Midterm Preparation Questions

EPFL EE-452 — Network Machine Learning — Spring 2024/2025

Instructions:

- This set of questions is not exhaustive. It serves to illustrate the style and difficulty level of questions that may appear in the midterm exam.
- The midterm will cover all material presented in lectures up to the date of the exam.
- The midterm is open-book: you may bring any printed or handwritten resources. Electronic devices are not permitted.
- The midterm exam is composed of two parts: a set of multiple-choice questions and a set of open questions.
- Multiple-choice questions should be answered with only one option. Selecting more than one option leads to 0 credit for that question.
- In the Open Questions, clearly justify your answers to receive full credit.

Part A: Multiple Choice Questions

A.1: Given the graph in Figure 1, what is the global clustering coefficient of the graph?

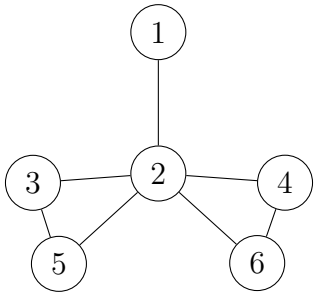

Figure 1: The numbers written in the nodes are the node IDs, not features.

- ☐ 2/11.
- ☐ 6/11.
- ☐ 1/7.
- ☐ 3/7.

A.2: Let χ_{10} be the eigenvector of the largest eigenvalue λ_{10} of the combinatorial Laplacian of the graph in Figure 2. Which of the following statements about the entries corresponding to nodes 9 and 10 is always true?

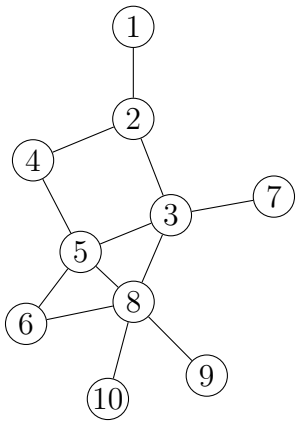

Figure 2: An unweighted and undirected graph. Numbers represent node IDs.

- ☐ They are both positive.
- ☐ They are both negative.
- ☐ They have opposite signs.
- ☐ They are equal.

A.3: Which of the following statements is true for unsupervised embedding extraction on graphs?

- ☐ The autoencoder framework offers great flexibility by enabling the unconstrained learning of node embeddings.

- ☐ The autoencoder framework stabilizes the underdetermined reconstruction of node embeddings, for example by regularizing them with a Laplacian eigenmaps component.
- ☐ Max pooling prevents capturing the very distinctive features of node embeddings in order to get graph embeddings robust to outliers.
- ☐ None of the above.

A.4: Which of the following statements about vanilla Graph Convolutional Networks (GCNs) is correct?

- ☐ GCNs require all input graphs to have exactly the same number of nodes.
- ☐ GCNs are sensitive to node ordering.
- ☐ GCNs do not handle node features as they only process structural information.
- ☐ GCNs inherently suffer from vanishing gradients regardless of network depth.

A.5: Oversmoothing in GNNs refers to the phenomenon where increasing the number of layers leads to homogeneous node representations. Based on this definition, which of the following is a consequence of oversmoothing?

- ☐ Nodes fail to capture information from distant parts of the graph due to limited number of layers.
- ☐ Node embeddings of the final layer lose local specificity, becoming ineffective for distinguishing node categories.
- ☐ Incomplete neighbor aggregation leaves out key details, impairing prediction accuracy.
- ☐ Nodes cannot accurately predict outcomes due to insufficient aggregation of neighbor information.

Part B: Open Questions

B.1: We will now derive the basic properties and characteristics of a **Random Graph Model**. There are multiple ways of modeling random graph generation, but we will consider the one suggested by Gilbert [1]. The Gilbert Random Graph Model, also known as the $G(N, p)$ model, is a well known model in graph theory where a graph is constructed by connecting each pair among the N nodes independently with a fixed probability p.

1. What is the probability of having a graph with L edges?
2. What is the average node degree of the network?
3. What is the average clustering coefficient of the graph?
4. Comment on how the average clustering coefficient evolves in terms of the number of nodes and the average node degree.

References

[1] E N Gilbert. Random graphs. *Ann. Math. Stat.*, 30(4):1141–1144, December 1959.