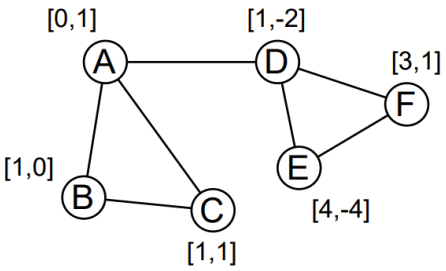
Final Exam (Graph Neural Networks –Fall 2024)

Full Name:
Student ID:

1. (10p) Consider an undirected graph G of six nodes A, B, C, D, E, and F given in the following figure. Each node has initial features that are the numbers standing next to it.



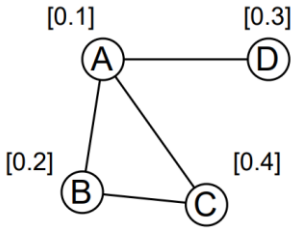
Assume that the hidden layer of an GCN model of all nodes at layer k can be calculated as:

$$H^{(k)} = \sigma\left(A \cdot H^{(k-1)}\right),$$

where $H^{(k)}$ denotes the output at layer k, σ is a ReLU function $\mathrm{ReLU}(x) = \max(0, x)$.
A is the adjacency matrix.

   a) Calculate the output of the GCN model at layer $k = 1$.
   b) What are the limitations of the GCN model compared to GraphSAGE?

2. (10p) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. Each node has an initial feature that is the number standing next to it (i.e., the initial feature of node 'A' is $h_A^{(0)} = 0.1$). According to GraphSAGE model with a MAX aggregation function, the feature of a node i at layer k can be updated as:

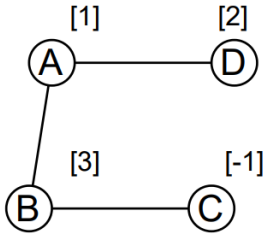


$$h_{N(i)}^{(k)} = \mathrm{AGGREGATE}\left(\left\{h_u^{(k-1)}, \forall u \in N(i)\right\}\right)$$
$$h_i^{(k)} = \mathrm{ReLU}\left(h_i^{(k-1)} || h_{N(i)}^{(k)}\right)$$

where || is a concatenation, $\mathrm{ReLU}(x) = \max(0, x)$,
N(i) is the neighbour nodes of node *i*.

   a) Calculate the feature of each node at $k = 1$.
   b) How does the GraphSage model address the scalability problem?

3. (10p) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. Each node has an initial feature that is the number standing next to it, as follows:



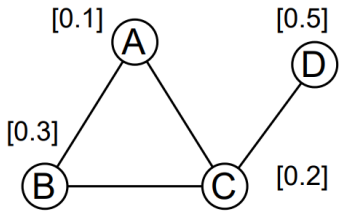
The output of an GCNII model of all nodes at layer (k) can be calculated as:

$$H^{(k)} = \sigma\left[\left((1-\beta)I_n\right) \cdot \left((1-\alpha)\tilde{A} \cdot H^{(k-1)} + \alpha H^{(0)}\right)\right]$$

where $H^{(k)}$ denotes the output at layer k, $\tilde{A}$ is the normalized matrix ($\tilde{A} = D^{-1}A$), I_n is the identity matrix, $\alpha = \beta = 0.5$, $\sigma(.)$ is a ReLU function $\mathrm{ReLU}(x) = \max(0, x)$.

   a) Calculate the representations of each node at layer $k = 1$.
   b) What are the reasons why the over-smoothing problem happens in graphs? And how does GCNII address the problem?

4. (10p) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. Each node has an initial feature that is the number standing next to it. According to GAT model, the weight matrix W is initialized as [0.1]. The feature of node 'i' at layer (k) can be updated as:



$$h_i^{(k)} = \sigma\left(\sum_{m \in N(i)} \alpha_{im} W h_m\right)$$

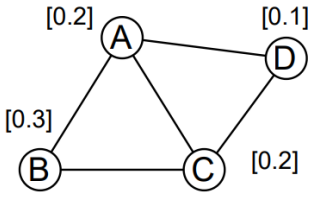
where: $\alpha_{im} = \frac{e_{im}}{\sum_{k \in N(i)} e_{ik}}$

and $e_{im} = \sigma\big(\text{MEAN}(Wh_i, Wh_m)\big)$

σ is a ReLU function $\text{ReLU}(x) = \max(0, x)$.

a) Calculate the feature of node 'A' at $k = 1$.
b) How can the attention mechanism address the node importance problem?

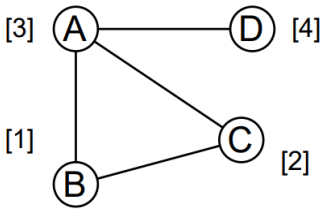
5. (10p) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. Each node has an initial feature that is the number standing next to it. According to GIN model, the parameter is a fixed scalar $\varepsilon = 0.5$, the feature of a node i at layer k can be updated as:



$$h_i^{(k)} = (1 + \varepsilon) \cdot h_i^{(k-1)} + \sum_{j \in N(i)} h_j^{(k-1)}$$

a) Calculate the feature of each node at $k = 1$.
b) What are the reasons why GIN model achieves 1-d Weisfeiler-Lehman?

6. (10p) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. According to GNN-AK model with a GCN model and MEAN aggregation function, the feature of a node v at k-th layer can be updated as:



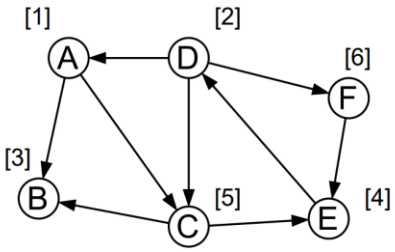
$$h_v^{(l+1)} = AGG^{(l)}(G[N_k(v)]), l = 0,1,\ldots,L-1$$

$$h_v = MEAN\left(h_u^{(L)} | u \in G[N_k(v)]\right)$$

where $G\left[N_k(v)\right]$ is a subgraph rooted at node v with k-hop distance, including v; and AGG(.) is a SUM function.

a) Calculate the feature of each node at $k = 1$ and $l = 0$.
b) Why considering a set of k-hop subgraphs will make the model more expressive than only neighbor aggregation?

7. (10p) Given a directed graph G of six nodes A, B, C, D, E, and F given in the following figure. Each node has an initial feature that is the number standing next to it.



   a) According to the Graphormer model, calculate the input node feature with its in-degree and out-degree. The definition of a Lookup table for node degree is as: T = {"0": 0.0, "1": 0.1, “2”: 0.2, “3”: 0.3, “4”: 0.4}.
   b) Why can injecting the structural information into graphs make the model more powerful in capturing the graph structures? What kind of structural information needs to be considered?

8. (10p) What are the main reasons why BPR loss (Bayesian Personalized Ranking loss) is commonly utilized in recommendation system models?

9. (10p) Why the TransR loss is critical in knowledge graph embedding models compared to the TransE loss?

10. (10p) What are the limitations of contrastive learning methods in molecular structure learning? And why are the subgraph-based methods better at molecular structure learning?