



INSTITUTO SUPERIOR TÉCNICO | UNIVERSIDADE DE LISBOA

NETWORK SCIENCE 2021/22

1st Exam | 26 November 2021 (2h)

The exam can be answered in Portuguese or in English in a standard IST exam sheet. All questions are equally valued.

1. i) Briefly explain what a complex system is; ii) suggest two examples of a complex system.

R: (several answers can be considered correct) A complex system is a system composed of many elements which may interact with each other. Examples include social insects, the brain, infrastructures such as power grid, transportation or communication systems, social and economic organizations, financial markets, etc. The fun part of it is that, typically, the global collective behavior is not easily predictable from the individual behavior of its elements. In these cases, such collective patterns or properties are also referred as “emergent”.

2. Provide examples of i) two networks of relevance in biological/medical sciences, and ii) two networks of relevance in social/economic sciences. For each case, please indicate what a node and link represent.

R: Metabolic networks, PPI networks, food webs, lending networks among banks, friendship networks, facebook, etc. All of these options shall be considered as correct.

3. Briefly explain a graph's clustering coefficient and how we can compute it by providing an algorithm and its time complexity in the worst case.

R: The clustering coefficient is the average of the local clustering coefficient C_i over all vertices in G . The local clustering coefficient C_i of a vertex quantifies how close its neighbors are to be a clique (complete graph). For each node i with degree k_i , one can compute the local cluster coefficient C_i of i as the number of links e_i between vertices within the neighborhood of i divided by the total number links that could possibly exist between them ($k_i(k_i-1)/2$). Hence, we are indirectly counting triangles. Algorithm (several options can be considered correct... here's one possible answer): Given a vertex, we have to check if each one of the $k_i(k_i-1)/2$ pairs of neighbors are linked (we should use a hash table or an adjacency matrix to check link existence in constant time). For each vertex, we have at most $N-1$ neighbors and, hence, the overall running time is $O(|N|^3)$.

4. Suggest a simple benchmark (i.e., a network model) useful to assess the accuracy of a community finding algorithm. You may briefly explain the idea of one of the benchmarks we discussed in our classes or propose a new one.

R: Imagine a simple variant of the Girvan-Newman benchmark. Start with N nodes partitioned into n_c communities of equal size. Each node is connected with probability p_{int} to the nodes in its community and with probability p_{ext} to the nodes in the other communities. These probabilities may be adapted to have a fixed average degree or not. The control parameter $\mu = k_{ext}/(k_{ext} + k_{int})$ captures the density differences within and between communities. We expect community finding algorithms to perform well for small μ . The performance of all algorithms should drop for large μ , when the link density within the communities becomes comparable to the link density in the rest of the network. Other possible answer would be to consider the Lancichinetti–Fortunato–Radicchi benchmark. This algorithm generates networks that have a priori known communities. Once again, these networks are used to compare different community detection methods. In this case, the algorithm is able to account for the heterogeneity in the distributions of node degrees and of community sizes. Finally, one can also consider a real-world dataset with known communities, such as the Zachary's karate club.

5. Briefly i) explain the Barabási-Albert model and ii) how you would modify it to explain the emergence of power-law degree distributions with a high-degree cutoff.

R (general idea): i) Growth + preferential attachment, ii) (e.g.) costs for high degree or imposing an age limit.

6. One of the puzzling features of a scale-free network is that the second moment of its degree distribution diverges for a large number of nodes ($N \rightarrow \infty$). Indicate two implications of this property. [60%+40%]

R (general idea): Network robustness, absence of an epidemic threshold, high speed of information spreading, (etc).

7. a) Suggest one possible disease spreading model underlying the dynamics associated with the ongoing COVID-19. Justify. b) How would you modify the model proposed in (a) to include age-specific parameters and interactions. c) Indicate the technique you would adopt to characterize the time evolution of the disease outbreak emerging from your model in (b) for different containment measures.

R: a) Several options are acceptable, such as a simple SIR model, a SEIR, SIRS or a SEIRS models. We may also pick one of those more complex models we have discussed in classes. In the SIR model, individuals may be in one of 3 compartments at each time t . $S(t)$ is used to represent the individuals not yet infected with the disease at time t , or those susceptible to the disease of the population. $I(t)$ denotes the individuals of the population who have been infected with the disease and are capable of spreading the disease to those in the susceptible category. $R(t)$ is the compartment used for the individuals of the population who have been infected and then removed from the disease, either due to immunization or due to death. Those in this category are not able to be infected again or to transmit the infection to others. If the SEIR model is considered, there's an additional compartment E representing infected individuals under latent period of the disease where the person is not infectious. One may also consider the SIRS model where immunity may last only for a given period of time.

b) Taking SIR model as an example, one should break each compartment into each age class, defining each quantities (β , $\langle k \rangle$ and δ) for each of this "age-structured" compartment models and interaction rates among them.

c) The model can be analyzed using a computer simulation, where the above transitions are explicitly introduced as a stochastic process. As an alternative, one may describe these models via ordinary differential equations and integrate numerically the resulting system of equations.

8. Briefly explain how we can use the degree correlation function ($k_{nn}(k)$) to measure the degree-assortativity of a graph.

R: Degree correlations capture the relationship between the degrees of nodes that link to each other. One way to quantify their magnitude is to measure for each node i the average degree of its neighbors. The degree correlation function, $k_{nn}(k)$, calculates this quantity for all nodes with degree k . In other words, $k_{nn}(k)$ is the average degree of the neighbors of all degree- k nodes. To quantify degree correlations we inspect the dependence of $k_{nn}(k)$ on k . In assortative networks hubs tend to connect to other hubs, hence the higher is the degree k of a node, the higher is the average degree of its nearest neighbors. Consequently for assortative networks $k_{nn}(k)$ increases with k . In disassortative network hubs prefer to link to low-degree nodes. Consequently $k_{nn}(k)$ decreases with k . For a neutral network, plotting $k_{nn}(k)$ as a function of k should result in a horizontal line.

9. Resort to the friendship paradox to suggest a vaccination strategy for scale-free networks, which is likely to perform better than random vaccination, and does not require complete information on the network topology

R: The friendship paradox is the idea that most people have fewer friends than their friends have, on average. This is particularly true in scale-free networks, as this average increases with the second moment of the degree distribution, due to the high probability of having a highly connected node in our neighbourhood. If that's true, we can use this principle to vaccinate the acquaintances of randomly selected individuals, indirectly targeting the hubs without having to know precisely which individuals are hubs. Example: 1) Choose randomly a fraction p of nodes (Group 0). Select randomly a link for each node in Group 0. Let us call Group 1 to this new set of nodes. Immunize the Group 1 individuals. Proceed like this, creating a group 2 if needed, from the neighbors of the vertices in group 1.

10. Please indicate whether each of the following statements is TRUE or FALSE.

Note: For each wrong answer, we discount a correct one.

- a) Indirect reciprocity is a mechanism for the evolution of cooperation based on the idea of repeated games between the same two players.
- b) In the Erdős–Rényi model, the average-path-length (APL) scales with $\ln N$, where N is the number of nodes.
- c) Newborns are enough to explain the emergence of recurrent epidemics (e.g., in Measles).
- d) The Threshold Voter Model encodes a simple form of complex contagion.
- e) A K-core is the largest subgraph whose nodes have at least K interconnections.
- f) The harmonic centrality of a node can only be computed for connected graphs.
- g) An evolutionary stable strategy is a strategy that, if adopted by all individuals in a population, cannot be replaced or invaded by a different strategy through natural selection.

R: a) F, b) T, c) F, d) T, e) T, f) F, g) T