# Complex Networks

---

*Teachers:* D. Garlaschelli, F. den Hollander, A. Plaat.
*Written examination:* Wednesday, 16 December 2015, 10:00–13:00.

---

Open book exam: the lecture notes may be consulted, but no other material. See the remark after Exercise 11.

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]** List the four main categories into which real-world networks are classified and give a short description of an example in each category.
2. **[7 points]**
   - 2a. Give the definition of typical distance in a graph $G$.
   - 2b. When is a sequence of random graphs $(G_n)_{n\in\mathbb{N}}$ called small world?
   - 2c. Is the sequence of Erdős-Rényi random graphs $(\mathrm{ER}_n(\lambda/n))_{n\in\mathbb{N}}$ small world for all choices of $\lambda \in (0,\infty)$?
3. **[7 points]**
   - 3a. Draw all the possible outcomes of $\mathrm{CM}_3((1,1,2,2))$.
   - 3b. What is the probability of each outcome? (Note that different pairings of half-edges may lead to the same outcome.)
   - 3c. Are all the outcomes simple?
4. **[16 points]**
   A real-world network is represented as a simple (i.e., with no multiple edges and self-loops) undirected graph $\mathbf{G}^*$ with $n = 5$ vertices. Chung and Lu decide they want to compare their model with the real-world network. They define their connection probabilities $\{p_{ij}\}$ (with $p_{ii} \equiv 0\ \forall i$) and, after computing them on the real network, they find that $p_{34} > p_{53}$, $p_{15} > p_{25}$, $p_{43} > p_{14}$, $p_{52} = p_{45}$.
   - 4a. Find the degree sequence $\vec{k}(\mathbf{G}^*)$ of the real-world network $\mathbf{G}^*$. Explain your result.
   - 4b. Draw $\mathbf{G}^*$. Explain your result.

4c. Write the matrix $\mathbf{P}$ having the numerical values of $\{p_{ij}\}$ as entries.

4d. Write the probability of occurrence of $\mathbf{G}^*$ under the model used by Chung and Lu.

4e. Draw all the graphs that have the maximum probability in the model.

4f. Is $\mathbf{G}^*$ found among the graphs with maximum probability? Explain why.

5. **[8 points]**

Whenever this exam asks to "provide an algorithm," you are requested to do so in a programming language that is close to Python, Java, C, or C++. Here "close" means that we are interested in the algorithmic essence. No points will be deducted for simple syntactic omissions such as missing semi-colons. The algorithmic meaning, however, should be clear beyond doubt. Do provide declarations of essential variables. Provide algorithmic detail with low level operations, no high level Python functions. See also definition of pseudo-code at the end of this exam.
Of an undirected graph the following edge list is given:

| $v_1$ | $v_2$ |
|---|---|
| 1 | 2 |
| 2 | 3 |
| 2 | 4 |
| 2 | 5 |
| 2 | 6 |
| 4 | 6 |

5a. Draw the graph.

5b. Provide the adjacency matrix of this graph in algorithmic Pseudo-code.

5c. What is the mean of the number of edges of the vertices? Provide an algorithm to compute this number.

5d. What is the time complexity of this algorithm? Why?

5e. Name two strategies to help you ascertain that your program is correct.

6. **[12 points]**

6a. Provide an algorithm for the Configuration Model based on an adjacency matrix. Describe any assumptions, imperfections or limitations of your program.

6b. What is the computational time complexity of the program?

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

variables with probability distribution function $f$ given by

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

conditioned to satisfy $K_1 + \cdots + K_N =$ even. 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 triangle. Let $X(t)$ denote the total number of infections at time $t$. Define $e(j) = \mathbb{E}(\tau_0 | X(0) = j)$, $j = 0, 1, 2, 3$, where $\tau_0 = \inf\{t \geq 0 \colon X(t) = 0\}$ is the time to extinction.

   8a. Write down the transition rates for the continuous-time Markov process $X = (X(t))_{t \geq 0}$.

   8b. Write down four equations linking $e(j)$, $j = 0, 1, 2, 3$.

   8c. Use these equations to compute $e(j)$, $j = 0, 1, 2, 3$.

9. **[16 points]**
A real-world network is represented as a simple (i.e., with no multiple edges and self-loops) undirected graph $\mathbf{G}^*$ with $n = 5$ vertices. Park and Newman, not satisfied with the model used by their colleagues Chung and Lu in Problem 4, decide they want to compare their own model with the real-world network. Park and Newman define their connection probabilities $\{p_{ij}\}$ (with $f(x_i) \equiv x_i$ and $p_{ii} \equiv 0\ \forall i$) and, after computing the hidden variables $\{x_i^*\}$ on the real network using the Maximum Likelihood principle, they find that $p_{34} > p_{53}$, $p_{15} > p_{25}$, $p_{43} > p_{14}$, $p_{52} = p_{45}$.

   *Note*: This problem refers to Problem 4, but is completely independent of the solution of Problem 4, so the two can be solved in any order.

   9a. Find the degree sequence $\vec{k}(\mathbf{G}^*)$ of the real-world network $\mathbf{G}^*$. Explain your result.

   9b. Draw $\mathbf{G}^*$. Explain your result.

   9c. If $\mathbf{P}$ denotes the matrix having the numerical values of $\{p_{ij}\}$ as entries, calculate the marginals of $\mathbf{P}$, defined as the $n$ row sums $\sum_{j=1}^{n} p_{ij}\ \forall i$.

   9d. Calculate the average nearest-neighbour degree $k_i^{\text{nn}}(\mathbf{G}^*)$ for each node $i$ of $\mathbf{G}^*$.

   9e. Calculate the clustering coefficient $C_i(\mathbf{G}^*)$ of each node $i$ for $\mathbf{G}^*$. (For nodes with degree 1, conventionally set the clustering coefficient to 0.)

   9f. Let $P(\mathbf{G}^*)$ be the probability of occurrence of the real-world network $\mathbf{G}^*$ in the model used by Park and Newman. Write $P(\mathbf{G}^*)$ as a function of the degree sequence of $\mathbf{G}^*$ and discuss qualitatively what happens to $P(\mathbf{G}^*)$ when the parameters $\{x_i\}$ are varied.

10. **[6 points]**

    10a. Provide an algorithm to compute the Empirical Average Nearest Neighbor degree. Assume as data structure an adjacency *matrix*.

    10b. What is the computational time complexity of the program? Why?

11. **[6 points]**

    11a. Provide an algorithm to compute the Empirical Average Nearest Neighbour degree. Assume as data structure an adjacency *list*.

    11b. What is the computational time complexity of the program? Why?

Wikipedia's entry on Pseudo-code says the following:
*It is an informal high-level description of the operating principle of a computer program or other algorithm.*

*It uses the structural conventions of a programming language, but is intended for human reading rather than machine reading. [...] The purpose of using Pseudocode is that it is easier for people to understand than conventional programming language code, and that it is an efficient and environment-independent description of the key principles of an algorithm. It is commonly used in textbooks and scientific publications that are documenting various algorithms.*