# 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?

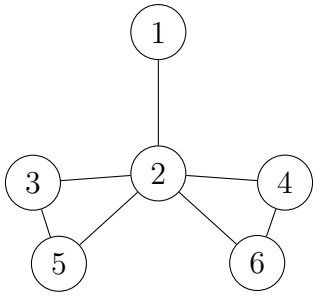


Figure 1: The numbers written in the nodes are the node IDs, not features.

- ☐ 2/11.
- ☐ 6/11.
- ☐ 1/7.
- ☐ 3/7.

**Solution:** Recall the formula for global clustering coefficient: $\frac{3\#triangles}{\#triplets}$. There are two triangles in the graph, so the dividend is $3 \times 2 = 6$. For the divisor, we can count the number of triplets which gives 14. Thus the final answer is 6/14.

**A.2:** Let $\chi_{10}$ be the eigenvector of the largest eigenvalue $\lambda_{10}$ 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?

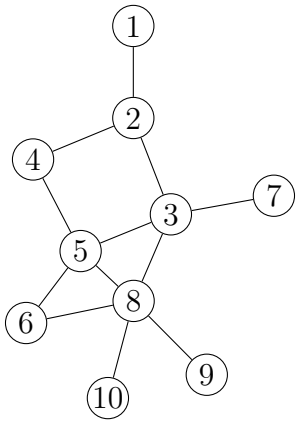


Figure 2: An unweighted and undirected graph. Numbers represent node IDs.

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

**Solution:** Let us denote $\chi_{10}[i]$ as the $i$-the entry of the eigenvector $\chi_{10}[i]$.

For a quick verification, we note that $L\chi_{10} = \lambda_{10}\chi_{10}$. By definition, $L = D - A$, with $D$ the diagonal matrix of node degrees and $A$ the adjacency matrix of the graph in Figure 2. Therefore, we have:

$$-\chi_{10}[8] + \chi_{10}[9] = \lambda_{10}\chi_{10}[9] \quad \text{and} \quad -\chi_{10}[8] + \chi_{10}[10] = \lambda_{10}\chi_{10}[10], \tag{1}$$

Therefore,

$$\chi_{10}[9] = \chi_{10}[10] = \frac{\chi_{10}[8]}{1 - \lambda_{10}}. \tag{2}$$

It just remains to confirm that $\lambda_{10} \neq 1$, which holds since $\lambda_{10} \geq \max_i \deg(v_i) = 5$, where $v_i$ denotes the $i$-th node of the graph.

*Proof of* $\lambda_{10} \geq \max_i \deg(v_i)$: Note that $\chi_{10}^\top L\chi = \chi_{10}^\top D\chi_{10} - \chi_{10}^\top A\chi_{10}$. Let us now consider the unit vectors $e_i$, which are identically zero except at the $i$-th entry, $e_i = [0, \ldots, 1, \ldots, 0]$. Note that for these $u_1^\top Du_1 = \deg(v_i)$and $u_1^\top Au_1 = 0$, where the second equality results from the fact that the diagonal entries of the adjacency matrix are 0. Thus, we have that $R(e_i) = \frac{e_i^\top Le_i}{\|e_i\|^2} = \deg(v_i)$, where $R$ denotes the Rayleigh quotient. Thus, $\lambda_{10} = \max_x\ x^\top Lx \geq \max_i \deg(v_i) = 5$.

*Intuition:* Note that the Rayleigh quotient definition of eigenvalues, $\chi_{10}$ is the (unit) vector that maximizes $\frac{\chi^\top L\chi}{\|\chi\|_2}$. Equivalently, it is the vector of highest variation over the graph. Since total variation is given by $\chi^\top L\chi = \sum_{i,j} ([\chi]_i - [\chi]_j)^2$, we see that both $[\chi]_9$ and $\chi_{10}[10]$ appear in the sum through their squared difference to entry of node 8, $\chi_{10}[8]$. It follows that they have to be equal in absolute difference to $\chi_{10}[8]$ to maximize variation. Thus, it could happen that $\chi_{10}[9]$ and $\chi_{10}[10]$ take two different values that would lead to the same absolute difference to $\chi_{10}[8]$. However, since we aim to minimize $\|\chi\|$ (as it appears in the denominator of the Rayleigh quotient), this ambiguity is resolved by selecting the value of smaller magnitude for both $\chi_{10}[9]$ and $\chi_{10}[10]$.

**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.

**Solution:**

- **False.** The reconstruction problem is underdetermined, so typically it requires enforcing some structure in the latent space.

- **True.**
- **False.** Max pooling allows to capture the very distinctive features, thus not robust against outliers.
- **False.** 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.

**Solution:**

☐ **False**. GCNs can deal with graphs with an arbitrary number of nodes.

☐ **False**. GCNs are, by design, permutation equivariant/invariant.

☐ **False**. GCNs can include all the node features in their message function.

☐ **True**. Vanilla GCNs indeed often face issues such as vanishing gradients or oversmoothing when depth increases

**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.

**Solution:** (A) describes underreaching, (C) and (D) generally describes a scenario of inadequate aggregation or ineffective message passing rather than oversmoothing explicitly. The correct answer is (B).

# 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.

**Solution:** Solution in several parts.

- The probability can be decomposed in three parts. First the probability of having created L links: $p^L$. The probability that no other link have been created: $(1-p)^{\frac{N(N-1)}{2}-L}$. The combinatorial factor to take into account the permutations: $\binom{\frac{N(N-1)}{2}}{L}$. Finally, the probability of seeing a graph with $L$ links is: $\binom{\frac{N(N-1)}{2}}{L}p^L(1-p)^{\frac{N(N-1)}{2}-L}$.
- The edge distribution follows a binomial distribution with probability $p$ and total number of trials being $\frac{N(N-1)}{2}$. Thus the average number of links in the graph is $E[L] = p\frac{N(N-1)}{2}$. The average degree of the network is then given by $k=\frac{2E[L]}{N} = p(N-1)$
- To calculate the clustering coefficient for a node $v$ we need to estimate the expected number of links $L_v$ between the node's neighbors. In a random network, the probability that two neighbors link one to another is $p$. As there are $\frac{N_v(N_v-1)}{2}$ possible links between the $N_v$ neighbors of node $v$, the expected value the number of links is: $E[L_v] = p\frac{N_v(N_v-1)}{2}$. Thus, the local clustering coefficient is $C_v = \frac{2E[L_v]}{N_v(N_v-1)} = p = \frac{k}{N-1}$
- For a fixed $k$, the clustering coefficient will decrease as the network grows with a rate $1/N$. Furthermore, the local clustering coefficient is independent of the node degree.

# References

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