Major Exam

SNA

G24AI1121

1)

a) Given Matrix:

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

here each row represents a node (vertex) : A, B, C, D

each column represents an edge: (e_1, e_2, e_3)

e_1 connects nodes A, B

e_2 connects nodes B, C

e_3 connects nodes B, D

Adj matrix (4x4) is

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1 & 0 & 0 \\ B & 1 & 0 & 1 & 1 \\ C & 0 & 1 & 0 & 0 \\ D & 0 & 1 & 0 & 0 \end{array}$$

from incidence matrix we place 1's for each pair Simmetrically in the adj matrix to reflect the unidirected connections.

b) B → Erdös-Rényi (Random Network model)

* This model assumes that edges are formed between pairs of nodes with uniform probability and each edge is added independently of others

* other models involve preferential attachment, local rewiring or fixed degree sequences.

c) C → Nash Equilibrium

* It represents a stable state where all players are making the best decision they can, given the decision of others.

d) B → Assortative Mixing

* Refers to the tendency of individuals in a social network to associate and bond with others who have similarities to themselves like age, profession, beliefs, hobbies etc.,

e) D → Because it Quantifies how often a node lies on the shortest paths between other nodes.

* Betweenness Centrality measures the extent to which a node lies on the shortest-paths between other nodes which makes it highly relevant in networks where information flows alongside specific paths as nodes with high betweeness can control communication between different parts of the network.

f) C → The presence of many nodes with very high degrees (hubs) that maintain Connectivity.

* Scale free networks have a power-law degree distribution where a few nodes have very high degrees while most have few.

* These hubs maintain the overall connectivity of the networ while random failures are unlikely to hit these hubs. targetted attacks on them can fragment the network explaining both robustness and vulnerability.

g] A → The number of intra-community edges is significantly higher than expected in a random network with the same degree sequence.

- Modularity measures the strength of division of a network into communities.

- High modularity indicates that nodes within the same community are densely connected and there are fewer edges between communities than expected by chance.

- Hence optimizing Modularity helps to identify community structures by comparing to a random baseline.

h] B → $\frac{2}{7}$

Jaccard Co-efficient

Neighbours of X : {A, B, C, D}

Neighbours of Y : {C, D, E}

Common neighbours : {C, D} (2)

Unique neighbours : {A, B, C, D, E} (5)

∴ Jaccard co-efficient = $\frac{2}{5}$ B

j) A → ICM uses edge probabilities independently
LTM uses a weighted sum active neighbours compared to a node threshold.

- In ICM activation depends on edge-specific probabilities while in LTM activation is determined by comparing the weighted sum of active neighbours to a node specific threshold.

j) B → Because aggregating features from dissimilar neighbours can blur the node's own representative features making classification harder.

- In networks with high heterophily connected nodes have differing characteristics.

- Standard GCN's aggregate neighbour features without accounting for their dissimilarity, which may dilute a node's own distinct features.

2] Stratagy to vaccinate 5% of the population:

To minimize the spread of infection in an SIR model using only 5% Vaccination Coverage we can Combine 1) Modularity optimization

2) Betweenness Centrality

to select most impactful individuals for vaccination

Betweenness Centrality:

* measures how often a node lies on the shortest paths between other nodes in the network.

* A node with high betweenness Centrality has more control over the flow of information/spread of disease as it acts as a bridge between different parts of the network.

* By vaccinating individuals with high betweenness Centrality we can block the flow of infection between various groups or Communities within the network.

Why choose Betweenness Centrality:

- Vaccinating high-betweenness individuals can interrupt the transmission paths most effectively.
- These individuals are likely to connect clusters of individuals who would otherwise have limited contact So vaccinating them can prevent the disease from spreading across large population of the network.

Modularity optimization:

- Involves identifying clusters of nodes that are more densly connected with eachother than with the rest of the network.
- By identifying Communities, we can target the vaccination of individuals in critical communities that may have higher liklihood of experiencing rapid outbreaks due to dense connectivity.

Why choose Modularity optimization:

In networks with community structure, vaccinating individuals in densly connected communities can be especially effective in limiting outbreaks.

* If a community is more tightly connected an infection can spread quickly within it so preemptively vaccinating key individuals within those communities can prevent large scale spread.

Combining the two Concepts:

i) Identify the communities within the network using a modularity optimization algorithm.

ii) within each community calculate betweenness centrality to identify individuals who serve as key bridges between communities.

iii) Prioritize vaccinating the individuals who have high betweenness centrality and are situated in communities with high interconnectivity which ensures that resources are allocated effectively to block transmission both within & between communities.

Justification:

This combined approach addresses both the inter community transmission and intra community transmission ensuring the vaccination strategy is both effective & efficient given the limited resources.

(3) To enhance the suggested collaborator's feature we can combine "link prediction algorithms" with node embedding techniques trained on academic networks.

i) Node Embedding with "Node2Vec":

* Node2Vec learns low-dimensional representation of nodes by simulating biased random walks on the graph.

* These embeddings capture both structural roles and community affiliations.

* Train Node2Vec on co-authorship networks and citation networks.

(iii) Link Prediction:

- Use the learned embeddings to predict the liklihood of future collaboration:

→ Combine embeddings of two nodes via dot product or neural nets.

→ Rank potential collaborators by predicted scores.

- This allows suggesting pairs of researchers likely to collaborate in the future based on both network proximity & embedding similarities.

→ Role of Homophily:

- Homophily is the tendency of nodes to connect with similar others.

→ In research this means scholars often collaborate within the same field or institution.

- Node2Vec naturally captures homophily as nodes with similar neighbours are placed close in embedding space.

- This improves recommendation accuracy, ensuring relevence of suggestions.

→ Promoting Cross - Disciplinary collaboration:

* Introduce a controlled diversity factor
  → penalize overly similar embeddings
  → prioritize researchers with complementary expertise by using content features.

Eg.

→ use metadata to find non-overlapping yet related fields.

→ Boost recommendations where embeddings are moderately similar but come from different disciplines.

4) a) Girvan - Newman Algorithm for Community Detection:

* Identifies communities by iteratively removing edges that are central to the network.

* the idea is that communities within a network are seperated by edges with high betweenness centrality

* By sistematically removing edges with high betweenness centrality the network progressively splits into smaller communities.

→ This process continues until the network is divided into distinct communities with the assumption that these communities are internally more connected and externally sparse.

b)

→ Calculates the betweenness centrality of all edges

→ Removes the edge with the highest betweenness

→ Recalculates the betweenness centrality of the remaining edges after each removal

→ Repeats this process until the network splits into distinct communities.

c) A major computational limitation of this algorithm is its high time complexity.

→ It requires to recalculate edge betweenness centrality after each edge removal.

$O(mn)$ n → nodes
m → edges

d) * The Louvain method is a more scalable approach for communities detection specifically designed to optimize modularity.

Two phases:

i, Local Modularity Optimization:

* Each node is assigned to its own community

* for each node the algorithm evaluates whether moving the node to a neighbouring community would increase the overall modularity of the network and move the node if it does.

ii, Community Aggregation:

* The communities formed in the 1st phase are aggregated into super nodes, and the algorithm repeats the modularity optimization process on this new graph of communities.

* This process is iterated, with each iteration resulting a finer level of community structure.

Advantages:

* Provides a direct measure of community quality.
* highly efficient in scaling to large networks as the size of network is reduced in each step.
→ low time complexity.

5)

a) Page rank algorithm:

★ Based on the idea that the importance of a node is determined by the importance of the nodes that link to it.

[in this case nodes are webpages]

* The intuition is that a webpage is more important if it is linked by other important pages.

* Pager rank treats the network as a directed graph where nodes represent webpages and edges represent hyperlinks between them.

→ The algorithm assigns a score to each node based on two main principles.

* incoming links
* Link quality

* PageRank is computed iteratively with each node's rank being based on the ranks of the nodes linking to it.

b) Role of damping factor (d) in the Random Surfer model:

a damping factor (d) in pages rank algorithm models the behaviour of a random surfer navigating the web.

• The damping factor serves to:

i) Control Random Behaviour:

The parameter 'd' represents the probability that the random surfer will continue following links. (usually 0.85)

the term (1-d) represents the probability that the surfer will jump to a random page, ensuring that every node in the graph has a non-zero probability of being reached even if it's isolated or has no links incoming.

ii, Ensure Convergence: without the teleportation step the algorithm could get stuck in pages with no outgoing links or sink into cycles.

The damping factor prevents this, by allowing surfers to jump to any page in the graph.

c] Dangling nodes:

* nodes in the graph that have no outgoing links.

* In the context of page ranking algorithm this presents a problem as if a node has no outgoing links, it can't contribute to the rank of any other node.

* This results in a loss of rank distribution causing the algorithm to not converge properly or assign incorrect rank values.

Handling Dangling nodes in Page Rank:

i, Redistribute Rank: Instead of allowing the rank to leak out of dangling nodes their rank is redistributed across all other nodes.

iii) Teleportation:

The damping factor already includes a mechanism that reduces the impact of dangling nodes by giving a rank to all nodes via teleportation

b)
a) Pure Strategy Nash Equilibria:

Payoff matrix

	strategy A	strategy B
strategy U	(3,2)	(0,1)
strategy L	(2,0)	(2,3)

player(1) (row):

- if player 2 plays A
  U : 3   L : 2 → best response = U
- if player 2 plays B
  U : 0   L : 2 → best response = L

player 2 (column):

- if player 1 plays U
  A : 2   B : 1 → best response = A
- if player 1 plays L
  A : 0   B : 3 → best response = B

So, the Nash Equilibria are where both players playing best responses to each other

(U, A) → player-1 plays U player 2 plays A

(L, B) → player-1 plays L player 2 plays B

Neither players can improve their payoff by deviating unilaterally.

b) Let player 1 play

- U with probability p
- L with probability 1-p

Expected payoff of player-2

If player 2 plays A

$$\text{payoff} = p \cdot 2 + (1-p) \cdot 0 = 2p$$

If player 2 plays B

$$\text{payoff} = p \cdot 1 + (1-p) \cdot 3$$

$$= p + 3 - 3p$$

$$= 3 - 2p$$

c) if $P = 0.7$

player 2's payoff if playing A

$2 \cdot 0.7 = 1.4$

player 2's payoff if playing B:

$3 - 2 \cdot 0.7 = 3 - 1.4 = 1.6$

$\therefore$ $1.6 > 1.4$ player 2 Will choose strategy B.

Expected payoff:

player-1's :

$B = 0.7 \cdot 0 + 0.3 \cdot 2 = 0.6$

player-2's : 1.6

$\boxed{\Rightarrow (0.6, 1.6)}$

7) Given

$N(B) = A, C$ [neighbours of node B]

feature vectors

$h_A^{(0)} = [1, 1]$; $h_C^{(0)} = [0, 3]$ $h_D^{(0)} = [2, 2]$

weight matrix:

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

Aggregate neighbours features:

$$h^{(0)}_{N(B)} = \frac{1}{3}\left(h^{(0)}_A + h^{(0)}_C + h^{(0)}_D\right)$$

$$= \frac{1}{3}\left(\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}0\\3\end{bmatrix} + \begin{bmatrix}2\\2\end{bmatrix}\right)$$

$$= \frac{1}{3}\begin{bmatrix}3\\6\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}$$

Transform:

$$W \cdot h^{(0)}_{N(B)} = \begin{bmatrix}0.5 & 0\\0.1 & 0.2\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$$

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

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

Activate Relu

$$h^{(1)}_B = \sigma\left(\begin{vmatrix}0.5\\0.5\end{vmatrix}\right) = \begin{bmatrix}0.5\\0.5\end{bmatrix}$$

$$h^{(1)}_B = \begin{bmatrix}0.5\\0.5\end{bmatrix}$$