Complex Networks

Teachers: M. Emmerich, D. Garlaschelli, F. den Hollander.
Written examination: Wednesday, 25 January 2015, 10:00–13:00.

Open book exam: the lecture notes may be consulted, but no other material.

Answer each question on a separate sheet. Put your name, student number and the number of the question you are answering on each and every sheet. Provide full explanations with each of the answers.

Each question is weighted by a number of points, as indicated. The total number of points is 100. The final grade will be calculated as a weighted average: 30% for homework assignments and 70% for this exam.

Success!

---

1. **[5 points]**
   Give two examples of economic networks and describe their main features.

2. **[7 points]**

   2a. Give the definition of the clustering coefficient of a graph G.

   2b. When is a sequence of random graphs $(G_n)_{n\in\mathbb{N}}$ called scale free?

   2c. For what choices of λ is the sequence of Erdős-Rényi random graphs $(\mathrm{ER}_n(\lambda/n))_{n\in\mathbb{N}}$ scale free?

3. **[7 points]**

   3a. Draw all the possible outcomes of $\mathrm{PA}_2(1,2)$.

   3b. What is the probability of each outcome?

   3c. Are all the outcomes simple?

4. **[16 points]**
   For a real-world graph $\mathbf{G}^*$ with $N = 10$ vertices and $L = 15$ undirected links, it is found that, once solving the maximum-likelihood equations, the Park-Newman (PN) model gives a matrix $\mathbf{P}_{PN} = (p_{ij})$ of connection probabilities equal to the corresponding matrix $\mathbf{P}_{ER}$ that would be obtained by using the Erdős-Rényi (ER) model (with maximized likelihood) for the same graph.

   4a. Calculate the matrix $\mathbf{P}_{PN}$.

   4b. Determine the degree sequence of $\mathbf{G}^*$.

   4c. Calculate the empirical degree distribution of $\mathbf{G}^*$.

   4d. Calculate the vector of hidden variables $\vec{x}^* = (x_1^*, \ldots, x_{10}^*)$ that maximizes the likelihood of the PN model, given $\mathbf{G}^*$.

4e. Calculate the z-score $z_{PN}[k_i] \equiv \frac{k_i(\mathbf{G}^*)-\langle k_i\rangle_{PN}}{\sigma_{PN}[k_i]}$ for each degree k_i under the PN model (where $\langle k_i\rangle_{PN}$ and $\sigma_{PN}[k_i]$ denote the expected value and the standard deviation of k_i under the PN model respectively, while $k_i(\mathbf{G}^*)$ denotes the empirical value of k_i in the real-world network).

4f. Calculate the expected average nearest-neighbor degree $\langle k_i^{nn}\rangle_{PN}$ for all nodes under the PN model.

4g. Imagine that the Local Rewiring Algorithm is applied 1000 times to $\mathbf{G}^*$, and denote the new graph by $\mathbf{G}^{**}$. Calculate the ratio between the probability of occurrence of $\mathbf{G}^*$ and that of $\mathbf{G}^{**}$ under the PN model.

5. **[6 points]**

5a. The following program initializes a graph in R:

```
G <- c(1,1, 2,1, 3,4, 4,1, 5,5)
```

Describe the graph by means of its adjacency matrix.

5b. How much memory in terms of the number of vertices $|V|$ and number of edges $|E|$ is required to store a *connected* undirected graph as an adjacency list and how much memory is required to store it as an adjacency matrix? Use the asymptotic notation (big O notation) to describe the worst case space complexity when using these data structures.

6. **[12 points]**

6a. State an algorithm for generating a self-avoiding (i.e., no duplicate vertices) random path

$$v_0 \xrightarrow{e_1} v_1 \xrightarrow{e_2} v_2 \xrightarrow{e_3} \dots \xrightarrow{e_k} v_k$$

of length k in a graph. Provide pseudocode. It is sufficient to output edges. What is the time complexity of the algorithm? (Hint: consider a Fisher Yates shuffle.)

6b. Describe briefly the advantages and the disadvantages of the following three algorithms for sampling from the configuration model: (1) random link stub connection; (2) repeated random link stub connection; (3) link stub reconnection (with local search operator).

7. **[8 points]** Consider the configuration model $\mathrm{CM}_n(\vec{K})$ with n vertices and degree sequence $\vec{K} = (K_1, \dots, K_n)$ whose components are i.i.d. random variables with probability distribution function f given by

$$f(1) = f(2) = f(3) = \tfrac{1}{3}, \qquad f(k) = 0 \text{ otherwise},$$

A "hacker" removes each vertex of degree 3 with probability $q \in (0,1)$, independently of other vertices. For what values of q does the "mutilated" random graph percolate for $n \to \infty$ (i.e., the largest connected component has a size of order n with a probability tending to 1 as $n \to \infty$)?

8. **[9 points]**
Consider the contact process with parameter $\lambda \in (0, \infty)$ on the complete graph with n vertices.

   8a. Let $X(t)$ denote the total number of infections at time t. Explain why $X = (X(t))_{t \geq 0}$ is a continuous-time Markov chain. Argue why the rate for the transition $j \to j-1$ equals j and the rate for the transition $j \to j+1$ equals $\lambda j(n-j)$.

   8b. Define $e(j) = \mathbb{E}(\tau_0 | X(0) = j)$, $0 \leq j \leq n$, where $\tau_0 = \inf\{t \geq 0\colon X(t) = 0\}$ is the time to extinction. Obviously, $e(0) = 0$. Show that

$$e(1) = \frac{1}{1+\lambda(n-1)} + \frac{1}{1+\lambda(n-1)} e(0) + \frac{\lambda(n-1)}{1+\lambda(n-1)} e(2).$$

   Write down a similar recursion relation linking $e(j)$ to $e(j-1)$ and $e(j+1)$ for $2 \leq j \leq n$ (with the convention that $e(n+1) = 0$).

   8c. Use the recursion relations to compute $e(j)$, $1 \leq j \leq n$, when $n = 3$ and $\lambda = 1$.

9. **[16 points]**
A real-world network $\mathbf{G}^*$ with $N = 5$ vertices and $L = 6$ undirected links is modeled via the Chung-Lu (CL) version of the Configuration Model. The resulting matrix of connection probabilities reads

$$\mathbf{P}_{CL} = (p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/4 & 1 & 1/2 \\ 1/2 & 0 & 1/6 & 2/3 & 1/3 \\ 1/4 & 1/6 & 0 & 1/3 & 1/6 \\ 1 & 2/3 & 1/3 & 0 & 2/3 \\ 1/2 & 1/3 & 1/6 & 2/3 & 0 \end{pmatrix}.$$

   9a. Determine the degree sequence $\vec{k}(\mathbf{G}^*)$.

   9b. Draw the graph $\mathbf{G}^*$ and calculate its adjacency matrix.

   9c. Calculate the local clustering coefficient c_i for all nodes in $\mathbf{G}^*$ (for nodes with degree 1, conventionally set the clustering coefficient to 0).

   9d. For which pairs of nodes in $\mathbf{G}^*$ is the CL model giving a higher connection probability than the Erdős-Rényi (ER) model with maximized likelihood?

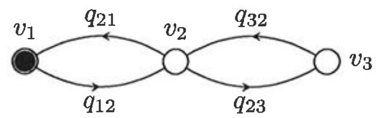
   9e. Calculate the probability $P_{ER}(\mathbf{G}^*)$ of occurrence of graph $\mathbf{G}^*$ under the ER model.

   9f. Draw the most likely graph(s) under the CL model.

   9g. Draw the most likely graph(s) under the ER model.

10. **[6 points]**

10a. A continuous-time Markov chain (CTMC) simulation of a piece of information spreading over a network with three nodes has to be simulated. The network has a chain structure, as indicated below, and the initial state is always that only the first node is infected (see picture below). The rate of a vertex informing a yet uninformed neighbour is $q_{i,j}$. What is the state space? Use a state space that is as small as possible, but still allows a simulation of the described process.



10b. What is the generator matrix of this process?

10c. What is the expected time until all nodes in the network are infected?

11. **[8 points]**

11a. Describe an efficient algorithm for computing the shortest path length from a source vertex to all other vertices in an undirected and edge-weighted, connected, and cycle-free graph (tree). What is the time complexity of your algorithm?

11b Discuss whether or not the following problem can be solved in polynomial time: Given a undirected, edge weighted graph $G = (V, E)$ with positive weights. What is the vertex with smallest *closeness centrality*? Discuss whether under the NP $\neq$ P conjecture this problem belongs to the complexity class P, NP, NP complete, or NP hard? (it might belong to multiple classes)