Complex Networks

Teachers: M. Emmerich, D. Garlaschelli, F. den Hollander.
Written examination: Wednesday, 21 December 2016, 10:00–13:00.

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

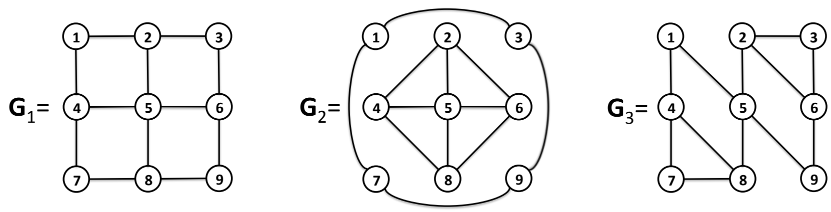
Answer each question on a separate sheet. Put your name, student number and the number of the question you are answering on each and every sheet. Provide full explanations with each of the answers.

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]**
   Why is the WWW a complex network? Describe its main features.
2. **[6 points]**
   2a. Draw the possible outcomes of the Erdős-Rényi random graph with 3 vertices.
   2b. What are the probabilities of each outcome when the edge probability is $\frac{1}{2}$?
   2c. Are all the outcomes simple?
3. **[9 points]**
   Consider the Configuration Model $\mathrm{CM}_n(\vec{k})$ with n vertices and degree sequence $\vec{k} = (k_1, \ldots, k_n)$.
   3a. For $n = 2$ and $n = 3$, give choices of $\vec{k}$, with $\sum_{i=1}^n k_i$ even and $1 \leq k_i \leq n$ for $1 \leq i \leq n$, that are not graphical.
   3b. Suppose that $K_1, \ldots, K_n$ are i.i.d. random variables drawn from a degree distribution f. Under what conditions on f is the conditional probability for $\mathrm{CM}_n(\vec{K})$ to be simple, given that $\sum_{i=1}^n K_i$ is even, strictly positive in the limit as $n \to \infty$?
   3c. Compute the probability in 3b when $f(k) = (1-q)q^k$, $k \in \mathbb{N}_0$.
4. **[16 points]**
   Consider the 3 graphs shown in the figure below. Assume that each of these graphs is studied using 3 alternative models: *i)* the Erdős-Rényi

(ER) random graph model (with parameter p^* fixed via maximum-likelihood); *ii)* the Chung-Lu (CL) specification of the configuration model; *iii)* the Park-Newman (PN) specification of the configuration model (with hidden variables $\vec{x}^*$ fixed via maximum likelihood).



4a. Which graph has the highest probability of occurrence under the ER model? Which graph has the lowest probability of occurrence under the ER model?

4b. Which graph has the highest probability of occurrence under the PN model? Which graph has the lowest probability of occurrence under the PN model?

4c. For which node in which graph is the empirical degree k_i^* closest to the corresponding expected degree $\langle k_i \rangle_{ER}$ calculated under the ER model?

4d. In which graph(s) are the four links with highest probability of occurrence under the CL model realized?

4e. In which graph(s) are the four links with *lowest* probability of occurrence under the PN model *not* realized?

4f. For which node in which graph is the empirical clustering coefficient c_i^* closest to the corresponding expected clustering coefficient $\langle c_i \rangle_{ER}$ calculated under the ER model?

5. [**6 points**]

5a. The following graph is described in R/igraph:

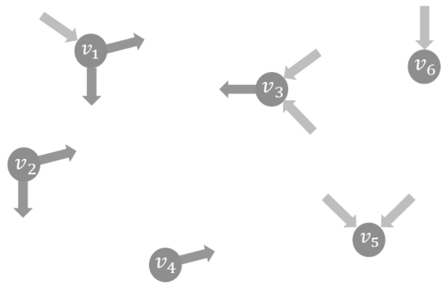
```
> g<-graph.formula(1-+2, 2-+3:1, 4)
```

Draw the graph and write up its adjacency matrix.

5b. Show that the storage of a graph $G = (V, E)$ with known and constant maximal degree of k requires only linear space complexity, i.e., a space complexity in $O(|V|)$.

6. [**12 points**]

6a. The figure shows vertices with stubs for incoming and outgoing edges:



Write a program (pseudocode) that randomly connects the output stubs with the input stubs, such that every input stub gets matched with exactly one output stub. Try to propose an efficient algorithm and report its time complexity. (Hint: you may use a Fisher Yates shuffle.) Assume that input and output stubs are numbered from 1 to n_{stubs}, and it is sufficient to output the edges as pairs of numbers.

6b. Discuss problems that may occur when we simply connect stubs (as described above) with the goal to generate graphs $G = (V, E)$ with

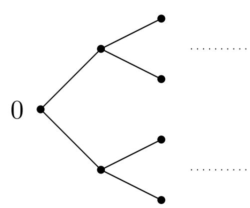
$$E \subseteq \{(v, u) \,|(v, u) \in V \times V \text{ and } v \neq u\}.$$

7. **[8 points]**

7a. Consider ordinary percolation on $\mathbb{Z}^d$. Let $p_c(\mathbb{Z}^d)$ denote the critical threshold. Prove that $d \mapsto p_c(\mathbb{Z}^d)$ is non-increasing.

7b. Why is it plausible that for all $p > p_c(\mathbb{Z}^d)$ there is a unique infinite cluster? (Hint: Use that the distribution of clusters is invariant under lattice translations.)

7c. Consider ordinary percolation on the infinite binary tree $\mathcal{T}_2$, where the origin is the root. (Two generations of the tree are drawn below.) Show that $p_c(\mathcal{T}_2) = \frac{1}{2}$.



8. **[8 points]**

8a. Explain why the contact process on a finite graph cannot sustain an epidemic, no matter how large the infection parameter λ is.

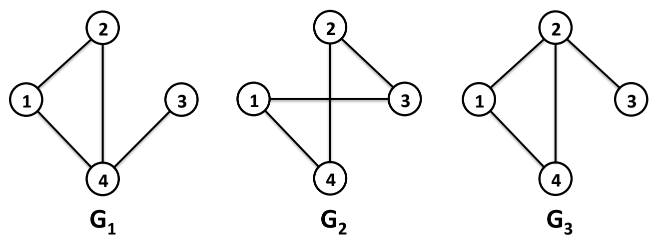
8b. Explain why, despite the fact stated in 8a, the average time the epidemic survives on the torus $\Lambda_N \subset \mathbb{Z}^d$ of side length N grows rapidly with N when $\lambda > \lambda_c(\mathbb{Z}^d)$ and slowly with N when $\lambda < \lambda_c(\mathbb{Z}^d)$.

8c. Is the behaviour stated in 8b the same for scale free graphs as the number of vertices n tends to infinity?

9. **[16 points]**
A real-world graph $\mathbf{G}^*$ is modeled using the Chung-Lu (CL) specification of the configuration model, and the resulting matrix of connection probabilities turns out to be

$$(p_{ij}) = \begin{pmatrix} 0 & 3/4 & 1/4 & 1/2 \\ 3/4 & 0 & 3/8 & 3/4 \\ 1/4 & 3/8 & 0 & 1/4 \\ 1/2 & 3/4 & 1/4 & 0 \end{pmatrix}.$$

9a. Among the graphs shown below, one is $\mathbf{G}^*$. Identify it.



9b. Calculate the expected degree $\langle k_i \rangle_{CL}$ of each node ($i = 1, 4$) under the CL model. Does $\langle k_i \rangle_{CL}$ coincide exactly with the empirical degree $k_i(\mathbf{G}^*)$ for all i?

9c. Calculate the expected average nearest neighbour degree $\langle k_i^{nn} \rangle_{CL}$ for all nodes under the CL model.

9d. Calculate the expected degree $\langle k_i \rangle_{PN}$ of each node ($i = 1, 4$) under the PN specification of the configuration model, after the latter is fitted to $\mathbf{G}^*$ using the maximum likelihood principle. Does $\langle k_i \rangle_{PN}$ coincide with $\langle k_i \rangle_{CL}$ for all i?

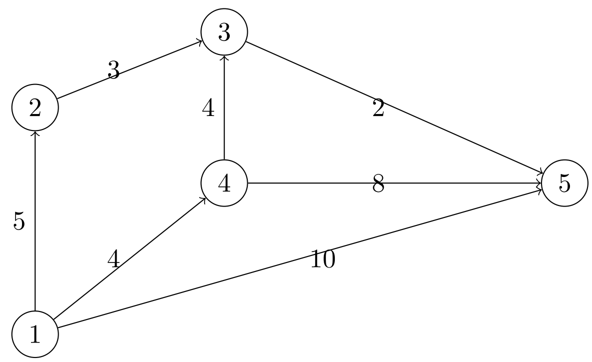
10. **[6 points]**

10a. Describe the state space and the generator matrix for the following continuous-time Markov chain simulation for an SI process on a graph with two vertices. Vertex 1 can infect vertex 2 with rate λ_{12} and heals with rate μ_1. Vertex 2 can infect vertex 1 with rate λ_{21} and heals with rate μ_2.

10b. Given that the system has just entered the state where vertex 1 is infected and vertex 2 is not infected, what is the average waiting time until the next event? What is the variance of that waiting time?

11. **[8 points]**

11a. Given the graph:



Describe the order in which the nodes leave the queue T of not yet visited nodes in Dijkstra's algorithm for computing the shortest path.

11b. What is the Page rank centrality of node 1 when a damping factor of $\alpha = 0.85$ is used? Suppose that someone states: "The *page rank centrality* is essentially the same as the degree of a vertex". Describe informally which important aspects of page rank centrality this statement omits or states wrongly.