INSTITUTO SUPERIOR TÉCNICO | UNIVERSIDADE DE LISBOA
NETWORK SCIENCE 2021/22
Problem Set 3

For our 3rd set of exercises we propose one of two options: i) solve the set of exercises proposed below; ii) leave these exercises for later, and use this class to work on your project.

1. (SF networks) Briefly explain the origin of the term scale-free when applied to complex networks.

2. (SF networks) Please detail why is it difficult to find real networks with power-law degree distributions ($P(k) \sim k^{-\gamma}$) with i) $\gamma<2$ and ii) $\gamma>>3$.

3. Why are scale-free networks referred as *ultra-small worlds*?

4. (BA model) Write down a simple computer program (or resort to one of the tools we have presented in previous labs), to generate a network with $N = 10^4$ nodes using the Barabási-Albert (BA) model with m=4. Use as initial condition a fully connected network with m = 4 nodes.
   a) Compute the average degree in time. Does it correspond to what you would expect theoretically?
   b) Measure the degree distribution at intermediate steps, namely when the network has 10^2, 10^3 and 10^4 nodes.
   c) Compare the distributions at these intermediate steps by plotting them together and fitting each to a power-law with degree exponent γ.
   d) Measure the average clustering coefficient as a function of N.
   e) Follow the 5th node added to your network and plot the evolution its degree in time. Can you explain this degree evolution?
   f) Repeat (e) replacing preferential attachment by uniform attachment. Can you explain the degree evolution?

5. (BA model) Compare the values of clustering coefficient and APL obtained from the Barabási-Albert (BA)model with those observed in real social networks (check M. J, Newman review paper available at the course webpage). Which growth model would you consider to generate a scale-free network with the same degree distribution as of the BA model, yet with a large clustering coefficient?

6. (SF models) Suggest an adaptation of the Barabási-Albert model to generate a bipartide scale-free network of size $N = 10^4$. How would you compare the i) degree distribution, ii) average degree and iii) clustering coefficient of a normal and a bipartide scale-free network or the same size?

7. Propose a model capable of generating a scale-free network with the same degree distribution as the BA model, yet with a given clustering coefficient *C*.

8. The coexistence of growth and preferential attachment in the Barabási-Albert model is at the core of the model. Are they both necessary for the emergence of the scale-free property?

9. Suggest a network model or principle able to generate a power-law degree distribution with saturation for low degrees.

Solutions

1. To best understand the meaning of the scale-free term, we need to familiarize ourselves with the moments of the degree distribution: $\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n P_k \approx \int_{k_{\min}}^{\infty} k^n P_k \, dk$. We expect that the lower moments are well-known to you. The first moment (n=1) refers the average of a distribution P_k. The second moment (n=2) helps us to compute the variance of the distribution, which tells us how spread k (the degree in our case) is. The third moment determines the *skewness* of a distribution, telling us how symmetric is P_k around the average.
   For P_k given by a scale-free degree distribution — $P_k = \left[(\gamma-1)k_{\min}^{\gamma-1}\right]k^{-\gamma}$ — we get $\langle k^n \rangle \approx C\frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1}$, and an expected k_{max} given by $\int_{k_{\max}}^{\infty} P_k \, dk < \frac{1}{N}$, i.e., $k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$. This means that the moments of the degree distribution, for very large N, become dominated by the term $N^{\frac{n-\gamma+1}{\gamma-1}}$. In other words, for $2 < \gamma \leq 3$: <k> is finite, <k^2> diverges. For $\gamma > 3$: <k> and <k^2> finite.
   <u>Now comes what is truly important</u>: Most power-law degree distributions rest in the $2 < \gamma \leq 3$ interval, such that the **first moment is finite but the second moment is infinite**… The divergence of <k^2> (and of σ_k) for large N indicates that **fluctuations around the average can be arbitrary large**. In other words, when we randomly choose a node, we do not know what to expect: The selected node's degree could be tiny or arbitrarily large. Hence networks with γ<3 do not have a meaningful internal scale, they are "scale-free"[1].
2. Let us return back to the expected maximum degree of a scale-free network of size N, given by $k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$. For γ<2, k_{max} grows faster than N, which turns to be impossible if one assumes, as it is the case, that I cannot have two links connecting the same pair of nodes. Now, what's the problem with having γ>>3? For $\gamma > 3$ both the first and the second moments are finite. For all practical purposes the properties of a scale-free network in this regime are difficult to distinguish from the properties a random network of similar size. To document the presence of a power-law degree distribution we ideally need 2-3 orders of magnitude of scaling, which means that k_{max} should be at least 10^2 - 10^3 times larger than k_{min}. By inverting $k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$ we get $N = \left(\frac{k_{\max}}{k_{\min}}\right)^{\gamma-1}$, which can be used to estimate the network size necessary to observe the desired scaling regime. Let's say that we aim at finding a network with γ=5 and require scaling that spans at least two orders of magnitudes (e.g. $k_{min} \sim 1$ and $k_{max} \simeq 10^2$). From the previous equation we need N>10^8, which is pretty hard to find.
3. For 2<γ<3 the average path length of a scale-free network scales with $\ln \ln N$, a significantly slower growth than the $\ln N$ derived for random networks (at the time, referred as small-world networks). Networks in this regime are called "ultra-small", as the hubs radically reduce the path length, even for very large networks, when compared with random networks.
4. This exercise expects that you would do some coding or resort to any platform (netlogo, mathematica, some gephi plugin, etc) where the Barabási-Albert (BA) model is already implemented. Below we provide some answers based on what one would expect from simple analytical arguments.
   a. The average degree should be constant in time and given by <k>~2m=8.
   b. The degree distribution P_k should converge to $P_k \sim k^{-3}$.
   c. See (b).

[1] Check the slides of class 05 or the 3rd chapter of A. L. Barabási, Network Science (2016).

d. Theoretically you should get a smooth decay with the network size N, following $\langle C \rangle \sim \frac{(\ln N)^2}{N}$.

e. The rate at which an existing node i , with degree k_i, acquires links as a result of new nodes connecting to it is given by

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j},$$

where m is the number of links of each new node. In other words, $\Pi(k_i)$ is the probability of being selected by the new node and m the number of chances of being chosen. Since $\sum_{j=1}^{N-1} k_j = 2E \approx 2mt$, we have

$$\frac{dk_i}{dt} = m\frac{k_i}{2mt} = \frac{k_i}{2t}.$$

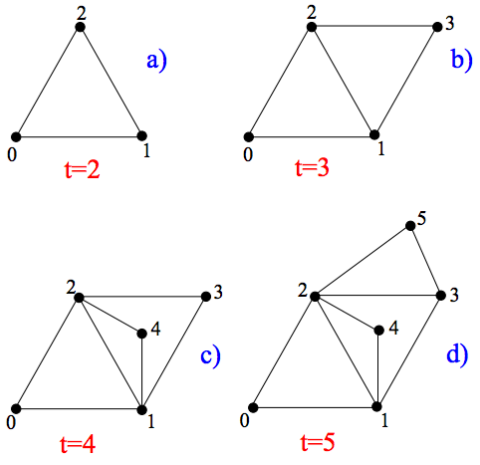
Solving this equation, we get $k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$ where t_i is the time where node i was added to the network. **Thus, your simulation is expected to show that the degree of a node *i* scales in time following**

$$k_i(t) \sim t^{1/2}.$$

In other words, the degree of each node increases following a power-law with the same dynamical exponent ½ for all nodes. The intuition why is sub-linear is simple: each new node has more nodes to link to than the previous node. Hence, with time the existing nodes compete for links with an increasing pool of other nodes.

f. For uniform attachment new nodes choose randomly the nodes they link to. This means that $\Pi(k_i) = \frac{1}{m_0 + t - 1} \sim \frac{1}{t}$ for large N=m$_0$+t. This leads to a degree dynamics given by $\frac{dk_i}{dt} = m\Pi(k_i) = \frac{m}{t}$. One can solve this equation to conclude that your simulations are expected to indicate that the degree of each node follows a logarithmic growth $k_i(t) \sim \ln t$ in time.

5. If you compute the clustering coefficient of a scale-free networks created through the BA model, you will notice that it would be several orders of magnitude below many of the real-world networks described in the literature. The minimal model (or the link selection model) proposed by Dorogovtsev and Mendes (PRE, 2001) solves this problem. As the BA model, at each time step we add a new node. Each new node selects a random link at random, and connects itself to the two ends of the chosen link. This procedure is repeated for each pair of links of the new node, creating a large number of triangular motifs. Contrary to the BA model, the model requires no knowledge about the overall network topology, hence it is inherently local and random. The principle behind the emergence of a scale-free network with the same exponent as the BA model is also simple. The higher is the degree of a node, the higher is the chance that it is located at the end of the chosen link. The fact that this dependence



is linear, gives rise to the same attachment probability $\Pi_i \sim \frac{k_i}{\sum_j k_j}$ as in the BA-model, and the same degree distribution.

6. One can easily adapt the BA model for bipartide networks. Add nodes sequentially as in the BA model. Each node will have one of two equally probable traits (e.g. a gender). You may also add a different trait in each time-step. When establishing the connections, each new node is connected to older nodes of the opposite trait (that's the trick) with a probability proportional to the degree of the older nodes. At the end, we will get a power-law degree distribution yet with a truncated maximum degree, as if the network would be half of its size: $k_{\max} = k_{\min}\left(\frac{N}{2}\right)^{\frac{1}{\gamma-1}}$.
7. Generate a SF network with the BA model. Repeatedly swap the ends of 2 randomly chosen vertices until the desired value of C is reached. In other words, you're optimizing the value of the clustering coeff. using a greedy algorithm such that clustering gets as close as possible to the desired value of the clustering coefficient.
8. YES! You may test two models, each containing only one of the two ingredients. A first model, that incorporates growth but lacks preferential attachment, generates a exponential degree distribution. A second model that lacks growth but incorporates preferential attachment can be built by starting with a fixed number of disconnected nodes. Then repeatedly add links among the N nodes. These links are added preferentially choosing high degree nodes as link destinations. Though the degree distribution early in the simulation looks scale-free, the distribution is not stable, and it eventually becomes nearly Gaussian as the network nears saturation. Thus, preferential attachment alone is not sufficient to produce a scale-free structure.
9. For instance, the growth + preferential attachment model with initial attractiveness generates saturation for small degrees.