Machine Learning for Graphs and Sequential Data Exercise Sheet 10

Graphs & Networks, Generative Models

Problem 1: An unweighted, undirected graph without self-loops represented by an adjacency matrix $A \in \{0,1\}^{N \times N}$ is given. Prove that the number of triangles in the graph is equal to $\frac{1}{6}\,\mathrm{trace}(A^3)$ and that this term is in turn equal to $\frac{1}{6}\sum_i \lambda_i^3$ where λ_i are the eigenvalues of the adjacency matrix A. *Hint:* Show first that A_{ij}^k is the number of walks of length k from node i to node j.

Problem 2: Given is an Erdös-Renyi graph consisting of N nodes, with the edge probability $p \in [0,1]$. Derive the probability p_k that a node in the graph has degree equal to exactly k.

Problem 3: Given is an Erdös-Renyi graph consisting of N nodes with edge probability $p \in [0,1]$. What is the expected number of triangles in this graph?

Problem 4: Given are 6 graphs $\{G_1, \ldots, G_6\}$, which exhibit the properties listed in Table 1. Five of them have been synthetically generated, while one is a real graph. Assign the graphs $\{G_1, \ldots, G_6\}$ to the following models (one each) and briefly justify each answer!

a) Erdös-Renyi model

b) Stochastic block model with 5 clusters

c) Stochastic block model with 10 clusters

d) Stochastic block model with core-periphery structure

e) Initial attractiveness model

f) Real graph

Hint: for information about the "eigengap" see Sec. 8.3 in this tutorial

Problem 5: Compare the two following graph generation processes.

- Graph G_1 is generated by a stochastic block model. It consists of N nodes partitioned into $K = 2$ communities. Both communities consist of exactly $N/2$ nodes, and $\boldsymbol{\eta} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$.
- Graph G_2 is an Erdös-Renyi graph of N nodes and edge probability p.

Given the probabilities a and b, for which values of p will the expected number of triangles in G_2 be *larger* than the expected number of triangles in G_1?

w/o self-loop

Problem 1: An unweighted, undirected graph without self-loops represented by an adjacency matrix $A \in \{0,1\}^{N\times N}$ is given. Prove that the number of triangles in the graph is equal to $\frac{1}{6}\,\mathrm{trace}(A^3)$ and that this term is in turn equal to $\frac{1}{6}\sum_i \lambda_i^3$ where λ_i are the eigenvalues of the adjacency matrix A. *Hint:* Show first that A^k_{ij} is the number of walks of length k from node i to node j.

$$A^0 = I_N$$

$$A^1 = A\cdot A^0 = A\cdot I = A \Rightarrow A^1_{ij} = \begin{cases} 1 & \text{connected } i,j \\ 0 & \text{else} \end{cases}$$

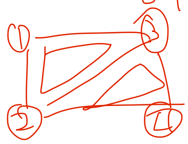
$$A^2 = A\cdot A^1 = \qquad \Rightarrow A^2_{ij} = \sum_{i=1}^{N} A_{iv}\cdot A^1_{vj}$$

$$\cdots$$

$$A^k = A\cdot A^{k-1} \qquad \Rightarrow A^{k-1}_{ij} = \sum_{i=1}^{N} A_{iv}\, A^{k-1}_{vj}$$

Note: undirected graph

Example:



$$\Rightarrow A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$k=3 \quad A^3 = A^3\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = A^2\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} = A\begin{bmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 2 & 4 & 4 & 2 \\ 5 & 4 & 4 & 5 \\ 5 & 4 & 4 & 5 \\ 2 & 4 & 4 & 2 \end{bmatrix}}_{A^3}$$

$$\frac{1}{2\cdot 3}\mathrm{trace}(A^3) = \frac{1}{6}[2+2+4+4] = 2$$

Recall: $\mathrm{trace}(A) = \sum_{i=1}^{N}\lambda_i \qquad A\vec{v} = \lambda\vec{v}$

$$A^3\vec{v} = A^2(\underbrace{A\vec{v}}_{\lambda\vec{v}}) = \lambda A^2\vec{v} = \lambda A(A\vec{v})$$
$$= \lambda^2 A\vec{v}$$
$$= \lambda^3\vec{v}$$

$$\frac{1}{6}\mathrm{trace}(A^3) = \frac{1}{6}\sum_{i=1}^{N}\lambda_i^3$$

Problem 2: Given is an Erdös-Renyi graph consisting of N nodes, with the edge probability $p \in [0,1]$. Derive the probability p_k that a node in the graph has degree equal to exactly k.

$e_{ij} \sim \text{Bernouli}(p)$ Have $N-1$ edges

$$\Rightarrow d_i \sim \text{Binomial}(N-1, p) = \binom{N-1}{k} p^k\, p^{N-1-k}$$

$k=3$

Problem 3: Given is an Erdös-Renyi graph consisting of N nodes with edge probability $p \in [0,1]$. What is the expected number of triangles in this graph?

define $\binom{N}{3}$ non-independent RVs

$$X_t = \begin{cases} 1, & \text{if triplet "t" forms triangle} \\ 0, & \text{otherwise} \end{cases}$$

$\sim \text{Bernouli}(p^3)$ identically distributed

$$\mathbb{E}_A\left[\sum_{t\in\tau} X_t\right] = \sum_{t\in\tau}\mathbb{E}[X_t]$$

$$= \underbrace{\binom{N}{3}}_{|\tau|}\mathbb{E}_A[X_t] = \binom{N}{3}p^3$$

Problem 4: Given are 6 graphs $\{G_1, \dots, G_6\}$, which exhibit the properties listed in Table 1. Five of them have been synthetically generated, while one is a real graph. Assign the graphs $\{G_1, \dots, G_6\}$ to the following models (one each) and briefly justify each answer!

a) Erdös-Renyi model

b) Stochastic block model with 5 clusters

c) Stochastic block model with 10 clusters

d) Stochastic block model with core-periphery structure

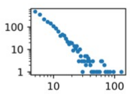
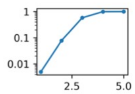
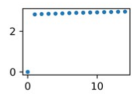
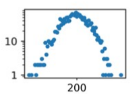
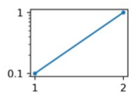
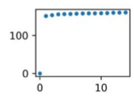
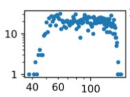
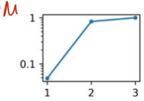
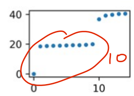
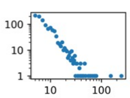
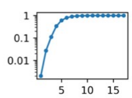
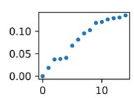
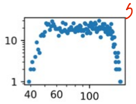
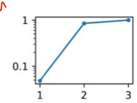
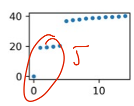
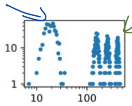
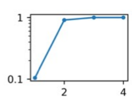
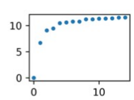
e) Initial attractiveness model

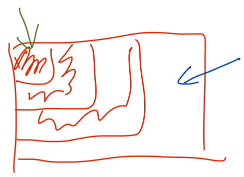
f) Real graph

a) degree ~ Possion

Hint: for information about the "eigengap" see Sec. 8.3 in this tutorial

Table 1: Graphs $\{G_1, \dots, G_6\}$

ID	Degree distribution	Hop plot	Smallest eigenvalues of Laplacian	Clustering coeff.
e G_1				0.013
a) G_2				0.100
c) G_3	 SBM		 10	0.145
f G_4				0.278
b) G_5	 SBM		 5	0.275
d) G_6				0.191



Problem 5: Compare the two following graph generation processes.

- Graph G_1 is generated by a stochastic block model. It consists of N nodes partitioned into $K = 2$ communities. Both communities consist of exactly $N/2$ nodes, and $\boldsymbol{\eta} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$.
- Graph G_2 is an Erdös-Renyi graph of N nodes and edge probability p.

Given the probabilities a and b, for which values of p will the expected number of triangles in G_2 be *larger* than the expected number of triangles in G_1?

$$G_2 = \binom{N}{3} p^3$$

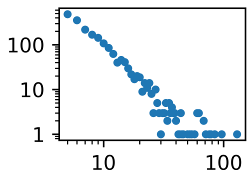
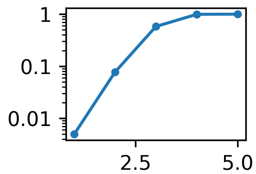
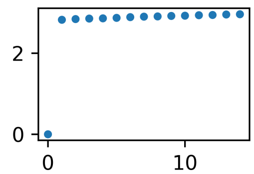
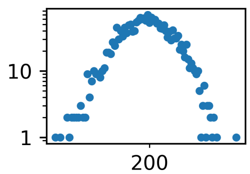
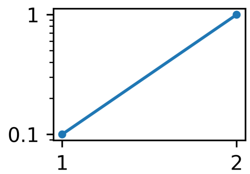
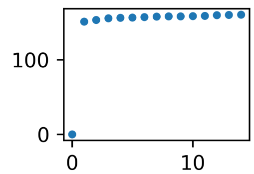
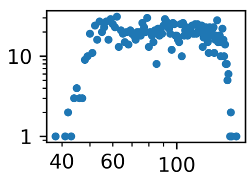
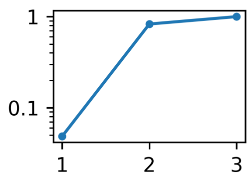
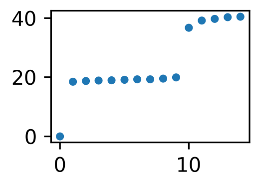
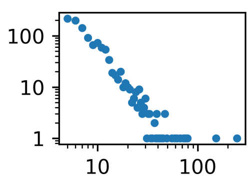
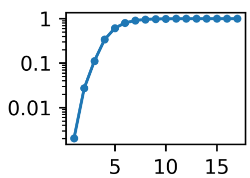
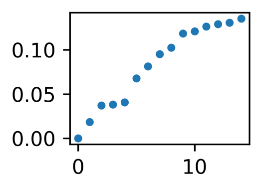
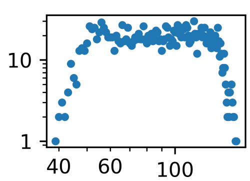
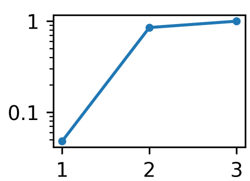
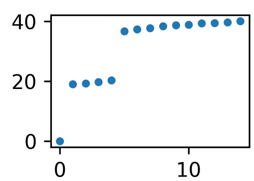
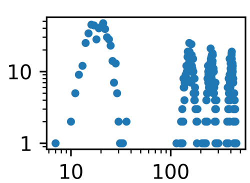
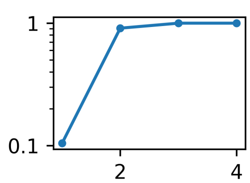
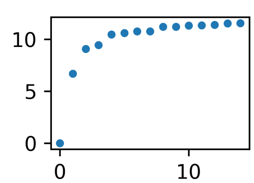
$$G_1 = \begin{array}{cc} & \\ \frac{N}{2} & \begin{bmatrix} a & b \\ b & a \end{bmatrix} \\ & \frac{N}{2} \end{array}$$

$$2 \cdot \left[\binom{N/2}{3} a^3 + \binom{N/2}{1} \binom{N/2}{2} a b^2 \right]$$

$$\Rightarrow \binom{N}{3} p^3 > 2\left[\binom{N/2}{3} a^3 + \binom{N/2}{1} \binom{N/2}{2} a b^2 \right]$$

$$p > \sqrt[3]{\frac{2\left[\binom{N/2}{3} a^3 + \binom{N/2}{1} \binom{N/2}{2} a b^2 \right]}{\binom{N}{3}}}$$

Table 1: Graphs $\{G_1, \ldots, G_6\}$

ID	Degree distribution	Hop plot	Smallest eigenvalues of Laplacian	Clustering coeff.
G_1				0.013
G_2				0.100
G_3				0.145
G_4				0.278
G_5				0.275
G_6				0.191