). The scatter plot, visualized on a Log-Log scale, illustrates a high degree of correlation between the two metrics; the red dashed line signifies perfect rank-order correspondence ($y = x$).

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.

Question 1 continued on next page...

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.