INSTITUTO SUPERIOR TÉCNICO | UNIVERSIDADE DE LISBOA

NETWORK SCIENCE 2020/21

1st Exam | Wednesday, 20 January 2021 (2h), 18h30

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

1. The "six degrees of separation" concept was examined and proposed in Stanley Milgram's 1967 "small-world experiment". **a)** Briefly explain the idea of the experiment and how it was implemented. **b)** Suggest a network measure able to assess the "six degrees of separation" concept. **c)** Suggest an algorithm to compute your network measure (considering directed, unweighted graphs) and **d)** indicate its complexity (indicate the network data structure you assumed).

R: a) In the experiment, Milgram sent several packages to 160 random people living in Omaha, Nebraska, asking them to forward the package to a friend or acquaintance who they thought would bring the package closer to a set final individual, a stockbroker from Boston, Massachusetts. Each individual could only mail the folder to someone they actually knew personally on a first-name basis. When doing so, each sender also instructed their recipient to mail the folder to one of the latter's first-name acquaintances with the same instructions, with the hope that their acquaintance might by some chance know the target recipient. Milgram reported that chains varied in length from two to ten intermediate acquaintances, with a median of five intermediate acquaintances (i.e. six degrees of separation) between the original sender and the destination recipient. b) Essentially, we need to compute number of people you have to communicate with (on average) to reach a complete stranger in a social network. The average path length (APL) provides just that. APL is the average distance (= length of shortest path) between all pairs of nodes in the network. c) It may be calculated through a BFS starting from every node. d) Assuming a linked list, it takes O(V(V+E)), where V is the number of vertices and E the number of edges.

2. One of the main goals of computational epidemiology is to model the spread of pathogens in populations. A predominant method of doing so is to abstract the population into compartments, which represent their health status with respect to the pathogen. Several models can be devised, depending on the data available, and the characteristics of the disease and the population. **a)** Suggest one possible model underlying the dynamics associated with the ongoing COVID-19 pandemic. Justify. **b)** List all free variables of your model.

R: a) Several options are acceptable, such as a simple SIR model, or a SEIR model. In the SIR model, individuals may be in one of 3 compartments in 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) The SIR model depends critically on three quantities: β, <k> and δ. Firstly, the product β<k> represents the average number of contacts per person per time <k>, multiplied by the probability of disease transmission in a contact between a susceptible and an infectious subject (β). Typically, if I(t) is the number (or fraction) of infected, β<k>I(t) represents the force of infection, i.e., the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals. Secondly, the SIR model depends on the recovery rate δ, which can be seen as the inverse of the infectious period (e.g., given in days). It models the transition rate from the compartment of infected individuals to the compartment of recovered individuals. The SEIR model would have an additional rate associated with the incubation time.

3. The basic reproduction number, or basic reproductive number, denoted as R_0, stands as a key parameter in epidemiology. Briefly explain **a)** what it means and **b)** how can we use it to estimate the critical proportion w_c of a population to be vaccinated in order to avoid a disease outbreak described as a SIR model in a well-mixed population. **c)** How would your reasoning in (b) change if the efficiency of your vaccine is not 100%?

R: a) The basic reproductive number of an infection is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. The definition assumes that no other individuals are infected or immunized in the population. b) R0 values are estimated having a compartment model in mind, and the estimated values are dependent on the model used and values of other parameters. For SIR model, $R_0 = \frac{\beta\langle k\rangle}{\delta}$, such that if $R_0 < 1$ the outbreak will die out, and if $R_0 > 1$ the outbreak will expand (see description of the parameters above). If a fraction w of the average number of contacts <k> is vaccinated, we will get a new reproductive ratio given by $R_{vac} = \frac{\beta\langle k\rangle(1-w)}{\delta} = R_0(1-w)$, such that the disease will die out for $R_{vac} < 1 \Rightarrow w > w_c = 1 - \frac{1}{R_0}$. As an example, an R_0=3.0 would imply a critical proportion of vaccinated individuals of 66% to halt the disease. c) If only a fraction φ of the vaccines are effective, then $w_c = \left(1 - \frac{1}{R_0}\right) \Big/ \varphi$.

4. How would you compute the degree distribution of a graph? In case of a scale-free degree distributions, how would you plot it?

R: Plot the degree distribution d(k) - the probability distribution of the degrees over the whole network. The process starts by computing the number of number Nk of nodes with degree k. From Nk we calculate d(k)=Nk/N. Use log-log plots; use cumulative degree distributions; avoid linear binning (e.g., use logarithmic binning).

5. The Barabási–Albert (BA) model is likely the most famous algorithm for generating scale-free networks. **a)** Describe the BA algorithm. **b)** Using the BA-model, specify the input values you would adopt to have a network with $N=10^4$ nodes, a power-law degree distribution, and an average degree of $<k>=12$. **c)** Depart from the BA model to suggest a model and associated principles underlying the self-organization of networks with power-law degree distributions and prominent cutoffs for high degrees.

R: a) The Barabási–Albert (BA) algorithm combines growth and preferential attachment for generating scale-free networks. The network begins with an initial connected network of m_0 nodes. New nodes are added to the network one at a time. Each new node is connected to $m<m_0$ existing nodes with a probability that is proportional to the number of links that the existing nodes: the probability p_i that a node i with a degree k_i receives a link from the new node is given by $p_i = k_i / \sum k_j = k_i / (2E)$. *b) The average degree is given by 2m, such that, for having a SF network with <k>=12 one needs to adopt m=6. The algorithm stops when the desired number of vertices is achieved. c) Often nodes have a limited lifetime or a threshold above which cannot receive more ties. If we consider that existing nodes with a degree $k>k_{cutoff}$ cannot receive links from newly introduced nodes, than we would obtain a power-law degree distribution with prominent exponential cutoffs for high degrees.*

6. Scale-free networks are said to be extremely robust against random failures. Explain why is this the case[1]. Also, please explain what is the *Achilles' Heel* of these topologies.

R: One can analyze the robustness of a graph as an inverse percolation problem. It can be shown that the critical fraction of nodes f_c to be removed above which the network falls apart depends strongly on the second moment of the degree distribution $<k^2>$ — $f_C = 1-\left(\langle k^2\rangle / \langle k\rangle - 1\right)^{-1}$. *For scale-free networks with $\gamma < 3$, the second moment diverges in the $N \to \infty$ limit (i.e., for large networks). If we insert this limit in the previous expression, we find that $f_c = 1$ for large networks. This means that, for very large networks, to fragment a scale-free network we must remove all of its nodes. In other words, the random removal of a finite fraction of its nodes does not break apart a large scale-free network. The Achilles' Heel of SF networks relies on its low resilience against targeted attacks. Let us remove first the highest degree node, followed by the next highest degree, and so on. In this case, it is sufficient to remove only a few hubs to break a scale-free network into disconnected components, as the 2nd moment of the degree distribution no longer diverges for large N.*

7. The *Prisoner's dilemma* is one of the most famous dilemmas of cooperation. In its simplest form, it considers symmetric pairwise one-shot interactions involving costs and benefits. Cooperation (*C*) corresponds to offering the other player a benefit *b* at a personal cost *c*, with *b*>*c*. Defection (*D*) means offering nothing. Consider a well-mixed population of individuals that interact following this game. **a)** Indicate the payoff matrix associated with this setting and find the evolutionary stable strategy/strategies of this dilemma. **b)** Let us consider that there is a punishment $K > c$ of those opting for defection against a cooperator. Suggest a payoff matrix associated with this setting. Are we still playing a Prisoner's dilemma? Indicate the evolutionary stable strategy/strategies of this new dilemma with punishment[1].

R: To understand the dilemma at stake let us write down the payoff matrix:

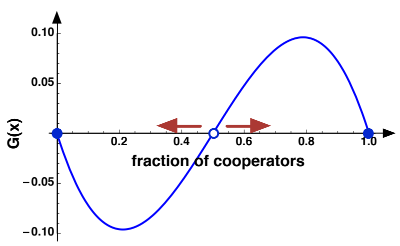
	C	D
C	b-c	-c
D	b-K	0

In a) K=0. In b) K>c. An Evolutionary Stable strategy ESS is a strategy that, if adopted by a population in a given environment, cannot be invaded by any alternative strategy that is initially rare. For K=0 (a), since a rare mutant choosing C will be always disadvantageous with respect to a resident population of Ds, then D will be the only ESS. However, for K>c, the same reasoning applies to resident populations of Cs. Thus, both C and D are evolutionary stable strategies in the presence of punishment. In other words, punishment transforms a prisoner's dilemma into a stag-hunt game, where cooperation may thrive.

You may also wish to show this formally assuming the replicator equation.

$$G(x) = \dot{x} = \frac{dx}{dt} = x(1-x)\left(f_C(x) - f_D(x)\right),$$

where x is the the fraction of Cs in the population. The average fitness of Cs and Ds is respectively given by $f_C(x) = x(b-c) + (1-x)(-c)$ and $f_D(x) = x(b-K)$. *There's one internal fixed point for* $f_C(x) - f_D(x) = 0 \Rightarrow x^* = c/K$. *Since,* $f_C(x) - f_D(x)$ *is a monotonous increasing function,* x^* *must be unstable. Please find on your right an illustration of the gradient G(x) for b=1.5, c=1 and K=2.*



---

[1] Showing a formal understanding of the problem worths 30%.

8. Please indicate whether each of the following statements is TRUE or FALSE.
Note: For each wrong answer we discount a correct one. The minimum mark is 0.

a) The epidemic threshold of a disease (e.g., in the SIS model) increases with the second moment of the degree distribution of the network of contacts.

b) Degree assortativity is the preference of network's nodes to attach to other nodes that have neighbors in common.

c) Modularity measures the quality of a given partition, by computing the difference between the network's real wiring diagram and the expected number of links within partitions if the network is randomly wired.

d) The friendship paradox is the phenomenon that states that most people have fewer friends than their friends have, on average.

e) The eigenvector centrality of a node is a measure number of shortest paths that pass through the node. It represents the degree of which nodes stand between each other.

R: a: False; b: False; c: True; d: True; e: False