Mid-term Exam (Graph Mining – Spring 2025)

Full Name:
Student ID:

- The formula and solution process should be presented with the answer.
- Answers must be written in English.

1. (Centrality Measures) Consider an undirected graph G of four nodes (15pt)
   a. Calculate betweenness centrality of node 4

Equation betweenness centrality: $B(v_i) = \sum_{s,t\in V}\frac{\sigma(s,t|v_i)}{\sigma(s,t)}$, where $\sigma(s,t)$ is the number of shortest paths from node s to node t, $\sigma(s,t|v_i)$ is the number of shortest paths from node s to node t that passing through node v_i.

Normalized betweenness centrality: $\bar{B}(v_i) = \frac{\mathrm{B}(v_i)}{(\mathrm{n}-1)\ (\mathrm{n}-2)/2}$ where n is number of nodes.

Ans:
Betweenness centrality of node 4

	$\sigma(s,t)$	$\sigma(s,t\|i)$	$\sigma(s,t\|i)/\sigma(s,t)$
1,2	1	0	0
1,3	2	1	½
1,4	1	0	0
2,3	1	0	0
2,4	1	0	0
3,4	1	0	0

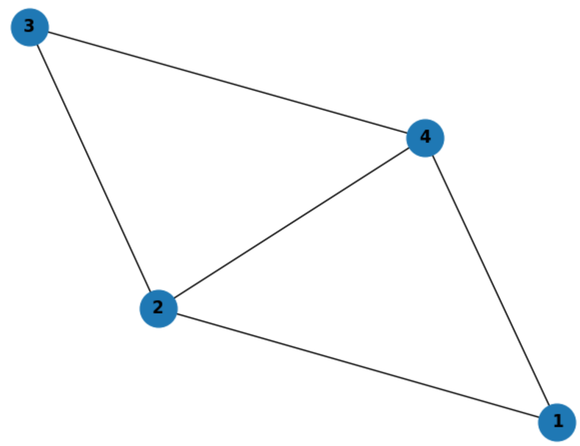
$$\bar{B}(v_i) = \frac{\mathrm{B}(v_i)}{(\mathrm{n}-1)\ (\mathrm{n}-2)/2} = \frac{1/2}{(4-1)\ (4-2)/2} = \frac{1/2}{3*2/2} = \frac{1}{6}$$

   b. Calculate closeness centrality of node 4

Equation closeness centrality: $C(v_i) = \frac{N-1}{\sum_{j=1}^{N-1} d(s,t)}$, where $d(s,t)$ is number of nodes in the shortest path between node s and node t, and N-1 is the number of nodes reachable from v_i.

Ans:

$$C(v_i) = \frac{N-1}{\sum_{j=1}^{N-1} d(v_j v_i,)} = \frac{4-1}{1+1+1} = \frac{3}{3} = 1$$



c. The betweenness centrality and closeness centrality are based on the shortest paths between nodes. What are the roles of the shortest paths in the two centralities, respectively.

Ans:

Centrality Measure	Role of Shortest Paths	Interpretation
Betweenness	Count how often a node lies on shortest paths between other nodes	Identify nodes that act as bridges or bottlenecks
Closeness	Sum the lengths of shortest paths from a node to all others	Identify nodes that can quickly reach all others

2. (Centrality Measures) Consider an undirected graph G of three nodes given in the following figure (15pt)

   a. Calculate Eigenvector, Katz centrality of node 1 with $\alpha = 1$, $\beta = 2$, t = 1

   Equation Eigenvector: x_i (t) = $\sum_{v_j \in N(v_i)} A_{ij} x_j(t-1)$, where A is adjacency matrix, t is time, with the centrality at time t = 0 being $x_j(0)$ = 1 $\forall j$

   Equation Katz: $Katz(G) = \alpha \sum_j A_{ij} x_j + \beta$, where α is damping factor, β is bias constant.

   Ans:

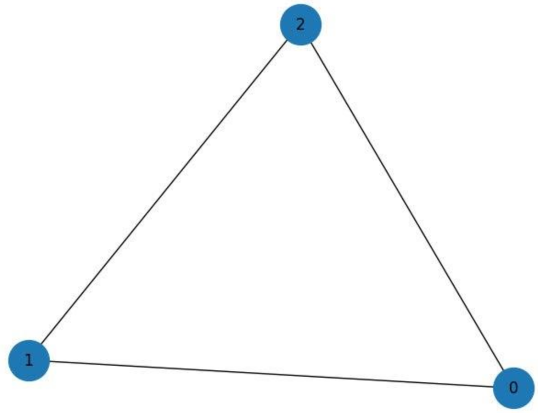
   Eigenvector: x_i (t) = $\sum_{v_j \in N(v_i)} A_{ij} x_j(t-1)$

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$x_j(0) = 1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$x_1(1) = \sum_{v_j} A_{1j} x_j(0) = (0 * 1) + (1 * 1) + (1 * 1) = 2$$

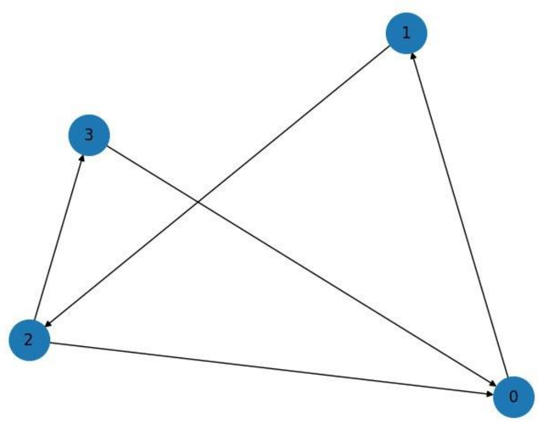
$$Katz(G) = \alpha \sum_j A_{ij} x_j + \beta = 1 * 2 + 2 = 4$$



b. Consider a directed graph G of four nodes given in the following figure, calculate PageRank centrality of node 1, with β = 0.75, given pagerank of (0) = 1, pagerank of (2) = 2, pagerank of (3) = 3

Equation PageRank centrality of node i: $x_i = \sum_{(j,i)\in E} \frac{x_j}{out \deg x_j} + \beta$, where x_j is PageRank score of all pages j that point to page i

$x_1 = \sum_{(j,1)\in E} \frac{x_j}{out \deg x_j} + \beta = \frac{1}{1} + 0.75 = 1.75$



c. In Katz centrality, what does the damping factor mean? What is the difference between Eigenvector centrality and Katz centrality are made by the damping factor?

Ans:

- In Katz centrality, the damping factor is a parameter that controls the influence of indirect connections in a network. It assigns exponentially decreasing weights to longer paths, ensuring that the contribution of distant nodes diminishes with path length. This approach allows Katz centrality to account for both immediate neighbors and the broader network structure, while preventing the measure from diverging.

- Damping factor in Katz centrality introduces a mechanism to account for the influence of indirect connections while ensuring the measure remains finite and applicable to a broader range of network structures compared to eigenvector centrality.

d. In PageRank centrality, what is the purpose of the inverse of source node degree? Then, what is the difference between the static damping factor in Katz centrality and the inverse of degree in PageRank?

Ans:

- In PageRank, the inverse of a source node's out-degree makes sure that all of its outbound links have the same amount of effect. This keeps nodes with a lot of outgoing connections from making their impact too strong.
- Katz centrality uses a damping factor to attenuate the influence of longer paths, allowing nodes to accumulate influence through both direct and indirect connections without normalizing by out-degree.

3. (Graph Visualization) Given undirected graph with 4 nodes A (0, 0), B (1, 0), C (1, 1), and D (0, 1) in the following figure. (10pt)

a. Calculate displacement vector for node A, where spring constant = 1, repulsive force constant = 2 and ideal edge length l = 2, log(1/2)= -0.3, log($\sqrt{2}/2$) $= -0.15$

Repulsive force

$$f_{rep}(u,v) = \frac{c_{rep}}{||p_v - p_u||^2}\overrightarrow{p_u p_v}$$

Attractive force

$$f_{spring}(u,v) = c_{spring}\log\left(\frac{||p_v - p_u||}{l}\right)\overrightarrow{p_u p_v}$$

$$f_{attr}(u,v) = f_{spring}(u,v) - f_{rep}(u,v)$$

Resulting displacement vector

$$F_u = \sum_{v\in V} f_{rep}(v,u) + \sum_{uv\in V} f_{attr}(v,u)$$

Ans:

$\overrightarrow{p_B p_A} = (0-1, 0-0) = (-1,0)$

$\overrightarrow{p_C p_A} = (-1,-1)$

$\overrightarrow{p_D p_A} = (0,-1)$

$$f_{rep}(v,A) = \frac{c_{rep}}{||p_A - p_v||^2}\overrightarrow{p_v p_A} = \frac{c_{rep}}{||p_A - p_B||^2}\overrightarrow{p_B p_A} + \frac{c_{rep}}{||p_A - p_C||^2}\overrightarrow{p_C p_A} + \frac{c_{rep}}{||p_A - p_D||^2}\overrightarrow{p_D p_A}$$

$$= \frac{2*(-1,0)}{1} + \frac{2*(-1,-1)}{2} + \frac{2*(0,-1)}{1} = (-3,-3)$$

$$f_{spring}(B,A) = c_{spring}\log\left(\frac{||p_B - p_A||}{l}\right)\overrightarrow{p_B p_A} = \log\left(\frac{1}{2}\right) = -0.3*(-1,0) = (0.3,0)$$

$f_{spring}(D,A) = (0,0.3)$

$f_{attr}(B,A) = f_{spring}(B,A) - f_{rep}(B,A) = (0.3,0) - (-2,0) = (2.3,0)$

$f_{attr}(D,A) = f_{spring}(D,A) - f_{rep}(D,A) = (0,0.3) - (0,-2) = (0,2.3)$

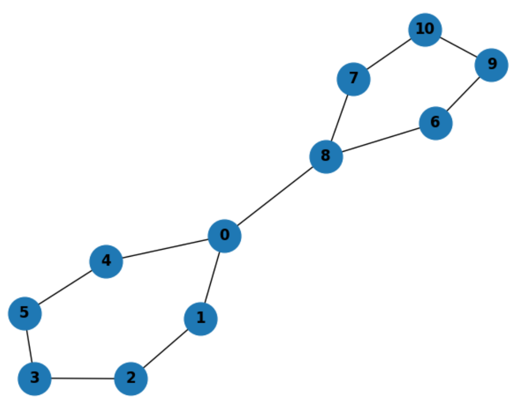
$$F_A = \sum\nolimits_{v \in V} f_{rep}(v, A) + \sum\nolimits_{Av \in V} f_{attr}(v, A) = (-3, -3) + (2.3,0) + (0,2.3)$$
$$= (-0.7, -0.7)$$

b. The attractive force affects only adjacent nodes. Then, why the repulsive force affects both adjacent and non-adjacent nodes? Also, in this regard, please explain which graph visualization is a good visualization.

Ans:

- In force-directed graph layouts, **repulsive forces** act between all node pairs to prevent overlap and distribute nodes evenly, while **attractive forces** act only between connected nodes to maintain structural integrity.
- A good visualization balances these forces, resulting in uniform edge lengths, minimal crossings, and clear clustering.

4. (Community Detection) Consider an undirected graph G of eleven nodes given in the following figure. There are two communities in the graph: A = {0, 1, 2, 3, 4, 5} and B = {6, 7, 8, 9, 10}. (10pt).



a. Calculate Min-cut, Normalized cut measurements of A and B.

Ans:

$\mathrm{assoc}(A), V = 3 + 2 + 2 + 2 + 2 + 2 = 13$

$\mathrm{assoc}(B), V = 3 + 2 + 2 + 2 + 2 = 11$

Min_cut (A, B) = 1 N_cut(A, B) $= \frac{cut(A,B)}{assoc(A),V} + \frac{cut(A,B)}{assoc(B),V} = \frac{1}{13} + \frac{1}{11} = \frac{24}{143} = 0.1678$

b. Calculate conductance of A and B using the equation (1).

$$conductance(A,B) = \frac{cut(A,B)}{\min\,(assoc(A,V), assoc(B,V))} \quad (1)$$

where $assoc(A,V)$ and $assoc(B,V)$ is the total connection from nodes in A and B to all nodes in the graph, respectively. $cut(A,B)$ is the number of cuts between 2 communities A and B.

Ans:

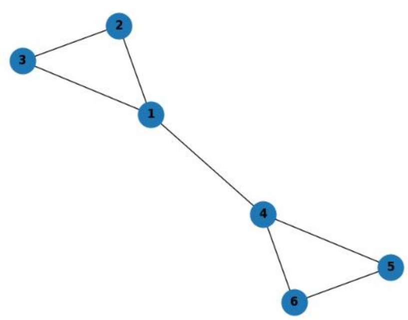
$$conductance(A,B) = \frac{cut(A,B)}{\min\,(assoc(A,V), assoc(B,V))} = \frac{1}{\min\,(13,11)} = \frac{1}{11}$$

c. What is the difference between Min-cut and Normalized cut? Please explain in terms of how the two measures consider intra-compactness and inter-connectivity of communities.

Ans:

	Min-cut	N-cut
Intra-compactness	Not considered	Encouraged through normalization
Inter-connectivity	Minimize total edge weight between subsets	Minimize normalized edge weight between subsets

5. (Community Detection) Consider an undirected graph G of six nodes given in the following figure. (10pt)



a. Calculate the global clustering coefficient C_i

Equation (1): $C_i = \frac{\#of\ closed\ triplets}{\#of\ connected\ triplets}$

Ans:

$$C_i = \frac{\#of\ closed\ triplets}{\#of\ connected\ triplets}$$

$$= \frac{(1,2,3), (4,5,6)}{(1,2,3), (4,5,6), (3,1,4), (2,1,4), (6,4,1), (5,4,1)} = \frac{1}{3}$$

b. Calculate the average clustering coefficient $\langle C \rangle$ in the graph G

Equation (2): $\langle C\rangle = \frac{1}{N}\sum_{i=0}^{N} C_i$

Ans:

Clustering node 1: $C_1 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{3.(3-1)} = 0.333$

Clustering node 2: $C_2 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{2.(2-1)} = 1$

Clustering node 3: $C_3 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{2.(2-1)} = 1$

Clustering node 4: $C_4 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{3.(3-1)} = 0.333$

Clustering node 5: $C_5 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{2.(2-1)} = 1$

Clustering node 6: $C_6 = \frac{2L_1}{d_1(d_1-1)} = \frac{2.1}{2.(2-1)} = 1$

$$\langle C\rangle = \frac{1}{N}\sum_{i=0}^{N} C_i = \frac{4.666}{6} = 0.777$$

c. What is the difference between global clustering and average clustering coefficients? Please explain in terms of how the two metrics measure density of edges within clusters or an entire graph.

Ans:

	Average Clustering Coefficient	Global Clustering Coefficient
Focus	Local neighborhoods	Entire network
Calculation	Average of local clustering coefficients	Ratio of closed triplets to all triplets
Node Weighting	Equal weight to all nodes	Higher-degree nodes have more influence
Sensitivity	Sensitive to local variations	Sensitive to overall network structure
Use Cases	Analyzing local cohesiveness	Assessing global interconnectedness

6. (Community Detection) Consider an undirected graph G of ten nodes given in the following figure with two communities: A = {0, 1, 2, 3, 4} and B = {5, 6, 7, 8, 9}. (5pt)

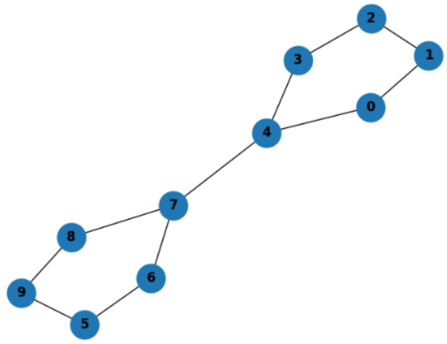
a. The community structure will be strong or weak if there are many pairs (i, j) with $A_{ij} > \frac{d_i d_j}{2m}$.

Ans: The community structure will be strong because many node pairs (i,j) within same community

b. Apply the Equation (1) to calculate the modularity Q of the two communities.

$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\cdot\delta(v_i, v_j) = \frac{1}{2\times 11}\left[\left(1-\frac{d_0 d_1}{22}\right)+\left(1-\frac{d_1 d_2}{22}\right)+\left(1-\frac{d_2 d_3}{22}\right)+\left(1-\frac{d_3 d_4}{22}\right)+\left(1-\frac{d_4 d_0}{22}\right)+\left(1-\frac{d_7 d_4}{22}\right)+\left(1-\frac{d_7 d_8}{22}\right)+\left(1-\frac{d_8 d_9}{22}\right)+\left(1-\frac{d_9 d_5}{22}\right)+\left(1-\frac{d_5 d_6}{22}\right)\right] =$

$\frac{1}{22}\left[\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*3}{22}\right)+\left(1-\frac{2*3}{22}\right)+\left(1-\frac{3*3}{22}\right)+\left(1-\frac{2*3}{22}\right)+\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*2}{22}\right)+\left(1-\frac{2*3}{22}\right)\right]=\frac{1}{22}\left[6*\left(1-\frac{2*2}{22}\right)+4*\left(1-\frac{2*3}{22}\right)\right]=\frac{43}{121}$



(1)

$$\delta(v_i, v_j) = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are in the same community} \\ 0 & \text{otherwise.} \end{cases}$$

where m is the number of total edges, A is the adjacency matrix of G, d_i is the degree of node v_i

c. Which edge is more or less valuable according to Modularity? What kinds of community structures are supposed by Modularity? Explain in terms of correlations of modularity with scale-free graphs.

Ans:

- More valuable edges: edges connecting nodes within the same community are more valuable, as they increase modularity score
- Less valuable edges: edge connecting nodes in different communities are less valuable, as they decrease modularity score
- Modularity favors community structures where nodes within the same group are densely connected, and connections between different groups are sparse.
- Scale-free Graphs: Modularity can detect communities in scale-free graphs, but the resolution limit may cause it to overlook smaller communities, especially in large or highly heterogeneous networks.

7. (Community Detection) Consider an undirected graph G of four nodes given in the following figure (5pt)

a. Calculate k-means clustering with K =2 in one interaction and initial centroids are node 1 and node 4, with Euclidean distance

Ans:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Centroid C1: [0,1,0,0]

Centroid C2: [0,0,1,0]

Euclidean distance for each node to centroids

Node 1 to C1: $\sqrt{(0-0)^2+(1-1)^2+(0-0)^2+(0-0)^2}=0$

Node 1 to C2: $\sqrt{(0-0)^2+(1-0)^2+(0-1)^2+(0-0)^2}=\sqrt{2}$

Node 2 to C1: $\sqrt{(1-0)^2+(0-1)^2+(1-0)^2+(0-0)^2}=\sqrt{3}$

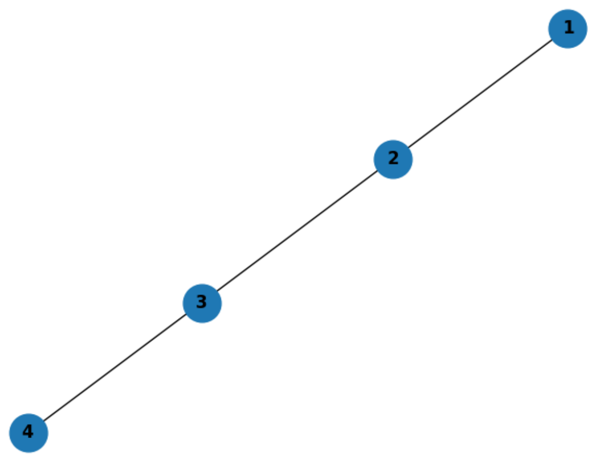
Node 2 to C2: $\sqrt{(1-0)^2+(0-0)^2+(1-1)^2+(0-0)^2}=1$

Node 3 to C1: $\sqrt{(0-0)^2+(1-1)^2+(0-0)^2+(1-0)^2}=1$

Node 3 to C2: $\sqrt{(0-0)^2+(1-0)^2+(0-1)^2+(1-0)^2} = \sqrt{3}$
Node 4 to C1: $\sqrt{(0-0)^2+(0-1)^2+(1-0)^2+(0-0)^2} = \sqrt{2}$
Node 4 to C2: $\sqrt{(0-0)^2+(0-0)^2+(1-1)^2+(0-0)^2} = 0$

Cluster C1: Node 1, Node 3
Cluster C2: Node 2, Node 4

b. What is the advantage of spectral clustering based on decomposition of adjacency matrices compared to applying k-means clustering directly to adjacency matrices as the question 7.a?
Ans: Spectral clustering provides a powerful alternative to traditional k-means, especially when dealing with complex, non-convex data structures. Its ability to capture the global structure of data makes it a valuable tool in the clustering.



8. (Link Prediction) Consider an undirected graph G of eight nodes given in the following figure. (10pt)

a. Calculate Jaccard's coefficient (JC), Adamic-Adar (AA) index of node 4 and node 6
Ans:

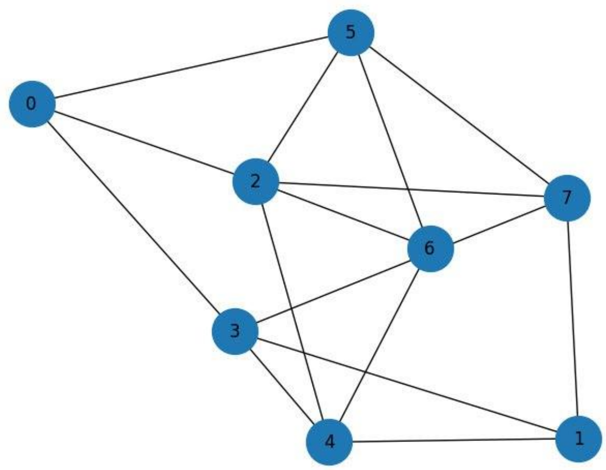
$$JC(4,6) = \frac{|\{2,3\}|}{|\{1,2,3,5,7\}|} = \frac{2}{4+5-2} = \frac{2}{7}$$

$$AA(4,6) = \sum_{u \in N(4) \cap N(6)} \frac{1}{\log\,|N(u)|} = \frac{1}{\log\,|N(2)|} + \frac{1}{\log\,|N(3)|} = \frac{1}{\log\,(5)} + \frac{1}{\log\,(4)}$$
$$= 1.71$$

b. JC and AA are commonly modifications of the common neighborhoods by applying weights to the number of common neighborhoods. Please explain their differences in terms of their methods for applying the weights and their underlying rationales.
Ans: JC treats all shared neighbors equally, focusing on the proportion of shared connections. In contrast, AA emphasizes the importance of less-connected common neighbors, providing a more nuanced similarity measure that can be particularly useful in networks where high-degree nodes are prevalent.

Equation JC: score (x, y) = $\frac{|N(x)\cap N(y)|}{|N(x)\cup N(y)|}$, where N(x), N(y) are neighbor nodes of node x, y respectively

Equation AA: score (x, y) = $\sum_{u\in N(x)\cap N(y)} \frac{1}{\log\ |N(u)|}$, with log(1) ≈ 0, log(2) ≈ 0.3, log(3) ≈ 0.4, log(4) ≈ 0.6, log(5) ≈ 0.7



9. (Link Prediction) Consider an undirected graph G of four nodes given in the following figure. (10pt)
   a. Calculate Katz Index with L = 2, β = 0.5

Equation: score (x, y) = $\sum_{l=1}^{L} \beta^l |paths_{xy}^{(l)}|$ = βA_{xy} + $\beta^2 A_{xy}^2$ + … + $\beta^L A_{xy}^L$, where $A^2 = A * A$, which A is adjacency matrix

Ans:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

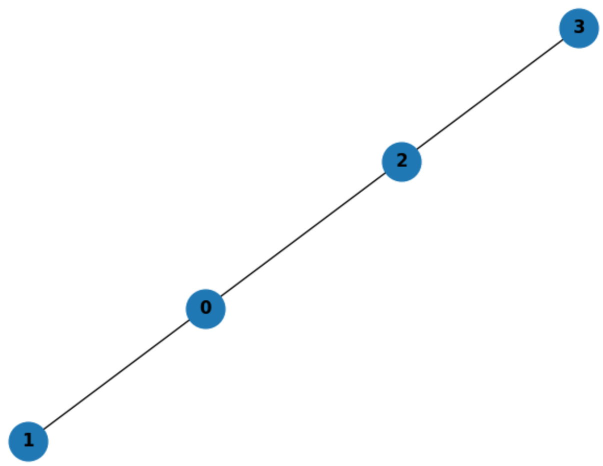
$$(x, y) = \sum_{l=1}^{L} \beta^l \left|paths_{xy}^{(l)}\right| = \beta A_{xy} + \beta^2 A_{xy}^2 = 0.5 \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} +$$

$$0.5^2 \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 \end{pmatrix} + 0.25 \begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 \end{pmatrix} + \begin{pmatrix} 0.5 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0.5 & 0 \\ 0.25 & 0 & 0 & 0.25 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0 \\ 0.5 & 0.25 & 0.5 & 0.5 \\ 0.25 & 0 & 0.5 & 0.25 \end{pmatrix}$$

b. Different from JC and AA, the Katz index is a global link prediction method. Why do we say it is global?

Ans: The Katz index is termed a global link prediction method because it considers all possible paths between two nodes in a network, not just their immediate neighbors. By summing over these paths and applying a damping factor to longer ones, it captures both direct and indirect relationships, reflecting the overall structure of the entire network



10. (Link Prediction) Consider an undirected graph G of three nodes given in the following figure. (10pt)

a. Calculate Hitting time of node 1 and node 2

Ans:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$P = AD^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} * \frac{1}{detD} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} * \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$H(1,2) = 1 + \sum_{m} \mathrm{p}_{mj} H(m,y) = 1$$

$$score(1,2) = -\mathrm{H}_{x,y} = -\frac{1}{|N(x)|}\sum_{k}(1 + \mathrm{H}_{k,y}) = -\frac{1}{1}(1 + \mathrm{H}_{2,2}) = -\frac{1}{1}(1 + 0) = -1$$

b. Please explain the common points and differences between the Hitting time and PageRank centrality. Also, what are the correlations between information propagation between nodes, transition probabilities, and the Hitting time.

Ans:

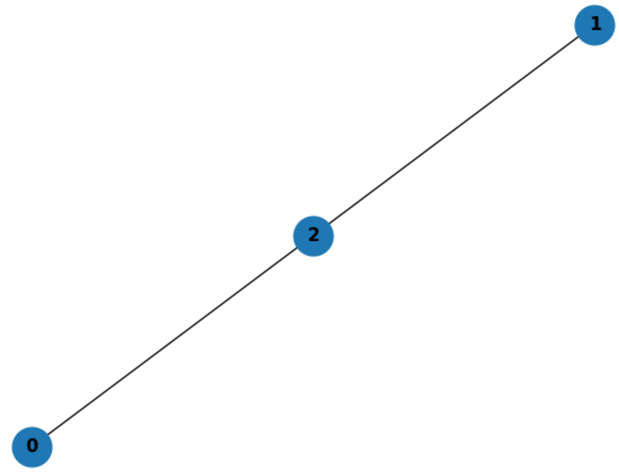
- Common:
  - Both concepts are based on the behavior of random walks
  - Utilize transition probability matrices to model the likelihood of moving from one node to another
  - Each provides insights into the structure and dynamics of networks, aiding in understanding node importance and connectivity
- Difference:

	Hitting Time	PageRank Centrality
Definition	Expected number of steps for a random walker to reach a specific node from a starting node.	Steady-state probability of a random walker being at a particular node.
Focus	Measure accessibility between specific pairs of nodes.	Assess overall importance or influence of nodes within the entire network.
Computation	Solves a system of equations based on transition probabilities.	Computes the principal eigenvector of the modified adjacency matrix.

- Relationship Between Information Propagation, Transition Probabilities, and Hitting Time:
  - Information Propagation: In network theory, information spreads through paths defined by the network's structure. The efficiency and speed of this propagation are influenced by the transition probabilities between nodes.
  - Transition Probabilities: These probabilities determine the likelihood of moving from one node to another in a random walk. Higher transition probabilities between nodes facilitate quicker information spread.
  - Hitting Time: This metric quantifies the expected time for information (modeled as a random walker) to reach a specific node from a starting point. Lower hitting times indicate faster information propagation to the target node

Equation Hitting time: score (x, y) = $-\mathrm{H}_{k,y} = -\frac{1}{|N(x)|}\sum_{k}(1 + \mathrm{H}_{k,y})$,

where H(k, y) = 1 + $\sum_{m} \mathrm{p}_{mj} H(m, y)$ when k≠y, otherwise H(k, y) = 0, p_{mj} is the element in the row m-th and column j-th of the matrix, P= AD^{-1},which P is a transition matrix, A is adjacency matrix and D is degree matrix.


1
2
0