Data driven social analytics - Homework2

Ana Mestre

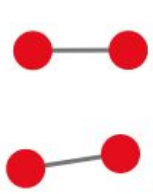
(1) What is the degree distribution for G(n, p)?

The distribution of the degree of any particular vertex is binomial.

(2) How can we keep average degree constant as $n \to \infty$?

To keep the average degree constant as the number of nodes grows, p(n) should be proportional to $\frac{1}{n}$.

(3) Are not those properties the same? (No isolated nodes vs connected network)



No because in the example on the left, the network is not connected (there are two separate components) and there are not isolated nodes. This example doesn't have the two different properties. Therefore, they are not the same.

(4) Are the previous three properties monotone?

Yes, they are monotone.

(5) Can you think about a non-monotone property?

Having an even number of links.

(6) When do we expect one isolated node (on average)?

That is where $e^{-(n-1)p} = \frac{1}{n}$. Solving this yields p(n - 1) = log(n), or right at the point where average degree (n - 1)p is log(n).

(7) Do ER random graphs have high or low clustering coefficient?

They have low clustering coefficient.

(8) What is the probability for an existing node to get a new link in step t?

An existing node has a probability m/t of getting a new link in step t.

(9) How many nodes have expected degree d less than 35 at time t = 100?

$d = m(1 + log(\frac{t}{i})) \to 35 = m(1 + log(\frac{100}{i}))$. We can solve that as a function of m.

(10) What is the solution of this equation?

$d_i = m + m \cdot log\left(\frac{t}{i}\right)$

(11) How many nodes have expected degree d less than 35 at time t = 100(m = 20)?

$$d = m\sqrt{\frac{t}{i}} \rightarrow 35 = 20\sqrt{\frac{100}{i}} \rightarrow i = 32.65$$

(12) Degree distributions of preferential attachment growth. Why 3?

The density function (derivative) is a power-law

$$f_t(d) = \frac{dF_t(d)}{dd} = 2m^2/d^3$$

We can find that the number 3 appears from analytical calculus from the previous formula.

(13) Why? [hint: look for the "friendship paradox"]

The second part captures the preferential attachment mechanism because higher degree nodes are more likely to be neighbours of nodes chosen in 1st part.