ECS 253 / MAE 253, Network Theory and Applications
Spring 2023
Advanced Problem Set # 1, Due April 19
Topic: Kinetic theory, and the Erdős-Rényi random graph

Problem 1: The Erdős-Rényi random graph – analyzing the phase transition

Consider an Erdős-Rényi random graph with N nodes and probability p for any edge to be present. Let N_G denote the number of nodes that are in the giant component. Thus, the fraction of nodes that are not in the giant component, $u = 1 - N_G/N$. Likewise, the probability that a node chosen uniformly at random is not in the giant component is u.

a) We first analyze the likelihood that an arbitrary node i is not in the giant component via its connection to another node node j. There are two possibilities that lead to the desired outcome: (i) i is not linked to j, (ii) i is linked to j, but j is not part of a giant component. Considering these facts and that there are $N-1$ possible choices for j show that:

$$u = (1 - p + pu)^{N-1}. \tag{1}$$

b) The average degree, $\langle k \rangle = p(N-1)$. Using $p = \langle k \rangle / (N-1)$ and the fact that $\ln(1+x) \approx x$ for small x show that:

$$\ln u = -\langle k \rangle (1-u). \tag{2}$$

c) Let $S = 1 - u$ denote the fraction of nodes in the giant component and show that the result in (b) leads to the equation:

$$S = 1 - e^{-\langle k \rangle S}. \tag{3}$$

d) Although Eq. (3) looks simple, it does not have a closed form solution. The easiest method to solve it is graphical. Plot the right hand side of Eq. (3) as a function of S for three choices of average degree $\langle k \rangle = 0.5, 1, 1.5$. Now plot also on the same figure the line $S = S$. Where the curve and the line intersect is where there are valid solutions to Eq. (3).

e) In fact the smallest value of $\langle k \rangle$ that leads to a non-zero solution for S (the critical level of connectivity for the emergence of a giant component) is when the derivative of the r.h.s. of Eq. (3) w.r.t S equals the derivative of the l.h.s. of Eq. (3) w.r.t S for $S = 0$. Using this, show that the giant component first emerges for $k_c = \langle k \rangle = 1$.

f) Now consider another interesting aspect, the value of p where we first achieve full connectivity and all nodes are in the giant component. The probability that a node selected uniformly at random does not have an edge into the giant component is $(1-p)^{N_G}$, and we sill consider the regime $N_G \approx N$. Show that the expected number of isolated nodes $I_G \approx Ne^{-NP}$. Then by using that formula and setting $I_G = 1$ calculate the value of p for the onset of full connectivity.

Problem 2: The Erdős-Rényi random graph – cluster size distribution

Here you will do some simple analysis of the Erdős-Rényi random graph evolution using kinetic theory. We model the growth process as cluster aggregation via the classic Smoluchowski coagulation equation. The following two references are classics:

- David J. Aldous, "Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists", *Bernoulli*, Vol 5 (1), 3-48, 1999.
- E. Ben-Naim and P. L. Krapivsky, "Kinetic theory of random graphs: From paths to cycles", *Phys. Rev. E* 71, 026129, 2005.

– Let $N_k(t)$ denote the total number of components of size k and time t.
– Let $c_k(t) = N_k(t)/N$ denote the density of components containing k nodes at t.
– We begin at $t = 0$ with $c_1(0) = 1$ (and thus $c_j(0) = 0$ for $j \neq 1$).
– We will drop the time subscript for simplicity, $c_k(t) \equiv c_k$, and analyze $\frac{dc_k}{dt}$. This approximates the impact of adding one edge as a continuous process and is the resulting average graph / "mean field" over all graphs (see Aldous 1999, for more details.)

a) The probability that a node chosen uniformly at random belongs to a component of size i is ic_i. Using this fact, write out the evolution equation for $\frac{dc_k}{dt}$. (You will have to consider all the ways that a new component of size k can be formed by adding one edge, and that the number of components of size k can decrease.)

b) By simple iteration, solve c_1, c_2 and c_3.

c) What formula does this suggest for the general process, c_k?

d) Note, to solve for c_k explicitly starting from the formula for $\frac{dc_k}{dt}$ requires generating functions and a clever Lagrange inversion formula (see Ben-Naim 2005 for details). The real formula is $c_k = \frac{k^{k-2}}{k!} t^{k-1} e^{-kt}$, but the one you find in part (c) is close. Using this explicit formula for c_k show that at the critical point (t=1) the density of components $c_k \sim k^{-5/2}$.