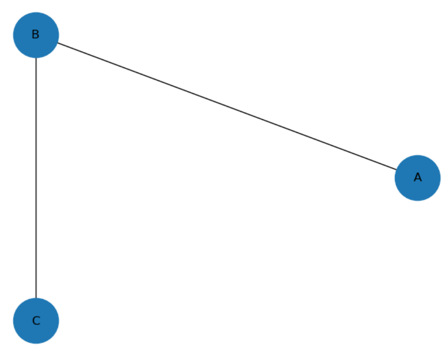
Mid-term Exam (Graph Neural Networks –Fall 2025)

Full Name:
Student ID:
Note: Students should write in English for evaluation.

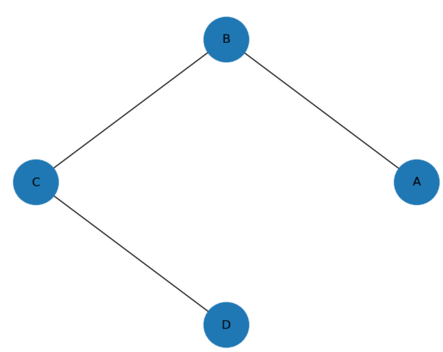
1. (Node2vec)
   a. Explain BFS, DFS.
   b. Consider an undirected graph G, with p = 2, q = 1. List all possible next nodes form node B and calculate unnormalized transition probabilities $\pi_{BAx} = \alpha_{pq}(A, x)$, where

$$\alpha_{pq}(A,x) = \begin{cases} \frac{1}{p}, if\ d(A,x) = 0 \\ 1, if\ d(A,x) = 1 \\ \frac{1}{q}, if\ d(A,x) = 2 \end{cases}$$



2. (LINE)
   a. What is second-order proximity?
   b. Consider an undirected graph G, with edge weights $w_{AB} = 2, w_{BC} = 1, w_{CD} = 3$, and node embedding $u_A = 1, u_B = 0.5, u_C = -0.5, u_D = -1$ and context node $u'_A = 0.5, u'_B = 1, u'_C = -1, u'_D = -0.5$. Calculate $p_2(A|B), p_2(B|B), p_2(C|B), p_2(D|B)$. Given $e^{0.25} = 1.28, e^{0.5} = 1.65, e^{-0.25} = 0.78, e^{-0.5} = 0.61$
   Second order:

$$p_2(j|i) = \frac{e^{u'_j u_i}}{\sum_k e^{u'_k u_i}}$$



3. (GCN)
   a. Explain how information is propagated in a GCN layer.
   b. Given a graph with an adjacency matrix A and initial node feature matrix $H^{(0)}$ as follows:

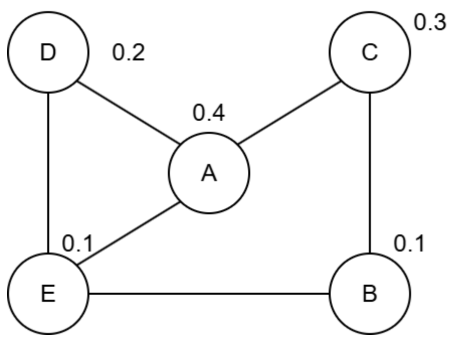
$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad H^{(0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 2 \\ 2 & 2 \\ 2 & 0 \end{bmatrix}$$

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 $\mathrm{ReLU}(x) = \max(0, x)$. Calculate the output of the GCN model at layer $k = 1$.

4. (GraphSAGE) Consider an undirected graph G of five nodes A, B, C, D, and E given in the following figure. Each node has initial features that are the numbers standing next to it (i.e., the initial feature of node 'A' is $h_A^{(0)} = 0.4$). According to GraphSAGE model with an AGGREGATE is a MEAN function, the feature of a node i at layer k can be updated as:

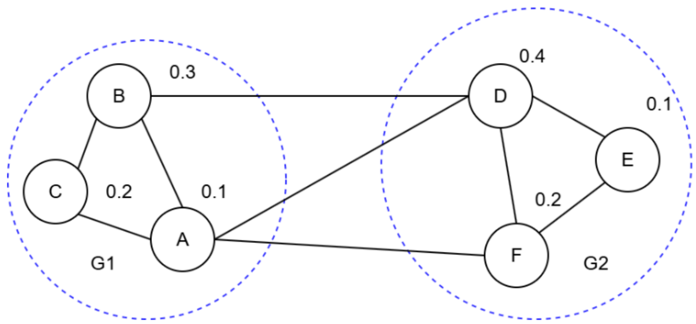


$$h_{N(i)}^{(k)} = \mathrm{AGGREGATE}\left(\left\{h_u^{(k-1)}, \forall u \in N(i)\right\}\right)$$
$$h_i^{(k)} = \mathrm{ReLU}\left(h_i^{(k-1)} || h_{N(i)}^{(k)}\right)$$

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

   a) What is the main idea behind the GraphSAGE model, and how does it differ from traditional GCNs?
   b) Calculate the feature of each node at $k = 1$.
   c) Calculate a graph-level embedding h_G by using a 'Mean' global pooling when $k = 1$.

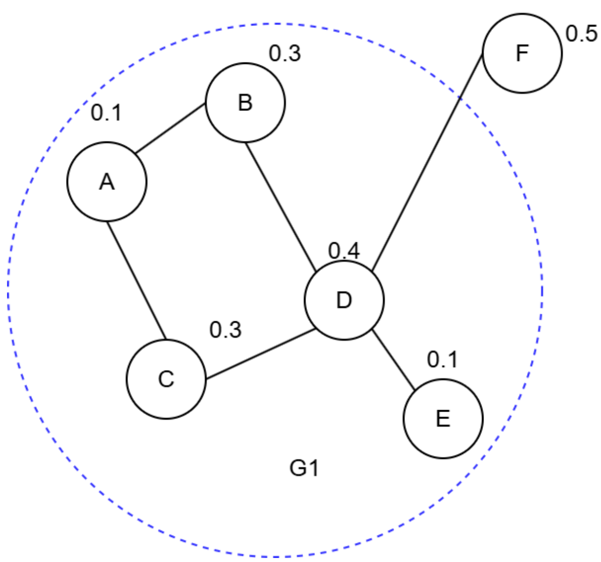
5. (ClusterGCN) Consider an undirected graph G of six nodes A, B, C, D, E and F given in the following figure. The graph G contains two cluster G_1 and G_2. Each node has initial features that are the numbers standing next to it. According to ClusterGCN model, the feature of a node i at layer k can be updated as:



$$h_{N(i)}^{(k)} = \text{MEAN}\left(\{h_u^{(k-1)}, \forall u \in N(i), G_u = G_i\}\right)$$
$$h_i^{(k)} = \text{ReLU}\left(h_i^{(k-1)} || h_{N(i)}^{(k)}\right)$$
where || is a concatenation.

a. What is mini-batch training?
b. Calculate the output representations of all nodes at layer $k = 1$.

6. (GraphSAINT) Consider an undirected graph G of six nodes A, B, C, D, E and F given in the following figure. The graph G has subgraph sampling G_1. Each node has initial features that are the numbers standing next to it. According to GraphSAINT model, the feature of a node i at layer k can be updated as:



$$h_{N(i)}^{(k)} = \text{MEAN}\left(\{h_u^{(k-1)}, \forall u \in N(i), G_u = G_i\}\right)$$
$$h_i^{(k)} = \text{ReLU}\left(h_i^{(k-1)} || h_{N(i)}^{(k)}\right)$$
where || is a concatenation.

a. How does GraphSAINT's sampling strategy differ from node-wise or edge-wise sampling methods like GraphSAGE?
b. Calculate the output representations of nodes A, B, C, D, and E at layer $k = 1$.

7. (JK Network) Given a graph with an adjacency matrix A and initial node feature matrix $H^{(0)}$ as follows:

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \qquad H^{(0)} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 1 & 2 \\ -1 & 1 \\ 2 & 3 \end{bmatrix}$$

Assume that the output of an JK network model of all nodes at layer (k) can be calculated as:

$$H^{(k)} = \max\left(\sigma\left(\tilde{A} \cdot H^{(0)}\right), \sigma\left(\tilde{A} \cdot H^{(1)}\right), \ldots, \sigma\left(\tilde{A} \cdot H^{(k-1)}\right)\right)$$

where $H^{(k)}$ denotes the output at layer $k, \tilde{A}$ is the normalized matrix ($\tilde{A} = D^{-1}A$), σ is a ReLU function $\text{ReLU}(x) = \max(0, x)$.

a) Explain how Jumping Knowledge Networks help mitigate the **over-smoothing problem** in deep GCNs.

b) Calculate $\tilde{A}$.

c) Calculate the output representations at layer $k = 2$.

s

8. (GCNII) Given a graph with an adjacency matrix A and initial node feature matrix $H^{(0)}$ as follows:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad H^{(0)} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

Assume that 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, $\alpha = \beta = 0.5$, σ is a ReLU function $\text{ReLU}(x) = \max(0, x)$.

a) What are the two major components introduced in GCNII to overcome the depth limitation problem in GCNs?

b) Calculate $\tilde{A}$.

c) Calculate the output representations at layer $k = 1$.

9. (DeepGCN) Given a graph with an adjacency matrix A and initial node feature matrix $H^{(0)}$ as follows:

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad H^{(0)} = \begin{bmatrix} 0 & -1 \\ -1 & 3 \\ 2 & -1 \\ 0 & -3 \end{bmatrix}$$

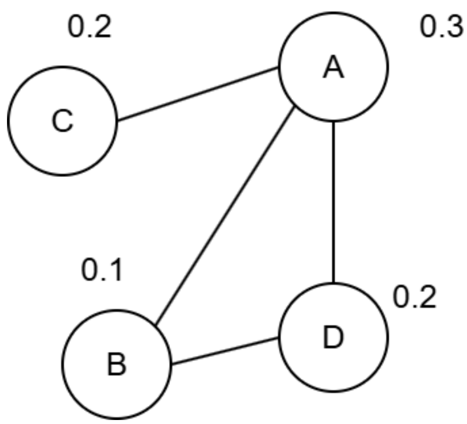
Assume that the hidden layer of an DeepGCNs model of all nodes at layer (k) can be calculated as:

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

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

a. How does DeepGCN mitigate the **over-smoothing problem** as depth increases?
b. Calculate the output of the GCN model at layer $k = 2$.

10. (GAT) Consider an undirected graph G of four nodes A, B, C, and D given in the following figure. Each node has initial features that are the numbers standing next to it (i.e., the initial feature of node 'A' is $h_A^{(0)} = 0.3$). According to GAT model, the weight matrix W is randomly initialized as [0.5]. 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

$$e_{im} = \sigma\left(\text{MEAN}(Wh_i, Wh_m)\right)$$

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

a) What problem in traditional GCNs does GAT aim to solve?
b) Calculate the attention coefficients e_{AB}, e_{AC}, and e_{AD}
c) Calculate the feature of node 'A' at $k = 1$.

11. (GATv2)
    a. What is the key limitation of GAT does GATv2 aim to fix?
    b. Calculate attention score e_{ij} to compare different between GAT and GATv2, given

GAT $e_{ij} = LeakyReLU(a\left(W\vec{h_i}, W\vec{h_j}\right))$

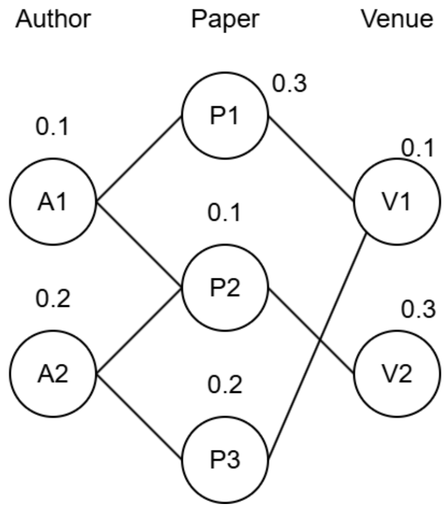
GATv2 $e_{ij} = aLeakyReLU(W[\vec{h_i}, \vec{h_j}])$

$$\vec{h_i} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \vec{h_j} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, a = [1,1,1,1]$$

LeakyReLU activation, where $\begin{cases} x = x, if\ x > 0 \\ x = 0.02x, if\ x < 0 \end{cases}$

12. (HAN) Consider a heterogeneous graph given in the following figure. There are three types of nodes in the academic network: Author (A), Paper (P), and Venue (V). Each node has initial features that are the numbers standing next to it (i.e., the initial feature of node 'A_1' is

$h_{A_1}^{(0)} = 0.2$). According to HAN model, the weight matrix W is randomly initialized as [0.5]. The feature of node 'i' at layer (k) can be updated as:



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

Where $\alpha_{im}^{\Phi} = \frac{e_{im}^{\Phi}}{\sum_{k \in N^{\Phi}(i)} e_{ik}^{\Phi}}$ and

$$e_{im}^{\Phi} = \sigma\left(\mathrm{MEAN}\left(W h_i^{\Phi}, W h_m^{\Phi}\right)\right)$$

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

a) What is heterogeneous graph?
b) List all the meta-path PAP and PVP. Calculate the attention coefficients of each meta-path PAP and PVP.
c) Calculate the feature of node 'P_1' at $k = 1$.