Mid-term Exam (Graph Mining – Spring 2024)

Full Name:
Student ID:

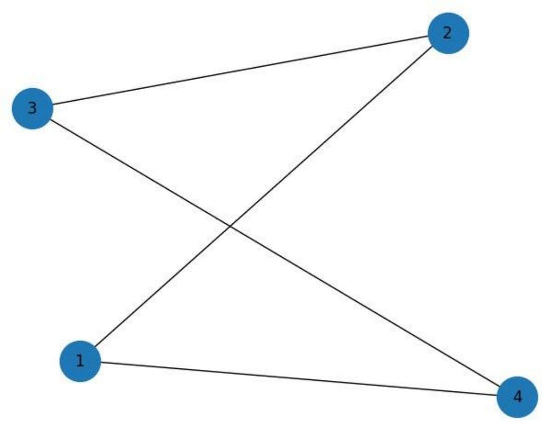
- The formula and solution process should be presented with the answer.
- Answers must be written in English.

1. Consider an undirected graph G of four nodes given in the following figure, calculate betweenness and closeness centrality of node 1 (5pt)

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

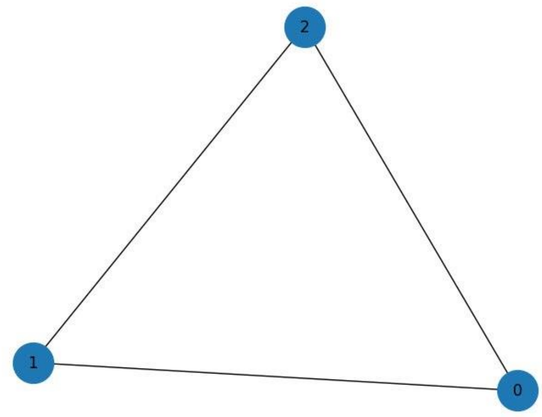
   Equation closeness centrality: $C(v_i) = \frac{N-1}{\sum_{j=1}^{N-1} d(v_j v_{i,})}$, where $d(v_j v_{i,})$ is number of nodes in the shortest path between node v_j and node v_i , and N-1 is the number of nodes reachable from v_i.



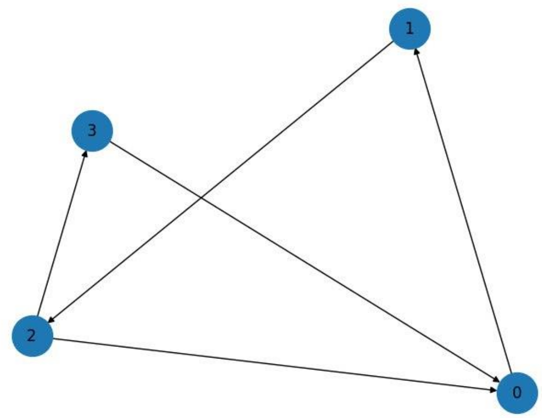
2. Calculate Eigenvector, Katz and PageRank centrality(10pt)
   a. Consider an undirected graph G of three nodes given in the following figure, calculate Eigenvector, Katz centrality of node 2 with $\alpha = 0.1$, $\beta = 1$, t = 1 (6pt)

      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 $x_j(0) = 1\ \forall j$

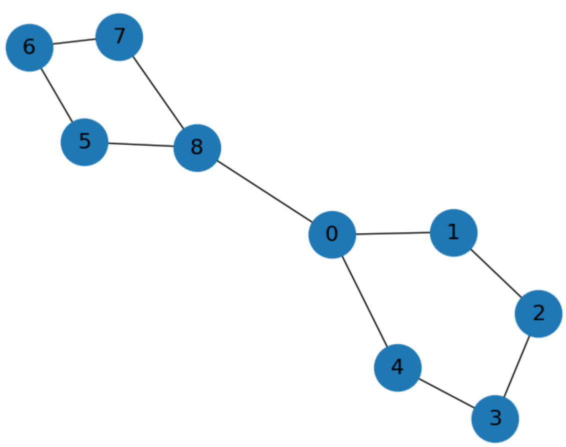
      Equation Katz: $Katz(G) = \beta(I - \alpha A^T)^{-1}.\mathbf{1}$, where α is damping factor, β is bias constant, I refers to the identity matrix, and **1** is a column vectors of ones. From Katz(G) results, write down the Katz centrality of node 2.



b. Consider a directed graph G of four nodes given in the following figure, calculate PageRank centrality of all nodes, with β = 0.85 (4pt)
Equation PageRank centrality of node i: $x_i = \sum_{(j,i)\in E} x_j + \beta$, where x_j is PageRank score of all pages j that point to page i



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



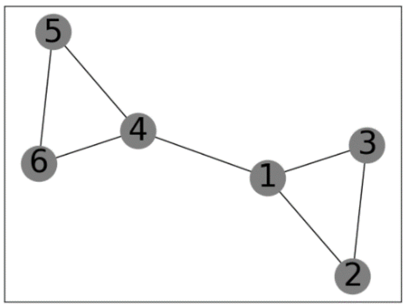
a. Calculate Min-cut and Normalized cut measurements of A and B.

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

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

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 cut between 2 communities A and B.

4. Consider an undirected graph G of six nodes given in the following figure. Apply the Equation (1) to calculate the clustering coefficient C_i of each node i and Equation (2) to calculate the average clustering coefficient $\langle C \rangle$ in the graph G. (10pt)

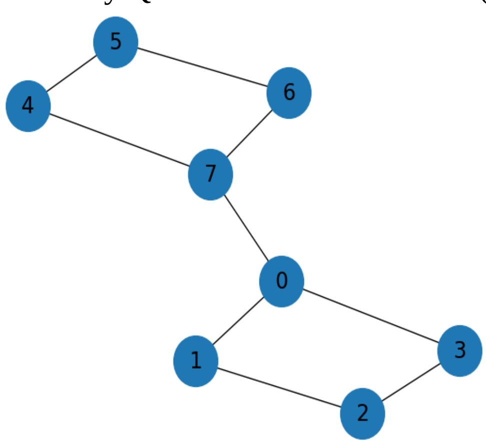


Equation (1): $C_i = \frac{2L_i}{d_i(d_i - 1)}$ (1)

where d_i is the degree of node i and L_i is number of edges between neighbors of node i.

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

5. Consider an undirected graph G of eight nodes given in the following figure with two communities: A = {0, 1, 2, 3} and B = {4, 5, 6, 7}. Apply the Equation (1) to calculate the modularity Q of the two communities. (10pt)



$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\cdot \delta(v_i, v_j) \quad (1)$$

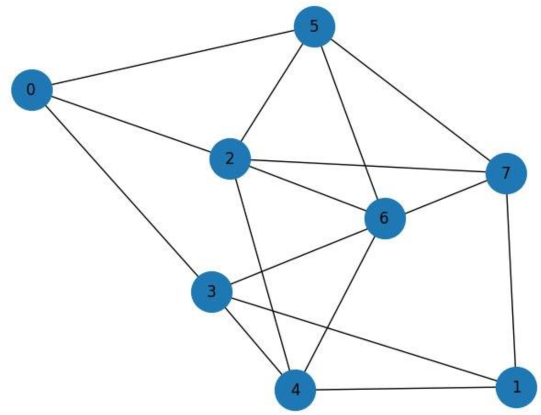
$$\delta(v_i, v_j) = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are in the same community} \\ 0 & \text{otherwise.} \end{cases}$$

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

6. Consider an undirected graph G of eight nodes given in the following figure, calculate Jaccard's coefficient (JC), Adamic-Adar (AA) index of node 2 and node 6 (10pt)

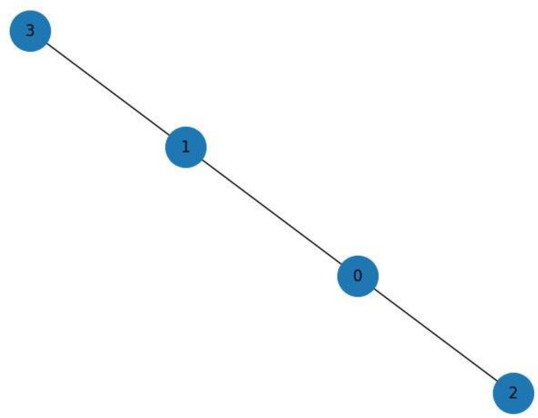
Equation JC: score (x, y) = $\frac{|N(x) \cap N(y)|}{|N(x) \cup N(y)|}$, where N(x), N(y) are neighbor nodes of node x, y respectively

Equation AA: score (x, y) = $\sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$, with log(4) ≈ 0.6



7. Consider an undirected graph G of four nodes given in the following figure, calculate Katz Index with L = 2, β = 0.5 (5pt)

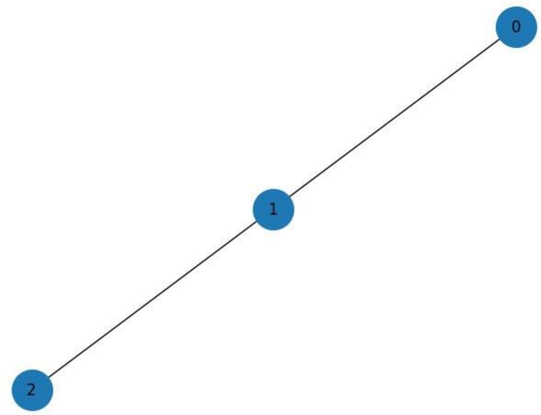
   Equation: score (x, y) = $\sum_{l=1}^{L} \beta^{l} |paths_{xy}^{(l)}|$ = $\beta A_{xy} + \beta^2 A_{xy}^2 + \ldots + \beta^L A_{xy}^L$, where $A^2 = A * A$, which A is adjacency matrix



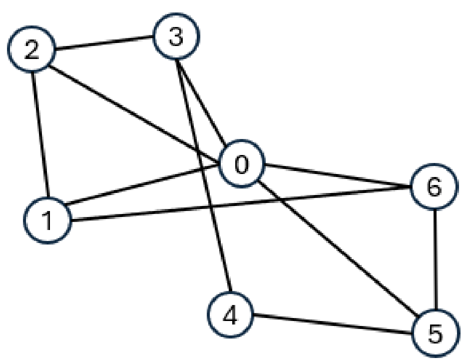
8. Consider an undirected graph G of three nodes given in the following figure, calculate Hitting time of node 1 and node 2 (5pt).

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

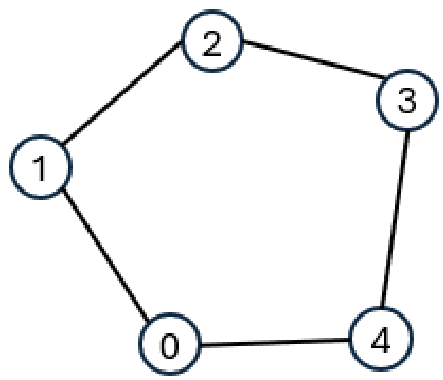
   where H(k, y) = 1 + $\sum_{m} \mathrm{p}_{mj} H(m, y)$ when k≠y, otherwise H(k, y) = 0, p_{mj} is the element in the row m-th and column j-th of the matrix, P= $\mathrm{A}D^{-1}$,which P is a transition matrix, A is adjacency matrix and D is degree matrix.



9. Consider two undirected graphs G_1 and G_2 below, calculate the graph edit distances from G_1 to G_2. The set of elementary operations: vertex insertion, vertex deletion, edge insertion, and edge deletion. In addition, the cost of insertion and deletion is 2 and 1, respectively. (5pt)

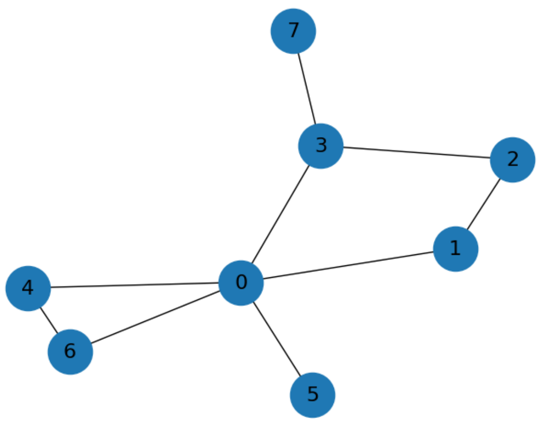


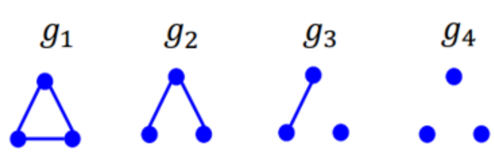
G_1



G_2

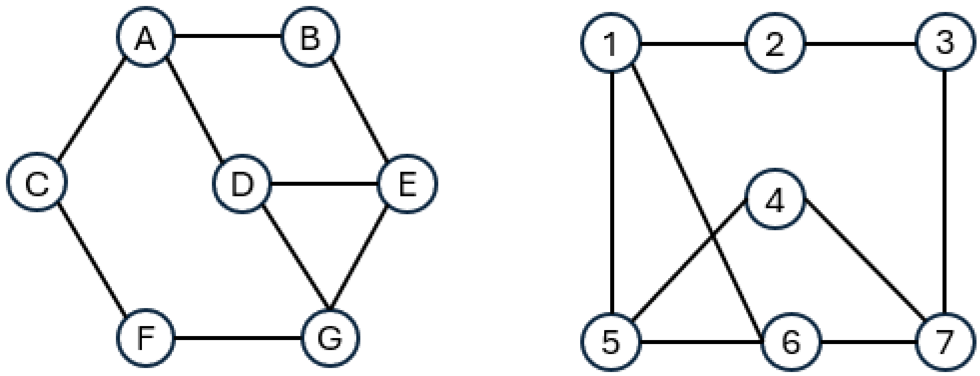
10. Consider an undirected graph G of eight nodes in the left side and four graphlets g_1, g_2, g_3, g_4 in the right side of following figure. Answer two question below: (10pt)





a. Count the number of the graphlets of size 3.
b. Make a feature vector for graph G based on the graphlet kernel with the 3-graphlets.

11. Consider two undirected graphs in the following figure: G_1 on the left and G_2 on the right. (10pt)
    a. Conduct Weisfeiler-Lehman (WL) relabeling process with the maximum degree 3.
    b. Make feature vectors for the graphs based on frequency of the WL subgraphs. Then calculate the similarity of graph G_1 and G_2 using Cosine Similarity equation (1).

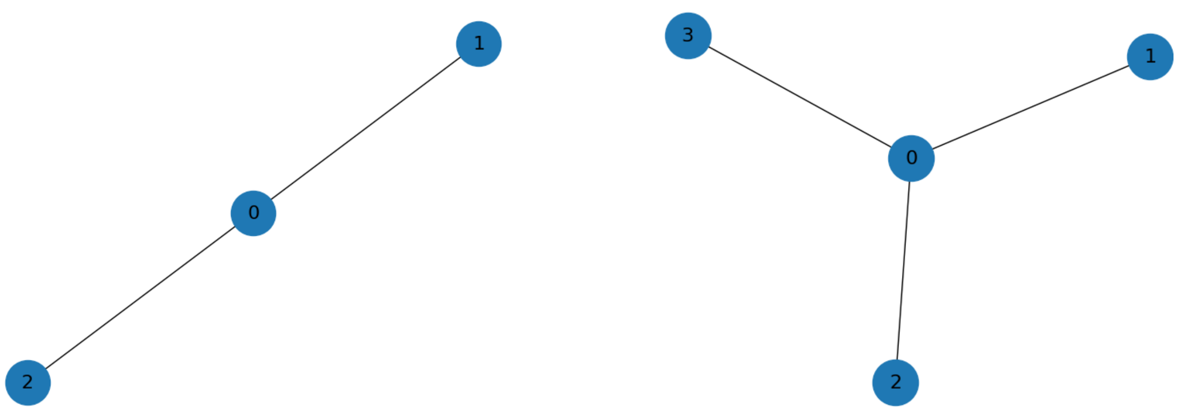


G_1 G_2

Cosine Similarity Equation (1): $cosine(WL_{G_1}, WL_{G_2}) = \frac{WL_{G_1}.WL_{G_2}}{\|WL_{G_1}\|\|WL_{G_2}\|}$.

where WL_{G_1} and WL_{G_2} is feature vectors of WL subgraph G_1 and G_2. "." denotes the dot product and "$\|\ \|$" denotes the Euclidean norm.

12. Consider an undirected graph G_1 and G_2 in the following figure. Make feature vectors of graphs G_1 and G_2 using the shortest path kernel and calculate similarity of graphs using the cosine similarity. (10pt)



G_1 G_2