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 $\frac{1}{6}\operatorname{trace}(A^3)$ 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.

We will start by showing the hint by induction. The base case $A^0 = I_N$ is obviously true. Let $k \geq 1$.

$$A^k_{ij} = \sum_{v=1}^{N} A_{iv}A^{k-1}_{vj} = \sum_{v=1}^{N} A^{k-1}_{vj} \cdot \begin{cases} 1 & \text{if there is an edge between nodes } i \text{ and } v \\ 0 & \text{otherwise} \end{cases}$$

It follows that A^3_{ii} is the number of walks of length 3 from node i to itself. In an undirected graph such a walk is exactly a triangle. We need to account for the fact that we can traverse the triangle in both directions and from each starting node and each triangle is counted $2 \cdot 3 = 6$ times. It follows that

$$\frac{1}{6}\sum_i A^3_{ii} = \frac{1}{6}\operatorname{trace}(A^3)$$

is the number of triangles.

We know that $\operatorname{trace}(A) = \sum_i \lambda_i$ is the sum of the eigenvalues of A. Let v be an eigenvector of A with eigenvalue λ. Then v is also an eigenvector of A with eigenvalue λ^3 since

$$A^3 v = A^2 \lambda v = A\lambda^2 v = \lambda^3 v.$$

In combination both statements show that $\frac{1}{6}\operatorname{trace}(A^3) = \frac{1}{6}\sum_i \lambda_i^3$.

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.

A node can have up to $N-1$ neighbors and the probability for a connection is p for each of them. Its degree d is then defined as the number of edges. In other words the degree of a node is equal to the number of successes in a Bernoulli experiment with $N-1$ trials and success probability p, so

$$d \sim \operatorname{Binomial}(N-1, p)$$

and

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

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?

There are $\binom{N}{3}$ possible triangles in the graph which we call $\mathcal{T}$. Let $X_t = 1$ if a triplet $t \in \mathcal{T}$ is a triangle and 0 otherwise. Then X_t is Bernoulli distributed with success probability p^3 because three edges need to exist for a triplet of nodes to form a triangle and they exist (or not) independently. This gives the expected number of triangles as

$$d\,\mathbb{E}_A\left[\sum_{(i,j,k)\in\mathcal{T}} X_{(i,j,k)}\right] = \sum_{(i,j,k)\in\mathcal{T}} \mathbb{E}_A\left[X_{(i,j,k)}\right] = \sum_{(i,j,k)\in\mathcal{T}} p^3 = \binom{N}{3}p^3$$

where the expectation is over the adjacency matrix of the Erdös-Renyi graph.

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

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

First, we can say that G_1 and G_4 correspond to e) and f) because of their power-law degree distribution alone. We can distinguish them by their clustering behavior. The generated graph should have fewer clusters because the "rich get richer" behavior of the initial attractiveness model makes it unlikely for the many low-degree nodes to connect to each other to form triangles and clusters. Therefore G_1 was generated by an IAM and G_4 is from a real graph.

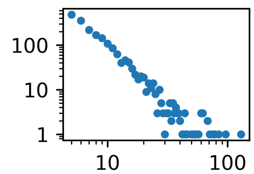
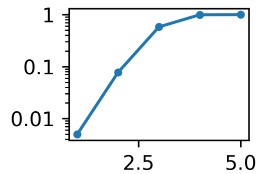
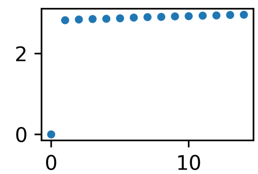
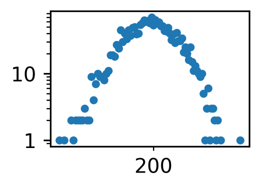
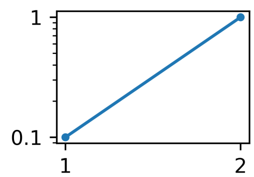
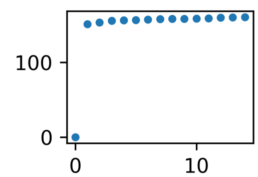
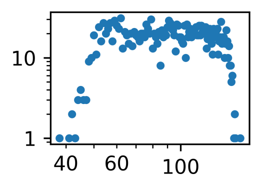
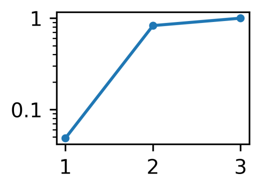
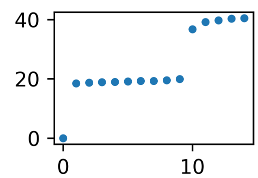
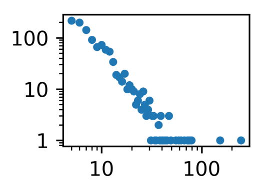
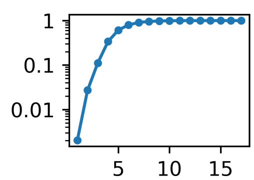
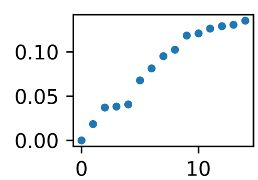
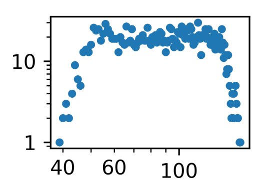
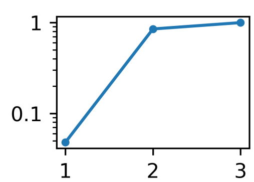
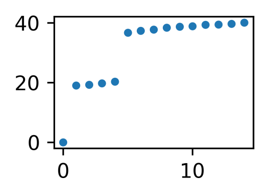
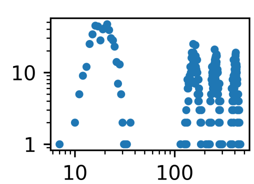
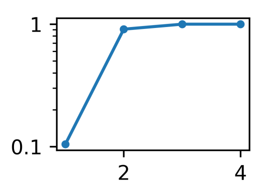
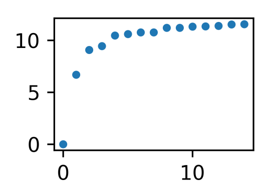
G_6 was generated from an SBM with core-periphery structure. The tell-tale sign are the clearly distinct levels of node degrees that correspond to the layers in the core-periphery structure.

Of the remaining graphs two have conspicuously similar profiles and we can recognize their source graph from their clustering structure. The eigenvalues of G_5 show a handful of small clusters while G_3 has a larger number of roughly equal sized clusters. We deduce that G_5 corresponds to an SBM with 5 clusters and G_3 to an SBM with 10 clusters.

Finally, G_2 must stem from an Erdös-Renyi model.

Problem 5: Compare the two following graph generation processes.

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

- 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?

We know from Problem 3 that G_2 has $\binom{N}{3}p^3$ triangles in expectation. For the SBM graph G_1 we know that there are $\binom{\frac{N}{2}}{3}a^3$ triangles in the first and second cluster in expectation because these are just Erdös-Renyi graphs with edge probability a. However, we also need to account for triangles between the clusters where two nodes are in cluster one and one in cluster two or vice versa. In the first case there are $\binom{\frac{N}{2}}{2}\binom{\frac{N}{2}}{1}$ such possible triangles because we choose two nodes from cluster one and one from the other. The probability of the triplet actually being a triangle is ab^2 because there needs to be one intra-cluster edges and two inter-cluster edges. This means that the expected number of triangles in this configuration is $\binom{\frac{N}{2}}{2}\binom{\frac{N}{2}}{1}ab^2$. Due to symmetry the other configuration (two nodes in cluster 2) contributes the same number of triangles. All in all G_1 has an expected number of triangles of

$$2 \cdot \left(\binom{\frac{N}{2}}{3} a^3 + \binom{\frac{N}{2}}{2} \binom{\frac{N}{2}}{1} ab^2 \right)$$

and the answer is

$$\binom{N}{3}p^3 > 2 \cdot \left(\binom{\frac{N}{2}}{3} a^3 + \binom{\frac{N}{2}}{2} \binom{\frac{N}{2}}{1} ab^2 \right) \Leftrightarrow p > \sqrt[3]{\frac{2 \cdot \left(\binom{\frac{N}{2}}{3} a^3 + \binom{\frac{N}{2}}{2} \binom{\frac{N}{2}}{1} ab^2 \right)}{\binom{N}{3}}}.$$