Problem Set 2. Complex Networks

IMPORTANT: Follow the four steps in the solution of each problem, see the document 'Problem Set Rubrics'.

Problem 1 (Exercises 4.8 and 4.9 in the book: Clustering coefficient of $ER_n(\lambda/n)$) Recall from (1.5.4) in Section 1.5 that the clustering coefficient of a graph $G = (V, E)$ with V the vertex set and E the edge set, is defined to be

$$CC_G = \frac{\mathbb{E}[\Delta_G]}{\mathbb{E}[W_G]},$$

where

$$\Delta_G = \sum_{i,j,k \in V} \mathbf{1}_{\{ij,ik,jk \text{ occupied}\}}, \quad W_G = \sum_{i,j,k \in V} \mathbf{1}_{\{ij,ik \text{ occupied}\}}.$$

Thus (since we are not restricting to $i < j < k$ in Δ_G and to $i < k$ in W_G), Δ_G is six times the number of triangles in G, W_G is two times the number of wedges in G.

(a) Interpret CC_G in terms of 'closed triangles' in social networks.

(b) Compute CC_G for $ER_n(\lambda/n)$ and describe its behavior as $n \to \infty$.

(c) Use (1) below and Exercise 4.7 in the book to conclude that as $n \to \infty$,

$$\frac{n\Delta_G}{W_G} \xrightarrow{d} \frac{6}{\lambda^2} Y,$$

where $Y \sim \text{Poisson}(\lambda^3/6)$. Relate this to (b).

Problem 2 Consider the $ER(n, \lambda/n)$ random graph and W_G as above.

(a) Is the distribution of W_G binomial? If yes, why? If no, why not and how is it related to a binomial distribution?

(b) Prove that

$$\frac{W_G}{n} \to \lambda^2 \quad \text{in probability as } n \to \infty. \tag{1}$$

Hint: Use the Chebyshev inequality and $\text{Var}(W_G) = \mathbb{E}(W_G)_2 + \mathbb{E}W_G - (\mathbb{E}W_G)^2$, where W_G can be expressed as a sum of indicators as in **Problem 1**. To compute $\mathbb{E}(W_G)_2$, use Theorem 2.5. Consider all possible cases when the two wedges contain in total 6,5,4 and 3 vertices. Note that only the term of the largest order of magnitude has to be calculated exactly. For the other terms, an upper bound is sufficient.

Problem 3 (Exercises 4.10 and 4.11 in the book) Consider the notion of monotonicity, increasing events and increasing random variables as defined in Section 4.1.1.

(a) Show that $|C_{max}|$ is an increasing random variable.

(b) Is $\{v \in C_{max}\}$ an increasing event?

Problem 4 (Exercises 4.10 and 4.11 in the book) Consider subcritical $ER_n(\lambda/n)$ graphs ($\lambda < 1$), let $|C_{n,\max}|$ denote the size of the largest connected component. Show that

$$\frac{|C_{n,\max}|}{\log n} \to \frac{1}{I_\lambda} \quad \text{in probability as } n \to \infty.$$

Use Theorems 4.4 and 4.5.

Problem 5. Consider a random variable X that has Pareto distribution

$$P(X > x) = Cx^{-\tau+1}, \quad \tau > 1, \quad x \geq x_0 > 0.$$

(a) Show that $\mathbb{E}(X^p) < \infty$ iff $p < \tau - 1$.

(b) Let $X_1, X_2, \ldots, X_n$ be independent copies of X, and let

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

be the order statistics, see (2.6.1). Use either a heuristic argument or Theorem 2.33 (or both) to describe the asymptotic behavior of $X_{(n-k)}$ as $n \to \infty$.

Problem 6. Numerical assignment. Generate the $ER_n(\lambda/n)$ graph. Investigate one of the following:

Option 1. Investigate the phase transition and emergence of the giant component. Consider different $\lambda > 0$ and different values of n.

Option 2. Denote by Δ_n the number of triangles in the $ER_n(\lambda/n)$ graph. Verify empirically that

$$\Delta_n \Rightarrow \text{Poi}\left(\frac{\lambda^3}{6}\right) \quad \text{as } n \to \infty.$$