Helsinki University of Technology Saramäki/Heimo
Laboratory of Computational Engineering

S-114.4150 Complex Networks, autumn 2006

Problem set #5 (due Thu. 7.12.)

1. In this exercise we study diffusion (or random walks) on undirected weighted networks. We start by imagining placing large number of random walkers onto the vertices of a network. These walkers are allowed, in each time step, to move along adjacent vertices along the edges connecting them. The edge, along which a walker chooses to move, is picked randomly with a probability proportional to the weight of the edge (*i.e.*, a walker moves from vertex i to j with probability $W_{ij}/\sum_k W_{ik}$). We assume that $s_i = \sum_k W_{ik} \geq 1, \quad \forall i$.

   a) Let us denote the fraction of walkers at vertex i at time t by $\rho_i(t)$. Derive the transfer matrix of the process, *i.e.*, the matrix T that satisfies
   $$\rho(t+1) = T\rho(t), \tag{1}$$
   where $\rho(t) = (\rho_1(t), \rho_2(t), \ldots)^T$.

   b) Show that T is diagonalizable. (Hint: Remember that it is enough to show that T is similar with a symmetric matrix. Why?)

   c) Show that $|\lambda_i| \leq 1$ for all i.

   d) Show that $\max_i \lambda_i = 1$.

2. Let G_i be time series of length T, $i = 1, 2, \ldots, N$ and let us define the correlation matrix of the time series by
   $$C_{ij} = \frac{\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle}{\sigma_i \sigma_j}, \tag{2}$$
   where $\sigma_i = \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$ and the angular brackets denote time average. Now we can construct a weighted network by defining
   $$W_{ij} = d(i,j) = \sqrt{2(1 - C_{ij})}. \tag{3}$$
   Show that W is a distance matrix, *i.e.*, that
   $$\begin{array}{ll} \text{(i)} & d(i,j) = 0 \quad \Leftrightarrow \quad G_i = G_j \\ \text{(ii)} & d(i,j) = d(j,i) \\ \text{(iii)} & d(i,k) \leq d(i,j) + d(j,k). \end{array} \tag{4}$$

3. In this exercise, we will think about the consequences of finite network size on observed degree distributions. Consider you have empirical data on a network of N vertices. Let us assume that (independent of your observations) you somehow know that this network has an underlying scale-free degree distribution $P(k) \propto k^{-\gamma}$ and the observed network is just a finite-size realization. Let us focus on the observed degree distribution.

   a) At some value $k = k_c$, your observed degree distribution $P'(k) = N(k)/N$ becomes discontinuous, meaning that for degrees smaller than this, you can always find at least one vertex. To get a better estimate of $P'(k)$ for $k > k_c$ the data is usually binned or shown as cumulative distribution.

      Let us approximate k_c by considering that in finite-size realizations of networks with $P(k) \propto k^{-\gamma}$ there is *on the average* at least one vertex for any $k \leq k_c$, and for $k > k_c$ *on the average* less than this. How does k_c behave as function of N? (You can assume that the prefactor of the distribution is $O(1)$, so forget it) If $\gamma = 3$, how large should your network be in order for the continuous part of the degree distribution to reach 100, ie $k_c = 100$?

   b) At $k > k_c$ the observed degree distribution $P'(k)$ is discontinuous. If the "real" distribution is again as above, what is the maximum degree k_{max} we observe? You can think e.g. that there is only one vertex with $k \geq k_{max}$.

   c) Consider now that the "real" degree distribution is scale-free only asymptotically, $P(k) \propto 1/[k(k+1)(k+2)]$. Nevertheless you take its tail and fit $P'(k) \propto k^{-3}$. At what degree does the resulting fractional error become less than 1%, i.e. $P(k)/P'(k) = 99/100$? (You may assume that the constant prefactors are the same for both distributions and so they cancel out). Given the above observations, what can you qualitatively say about the network size needed for getting an accurate estimate of the asymptotic power law exponent?

4. Suppose we have the following simple model of growing networks: we start with a small seed, e.g. a triangle with all weights equal to one. Then, at each time step, we pick an edge jk with probability proportional to its weight, $\pi_{jk} = w_{jk}/\sum w$. Then we connect a new vertex n to vertices j and k with links of weight $w_{nj} = w_{nk} = 1$, and increase w_{jk} by an amount δ. This process is then repeated.

a) Write down a rate equation for the evolution of the average weight of an edge which was "born" at $t = i$, $\partial \langle w(i,t) \rangle / \partial t$. Solve it for $\langle w(i,t) \rangle$. Here, i is just the index of an edge and marks the time $t = i$ when the edge was "born". You may assume that the total weight in the network at $t = 1$ is negligible, i.e. $\sum w \approx (2 + \delta)t$.

b) Assume that this "mean-field" equation is exact, i.e. there are no fluctiations. Then the weight distribution can be written as

$$P(w) = \frac{1}{2t} \int_{i=1}^{i=t} \delta \left(w - \langle w(t,i) \rangle \right) di$$

(There are 2 edges born at every time step). The same trick as with the degree distributions can be used here:

$$\begin{aligned} P(w) &= \frac{1}{2t} \int_{i=1}^{i=t} \delta \left(w - \langle w(i,t) \rangle \right) di \\ &= -\frac{1}{2t} \int_{w(i=t,t)}^{w(i=1,t)} \delta \left(w - \langle w(i,t) \rangle \right) \frac{di}{d \langle w(i,t) \rangle} d \langle w(i,t) \rangle \end{aligned}$$

which, due to how the delta function works, equals

$$P(w) \quad = \quad -\tfrac{1}{2t} \tfrac{di}{d\langle w(i,t) \rangle} \text{ evaluated at } \langle w(i,t) \rangle = w.$$

The weight distribution should be a power law - what is its exponent?

If in trouble, check out *Phys. Rev. Lett.* **92**, 228701 (2004) for a similar kind of derivation with the BBV model.