**Lab – 06 answers**

**Q1. Observations:**

1. **Density**
   * Stays roughly constant around 0.2, because we fixed edge probability p = 0.2.
   * Even as the number of nodes increases, the fraction of possible edges used remains the same.
2. **Degree Distribution (Histograms)**
   * For small N (20, 50) → degree values are lower and more clustered.
   * For large N (100, 200) → average degree grows, and distribution spreads out like a bell curve.

**Observations (for N = 20, 50, 100, 200)**

* As **N increases**, the **density decreases** because the number of possible edges grows much faster than the actual number of edges.
* The **degree distribution histogram widens** as N increases:
  + For small graphs, most nodes have similar degrees.
  + For large graphs, some nodes have many edges while others have few, so the distribution spreads out.

**Question -02**

**a) Learning Methods**

* Supervised Learning: All training data is labeled. The model learns directly from input–output pairs.
* Self-Supervised Learning: Labels are generated automatically from the data itself (e.g., predicting missing words in a sentence).
* Semi-Supervised Learning: Only part of the dataset is labeled. The model uses both labeled and unlabeled data during training.

**b) Transductive vs Inductive Learning**

* Transductive Learning: The model has access to the full graph (both labeled and unlabeled nodes) during training. It can only predict labels within this graph.
* Inductive Learning: The model must generalize to unseen data or graphs. It can make predictions on nodes/graphs that were not present during training.

**Q3. Experiments on Karate Club GCN**

**a) Increasing Epochs (50 → 500)**

* At first, validation accuracy improves steadily.
* After some point, accuracy plateaus and may even slightly drop due to overfitting.
* **Observation: Training longer improves accuracy up to a limit, but does not always guarantee better generalization.**

**b) Removing Self-Loops**

* With self-loops, each node preserves its own features when aggregating neighbor information.
* Without self-loops, important self-information is lost.
* **Observation: Validation accuracy decreases when self-loops are removed.**

**c) Increasing Layers (3 → 8)**

* With more layers, nodes aggregate information from larger neighborhoods.
* However, too many layers cause over-smoothing, where all node features become too similar.
* **Observation: Accuracy drops beyond 3–4 layers.**

**d) Hyperparameters & Skip Connections**

* In\_channels / out\_channels: Control hidden layer sizes. Larger sizes may increase accuracy but risk overfitting.
* Skip Connections: Allow the model to reuse earlier features and prevent over-smoothing.
* **Observation: Skip connections slightly improve accuracy compared to plain deeper GCNs.**

**4. Types of GNNs**

* Message Passing Neural Network (MPNN): General framework where nodes exchange messages with neighbors and update their states. Almost all GNNs fall into this category.
* Graph Convolutional Network (GCN): Aggregates the mean of neighbor features. Simple, efficient, but treats all neighbors equally.
* Graph Attention Network (GAT): Uses an attention mechanism to assign different weights to neighbors, so important neighbors contribute more.
* GraphSAGE: Samples a subset of neighbors instead of using all. Useful for inductive learning, where the model needs to generalize to unseen graphs.