working world

Major Exam
Social Network Analysis

Name: Aditya Sharma
Roll No: G24AI1098

1 (a) Convert the incidence Matrix to Adjancency Matrix

Given

$$\begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

Each colum represents an edge
Edge 1 = B/w node 1 and 2
Edge 2 = B/w node 2 and 3
Edge 3 = B/w node 2 and 4

So edges are (1,2) (2,3) (2,4)

∴ Adjancency Matrix =)

$$\begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

(b) Ans B is correct [Erdös-Rényi (Random Network) Model) because in this model, edges between each pair of nodes are created ~~indepen~~ independently with a fixed probability.

(c) Ans C is correct ⇒ Nash Equilibrium
In Nash Equilibrium, no player can benefit by changing strategies unilaterally.

(d) Ans B is correct ⇒ Assortative Mixing
This is the tendency to connect with similar others in network.

(e) Ans D is correct ⇒ Because it quantifies how often a node lies on the shortest paths between other nodes.

(f) Ans C is correct ⇒ The presence of many nodes with very high degrees (hubs) that maintain connectivity

(g) Ans A is correct ⇒ The number of intra-community edges is significantly higher than expected in a random network with the same degree sequence.

(h) Ans is **B** is Correct ⇒ 2/5

X : {A, B, C, D}

Y : {C, D, E}

intersection = {C, D} = 2

Union = {A, B, C, D, E} = 5

∴ ⇒ 2/5

(i) Ans **A** is correct ⇒ ICM uses edge probabilities independently ; LTM uses a weighted sum of active neighbours compared to a node ~~threshold~~ threshold This capture a core difference in activation logic

(j) Ans **B** is correct ⇒ Because aggregating features from dissimilar neighbors can blur the node's own representative feature, making classification harder.

2. Our Aim is to Vaccinate 5% of the population to minimize spread.

Strategy we can follow:

Using Betweenness Centrality and degree centrality together

- Degree Centrality: Nodes with many direct connections spread the disease faster
- Betweenness centrality: Nodes that act as Bridge between different groups can propogate the disease across communities

Steps:

(i) Compute degree centrality for all nodes and select nodes with highest degree.

(ii) Compute Betweenness centrality and select the bridge node

(iii) Rank nodes using a weighted combination of both scores.

(iv) Select top 5% of the node ~~[illegible]~~ based on the ranking.

Justification
(i) high degree nodes infect many people directly
(ii) high betweenness nodes enables spread across communities
(iii) Combining both ensures vaccination of super spreaders and the bridges, effectively limiting both local and global spread.

3 Suggeted Collaborators Using Link Prediction + node embedding.

Approach:

(i) Graph construction: Build a graph using co-authorship and citation relationship.

(ii) Node Embedding:

a) Train Node2Vec on the graph to learn vector representation

b) Capture both homophily (similar fields) and structural roles.

(iii) Link Prediction

a) Use similarity (cosine / dot-product) between embeddings to predict potential collaboration

(b) Combine with traditional link prediction scores like Jaccard, Adamic-Adar.

Homophily: Reasearchers from similar field have similar neighbors and thus similar embeddings. This leads to more accurate collaboration prediction within fields.

Promote Cross - Discipilinary Collaboration
Diversify top suggestion by adding an "explore" component
→ Penalize overly similar field
→ Encourage connections with structurally similar but topically different nodes.
→ Use clustering or topic models to find diverse candidate.

4 (a) Three core idea of Girvan - Newman Algorithm
(i) Repeatedly remove edges with highest edge betweenness to break the graph into communities.
(ii) Betweenness identifies bridges between communities
(iii) As bridges are removed, the graph decomposes into densly ~~[illegible]~~ connected communities.

(b) Uses of Edge Betweeness

(i) After each edge removal, recompute edge Betweenness.

(ii) This iterative recomputation reflects the evolving structure of the network

(c) Major Computational Limitation

(i) High time complexity: Recomputing edges ~~best~~ betweenness is $O(nm)$ per step.

(ii) Not scalable to large network.

(d) Louvain Method.

(i) Greedy modularity optimization

Two phases.

a) Assign nodes to communities to locally maximize the modularity.

b) Collapse communities into super nodes and repeat.

(ii) Very fast, scalable and handles larger graph well

Ans 5 (a) The page rank algorithm originally ~~developed~~ developed by Google, is used to measure the importances of node (eg. Web pages) in a directed network.

Imagine a random ~~webpage~~ web surfer clicking on links indefinitely. The more often the surfer ends up on a page, the more important that page is

Let $PR(i)$ be the PageRank of node (i) and $L(j)$ be the number of outbounds links from node j then

$$PR(i) = \sum_{j \in In(i)} \frac{PR(j)}{L(j)}$$

(b) Damping factor d represents the probability (usually $d = 0.85$) that the random surfer follows a link. With probability $(1-d)$ the surfer ~~f~~ jumps to any random page.

Equation with Damping

$$PR(i) = \frac{1-d}{N} + d \sum_{j \in In(i)} \frac{PR(j)}{L(j)}$$

Where N : Total number of nodes

$\frac{1-d}{N}$: Represents random jumps.

(c) Nodes with no outgoing links are Dangling nodes in a web context; dead ends eg. (PDF pages
If a node has no outbound link, it can "absorb" rank. It break the assumption that PageRank is redistributed in across nodes.

Cause the PageRank vector to lose mass and not remain stochastic.

To handle this in PageRank

(i) Pretend that dangling nodes link to all nodes (including themselves)

(ii) Distribute ~~this~~ their rank uniformly accross the network

⇒ If node(j) is dangling.

$$\frac{PR(j)}{N} \text{ is added to all nodes.}$$

Which modifies PageRank Equation

$$PR(i) = \frac{1-d}{N} + d \sum_{j \in In(i)} \frac{PR(j)}{L(j)} + d \sum_{j \in D} \frac{PR(j)}{N}$$

Where D: set of Dangling nodes.

Ques 6 (a) A pure strategy Nash Equilibrium is a strategy profile (choice for each player) where

No player can increase their playoff by unilaterally changing their choice.

Let's analyse all 4 outcome.

(i) ⇒ (U, A)

where player 1's play off : 3 (Better than 2)

player 2's play off : 2 (Better than 0)

No one has incentive to deviate → Nash Equilibrium

(ii) ⇒ (U, B)

player 1 gets 0 → could switch to L → gets 2

would deviate

∴ Not a Nash Equilibrium

(iii) (L, A)

Player 1 gets 2 switching to U gives 3

would deviate –

∴ Not a Nash Equilibrium.

(iv) (L, B)

Player 1 gets 2 → switch to U gives 0 } No deviate

Player 2 gets 3 → switch to A gives 0 }

∴ Nash Equilibrium.

$\therefore$ Two pure strategy Nash Equilibria
(i) (U, A)
(ii) (L, B)

(b) Expected pay off for Player 2
Let p be the probability that player 1 plays U and (1-p) is for they Play L

Now if player 2 plays A
$E_2(A) = p \times 2 (U,A) + (1-p) \times 0 (L,A) = 2p$

If player 2 plays B
$E_2(B) = p \times 1 (U,B) + (1-p) \times 3 (L,B)$
$= p + 3(1-p) = p + 3 - 3p$
$\Rightarrow 3 - 2p$

hence
$E_2(A) = 2p$
$E_2(B) = 3 - 2p$

(c) Expected outcome when $p = 0.7$
For A = $E_2(A) = 2p \Rightarrow 2 \times 0.7 \Rightarrow 1.4$
For B = $E_2(B) = 3 - 2p = 3 - 2 \times 0.7 = 3 - 1.4 = 1.6$
Player 2 should chose strategy B as $1.6 > 1.4$

Ques 7. Given: A directed graph with edges.

$A \to B,\ C \to B,\ D \to B$.

So neighbors of $B = \{A, C, D\}$

→ Initial feature vectors.

$$h_A^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ h_C^{(0)} = \begin{bmatrix} 0 \\ 3 \end{bmatrix},\ h_D^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

→ Weight Matrix W:

$$W = \begin{bmatrix} 0.5 & 0 \\ 0.1 & 0.2 \end{bmatrix}$$

→ Activation function ReLU applied element-wise as $ReLU(x) = \max(0, x)$

Step 1 Aggregate Neighbor features.

We compute the average of the feature from B's neighbor

$$h_N^{(0)} = \frac{1}{|N(B)|} \sum_{u \in N(B)} h_u^{(0)}$$

Here $N(B) = \{A, C, D\}$

$$h_A^{(0)} + h_C^{(0)} + h_D^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{matrix} 3 \\ 6 \end{matrix}$$

$$\Rightarrow h_N^{(0)} = \frac{1}{3} \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Step 2 Linear Transformation

Apply the weighted matrix W to the aggregated vector

$$W h_N^{(0)} = W \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0.1 & 0.2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} (0.5)(1) + (0)(2) \\ (0.1)(1) + (0.2)(2) \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 \\ 0.1 + 0.4 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Step 3. Activation (ReLU)

$$\text{ReLU}\left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right) = \begin{bmatrix} \max(0, 0.5) \\ \max(0, 0.5) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Final Ans $\Rightarrow$ $h_B^{(1)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

This is the updated feature vector for node B