G24AI1109
ANMOL CHUGH

MAJOR EXAM

SOCIAL NETWORK ANALYSIS

DATE
PAGE

Q1) Incidence matrix
text

1 0 0
1 1 1
0 1 0
0 0 1

Now lets
nodes = 1, 2, 3, 4 – Assumption
Edges = E_1, E_2, E_3 – "

E_1 connects nodes 1 & 2
E_2 " " 2 & 3
E_3 " " 3 & 4.

	1	2	3	4
1	0	1	0	0
2	0	0	1	~~0~~
3	0	1	0	0
4	0	1	0	0

Q1b) B) Erdős – Rényi (Random Network) Model
Justification :- This model assumes each pair of nodes is connected with equal & independent probability.

Q1c) C) Nash Equilibrium
Justification :- At Nash Equilibrium, no player can gain by changing only their own strategy.

Q1) d) B) Assortative Mixing
Justification :- As it describes the preference of nodes to connect with similar nodes (e.g same degree).

Q1) e) D) Because it quantifies how often a node lies on the shortest paths between the nodes.
Justification :- It is because betweenness centrality is more useful in determining control over information flow through shortest paths.

Q1) f) c) The presence of many nodes with very high degrees (hubs) that maintain connectivity.

Justification : Hubs ensure network connectivity despite random failures but are vulnerable to targeted attacks.

Q1) g). A) The number of intra-community edges is significantly higher than expected in a random-network with same-degree sequence.

Justification : High modularity partitions have dense intra-Community connections & sparse inter-community connections.

Q1) h) B) 2/5

Justification :- ~~In both C & X, Y, are common~~, In both X, Y → C & D are common making them the intersection which makes intersection = 2 while union = 5.

∴ Coefficient = I / U

= 2/5.

Q1) i) A) ICM uses edge probabilities independently, LTM uses a weighted sum of active neighbors compared to a node threshold

Justification:- ICM - Each active node tries to activate neighbor independently with some probability

LTM :- A node activates only if the total influence from neighbors exceeds a threshold

6

Q1) i) B) Because aggregating features from dissimilar neighbors can blur the node's own representative features, making classification harder.

Justification :- In heterophilic networks, nodes often connect to dissimilar nodes. Standard GCNs assume homophily (nodes connected to similar nodes), so when aggregating neighbor features, the model might get "confused" due to irrelevant information from dissimilar neighbors.

Q2) We can vaccinate 5% of the population to minimize infections using two network analysis concepts.

1) Betweeness Centrality :-

So as per understanding it identifies nodes on the most shortest paths between others, acting as bridges for disease spread.

It's application can help prioritize nodes with highest scores, as vaccinating them disrupts transmission paths.

It can be explained as the SIR model disease spreads through contact. Nodes with high betweeness are critical for connecting different parts of network this will reduce the spread.

2) Degree centrality:-

Understanding - It measures the numbers of direct connections (contacts) a node has.

Application - After high betweenness we focus on high degree centrality to maximize the number of direct contacts.

It can be justified as high degree nodes can be heavy spreaders in the SIR model. Therefore vaccinating them can reduce initial spread.

Q3) Combine Link Prediction.

Approach.

* Note Embeddings (Node2Vec):

•) Methodology:- Train Node2Vec on a co-authorship/citation network. Nodes are researchers, edges are co-authorships or citations. Node2Vec uses random walks to generate low-dimensional embeddings capturing structural & neighborhood similarities

* Link Prediction Algorithm:

•) Methodology:- Use metrics like Jacard Coefficient on the co-authorship to predict potential collaborations

Jacard Coefficient

$$= \left| \frac{\text{Neighbors}(i) \cap \text{Neighbors}(j)}{\text{Neighbors}(i) \cup \text{Neighbors}(j)} \right|$$

Now we combine them as follows:-

1) Use Node 2 Vec to filter researchers with similiar embedding.

2) Apply Link prediction on filtered set to rank collaborators based on network structure.

3) Now recommendations of researchers are both topically similiar (embeddings) & likely to collaborate (link prediction)

ROLE OF HOMOPHILY

•) Researchers tend to collaborate within same fields:

•) Node2Vec embeddings reflect homophily due to co- authorship / citation patterns. This ensures recommendations align with existing research intrests.

•) However, dependency on homophily may limit cross-disciplinary collaborations.

Promoting Cross-Disciplinary Collaboration

•) Introduction of a diverse penalty in rankings.

•) Occassionally recommend researchers with high link prediction scores but moderate embedding dissimiliarity, encouraging collaboration

•) This will balance homophily-driven recommendations

Q4) a) Core idea of Girvan–Newman Algorithm:

•) It detects communities by iteratively removing edges with the highest edge betweeness centrality, splitting the network into disconnected components (communities).

•) Edges with high betweeness connect different communities as they lie on many short paths

b). Use of Edge between Centrality :-

Steps:-

1) Compute edge's betweeness for all edges.

2) Remove the edge with highest betweeness.

3) Recalculate betweeness for remaining edges

4) Repeat until network splits into desired communities

c) Major computational Limitation:

- Computing edge betweenness is computationally expensive requiring shortest-path calculations for all node pairs.

- It is complex as it is repeated multiple times making it unfeasible for large graphs.

- Also scales poorly for networks with millions of nodes/edges.

d) Louvain Method as a Scalable Alternative

- Optimizes modularity by iterative merging of nodes into communities to maximize intra-community edges.

Process:- 1) Start with each node as its own community.

2) For each node, try moving it to a neighbor's community if it increases modularity

3) Repeat until no modularity gain.

4) Aggregate community into supernodes & repeat on coarser graph

DATE
PAGE

Q5) a) Intuition behind PageRank:

•) It assigns importance to nodes in a directed graph (eg web pages) based on idea that important nodes are linked to other important nodes.

•) Due to which the intuition becomes Nodes with many incoming links from high-Page Rank nodes are more important.

b) Role of Damping factor d:

•) It is the probability that the surfer follows an outgoing link; 1 - d is the probability of jumping to a random node.

•) It ensures convergence by introducing randomness, preventing the surfer from getting stuck in loops.

•) Higher d emphasizes link structure; lower d makes scores more uniform.

c) Dangling Nodes Problems & Solution:

•) It has no outgoing links causing the random surfer to "stop"

•) The transition matrix becomes non-convergent to steady state.

Sol^n

- Assume dangling nodes link to all nodes uniformly.

- Add a term to PageRank vector update, redistributing the dangling nodes probability mass equally across all nodes.

- formula adjustment: $PR = d \cdot (A \cdot PR + D) + \frac{1-d}{N} \cdot 1$ where D accounts for dangling nodes.

Q6) a) A strategy pair where player can improve their pay off by unilaterally changing their strategy.

Check each pair:

(U, A): pay off (3, 2)

- player 1: Switch to L → 2 < 3, no improvement.
- player 2: Switch to B → 1 < 2, "
- Nash Equilibrium.

(U, B): Payoff (0, 1)

- Player 1: Switch to L → 2 > 0, improvement.
- Not Nash.

(L, B): Payoff (2, 3).

- Player 1: Switch to U → 0 < 2, no improvement.

- player 2 : Switch to A → 0 < 3,
no improvement
Nash Equilibrium

There are two pure strategy Nash Equilibria
(U, A) & (L, B)

Q6) b) •) Player 1 plays U with probability P, L with probability 1-p

•) Player 2's strategy

Expected payoff for player 2 is choosing strategy A :

$$E[A] : p \times 2 + [(1-p) \times 0] = 2p$$

Expected payoff for player 2 if choosing strategy B :

$$E[B] = p \times 1 + (1-p) \times 3$$
$$= p + 3 - 3p$$
$$= 3 - 2p$$

Q6) c) For player 2 :

$$E(A) = 2p = 2 \times 0.7 = 1.4$$
$$E(B) = 3 - 2p = 3 - 2 \times 0.7 = 1.6$$

Since $E[B] > E[A]$, player 2 would choose strategy B for player 1 (0.7 0, 0.32)

Expected payoff = $0.7 \times 0 + 0.3 \times 2$
$= 0.6$ (When player 2 choose Strategy B)

The expected outcome for Player 1 receives payoff of 0.6 & player 2 receives a payoff of 1.6.

Q7) Given.
Directed Edges $A \to B$
$C \to B$
$D \to B$.

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

Step 1 : Aggregate neighbor.

$h_A^{(0)} + h^{C(0)} + h^{D(0)} =$

$\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

$\begin{bmatrix} 3 \\ 6 \end{bmatrix}$

$h_{N(B)}^{(0)} = \frac{1}{3}\begin{bmatrix} 3 \\ 6 \end{bmatrix} = \frac{1}{2}$

Step 2

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

$$= \begin{bmatrix}0.5\\ 0.1\times 1+0.2\times 2\end{bmatrix}$$

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

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

Step 3 :

$$\mathrm{ReLU}([0.5, 0.5]) = [0.5, 0.5]$$

Final Answer $h_B^{(1)} = [0.5, 0.5]$