PS 2

1

1: Erdos - Renyi.

The probability of a <u>specific</u> simple graph G with M edges is

$$P(G) = p^M (1-p)^{C_2 - M},$$

where C_2 is the binomial coeff. $C_2 = \binom{N}{2}$

1a The total probability of finding a graph with M edges will be

$$p(M) = \#\text{possible graphs with } M \text{ edges} \cdot P(\text{specific simple graph})$$

$$= \binom{C_2}{M} \cdot P(G) = \binom{C_2}{M} p^M (1-p)^{C_2 - M}$$

Now $\langle M \rangle = \sum_M^{\binom{N}{2}} M \cdot p(M)$

$$\Rightarrow \langle M \rangle = \sum_M^{\binom{N}{2}} M \binom{C_2}{M} p^M (1-p)^{C_2 - M} =$$

$$= \sum_M M \frac{C_2!}{M!(C_2 - M)!} p^M (1-p)^{C_2 - M} =$$

$$= \sum_M \frac{C_2!}{(M-1)!(C_2 - M)!} p^M (1-p)^{C_2 - M} =$$

$$= C_2 \cdot p \sum_M \frac{(C_2 - 1)!}{(M-1)!(C_2 - M)!} p^{M-1} (1-p)^{C_2 - M} =$$

$$= C_2 p \sum_M \binom{C_2 - 1}{M - 1} p^{M-1} (1-p)^{C_2 - M} = \cancel{C_2 p}$$

Note that $\sum_{K=0}^{n} \binom{n}{K} x^K y^{n-K} = (x+y)^n$ (Newton's binomial) (2)

$$\ldots = C_2 \cdot p \cdot (p + 1 - p)^{C_2 - 1} = C_2 \, p.$$

Note: We could have found the same results by considering each link as a Bernoulli variable

b).

Each node can make $N-1$ possible connections with probability p

$\Rightarrow$ mean degree $= (N-1) \cdot p$

c) Show, under appropriate assumptions, that the degree distribution for an ER graph satisfies

$$p_K \simeq e^{-c} \frac{c^K}{K!} \quad N \to +\infty \quad , \text{ where } c = (N-1)p$$

Let's focus on a specific node. The prob. that it will be connected to some particular K other nodes is

$$p^K (1-p)^{N-1-K}$$

There are $\binom{N-1}{K}$ ways to pick the other K node, so the probability of having K neighbors is

$$p_K = \binom{N-1}{K} p^K (1-p)^{N-1-K} \quad = \text{binomial distribution}$$

We assume that $\langle K \rangle$ remains constant when $N \to +\infty$: $(N-1)p = c \Rightarrow p = \frac{c}{N-1} \to 0$

We can write

$$\ln\left[(1-p)^{N-1-K}\right] = (N-1-K)\ln\left(1-\frac{c}{N-1}\right) \sim (N-1-K)\cdot\left(-\frac{c}{N-1}\right) \sim -c$$

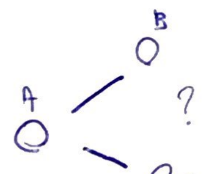
$$\binom{N-1}{K} = \frac{(N-1)!}{(N-1-K)!K!} \approx \frac{(N-1)^K}{K!}$$

$$\Rightarrow p_K = \frac{(N-1)^K}{K!}p^K e^{-c} = \frac{(N-1)^K}{K!}\left(\frac{c}{N-1}\right)^K e^{-c} = \frac{c^K}{K!}e^{-c}$$

The famous Poisson distribution

d) Numerical

e) The ER graph is homogeneous (each node is equal to the other), so we expect ~~on average~~ that over the ensemble the global clustering coef. will be equal to the local clustering coeff.



Focusing on a specific node A, and assuming it is connected to nodes B and C, what is the expect. value of its clustering coeff.?

$C_A = 1\cdot p + 0\cdot(1-p) = p \Rightarrow p$ is ER's expected clust. coeff.

See notebook for numerical verification.

f) Show that the diameter of an ER graph is $B + \frac{\ln N}{\ln c}$ as $N \to +\infty$, where B is a const. and $c = (N-1)p$.

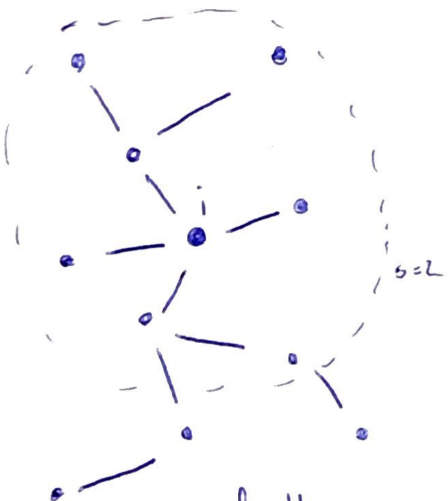
Crude solution

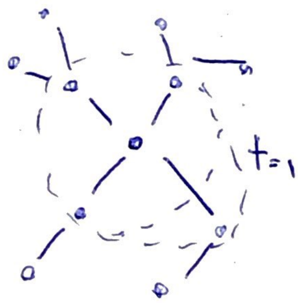
The exp. number of nodes at distance s from a given node is equal to c^s. The diameter will be the value of s s.t.

$$c^d = N \Rightarrow d \ln c = \ln N \Rightarrow d = \frac{\ln N}{\ln c}$$

Less Crude

Consider two starting nodes i and j. Consider balls around the nodes that have, respectively, grown s steps from i and t steps from j.





We will have picked up c^s and c^t nodes respectively, and we want to be in the regime where

$$c^s \ll N, \quad c^t \ll N, \quad N \to +\infty$$

Now if there is an edge between the surfaces, then there must be an edge between neighborhoods of any larger ball. Conversely, if there is no edge, there will be no edge between smaller balls, and the the shortest path between i and j must have a length $> l = s+t+1$.

$\Rightarrow$ If there isn't an edge between the balls, $d_{ij} > l \Rightarrow P(\text{no edge between balls}) = P(d_{ij} > l)$

Considering all nodes on each surface, we have $c^s \cdot c^t$ possible links

$$\Rightarrow P(d_{ij} > l) = (1-p)^{c^{l-1}} = \left(1 - \frac{c}{N-1}\right)^{l-1}$$

Taking the log

$$\ln\left[P(d_{ij} > l)\right] = c^{(l-1)} \ln\left(1 - \frac{c}{N-1}\right) \sim \cancel{(l-1)}\, c^{(l-1)} \cdot \left(\cancel{1} - \frac{c}{N}\right) \sim \frac{c^l}{N}$$

The diameter is the smallest value of ℓ for which $P(d_{ij} > \ell) = 0$. The probability $P(d_{ij} > \ell) \to 0$ only if c^ℓ grows faster than N, so we write

5

$$c^\ell = b N^{1+\varepsilon}$$

$$\Rightarrow \ell = \frac{\ln b}{\ln c} + \lim_{\varepsilon \to 0^+} \left[\frac{(1+\varepsilon)\ln N}{\ln c} \right] = B + \frac{\ln N}{\ln c}$$

2) Configuration model

At each time step, the probability of the node edge $\underline{\text{not}}$ being chosen is

$$q_1 = \underbrace{\left(\frac{M-1}{M}\right)}_{\text{not chosen as first edge}} \cdot \underbrace{\left(\frac{M-2}{M}\right)}_{\text{not chosen as second edge}} = \frac{M-2}{M}$$

$\Rightarrow$ For K steps

$$q = \left(\frac{M-2}{M}\right)^K$$

If we want $q < \frac{1}{10M}$

$$\Rightarrow \left(\frac{M-2}{M}\right)^K < \frac{1}{10M} \Rightarrow K \ln\left(1 - \frac{2}{M}\right) < -\ln(10M)$$

$$\overset{\approx}{\Rightarrow} -\frac{2K}{M} < -\ln(10M) \Rightarrow K > \frac{2\ln(10M)}{M}$$

b) see notebook.

3 a

Consider a random network where an edge exists between v_i and v_j with an independent probability equal to $f(|i-j|) \equiv \alpha \lambda^{|i-j|-1}$, $\alpha > 0$; $\lambda \leq 1$

6

How many edges on the right? Let's focus on node 0,

$$E[\text{\# edges on the right}] = \sum_{j=1}^{+\infty} P(\text{link } 0\text{-}j) = \sum_{j=1}^{+\infty} \alpha \lambda^{j-1} =$$

$$= \alpha \sum_{k=0}^{\infty} \lambda^k = \frac{\alpha}{1-\lambda}$$

The problem is symmetrical, so we expect to have the same number of nodes on the left and on the right

$$\Rightarrow \text{Expected degree} = \frac{2\alpha}{1-\lambda}$$

4 |

- Let A be the $n \times n$ adjacency matrix for an undirected ~~matrix~~ network, and $D = \operatorname{diag}(A^2)$.
- A walk on this matrix is backtracking if it contains a sequence "uvu"; if it doesn't it's non-backtracking.
- We want to compute the NB~~XX~~TW centrality.
- We define the matrix polynomial $M(\alpha) = I - \alpha A + \alpha^2 (D - I)$, and claim that Q is the non-backtracking version of the centrality matrix,

$$Q = (1-\alpha^2) M(\alpha)^{-1},$$

where α is small enough so that the spectrum of $M(\alpha)$ is > 0

- Let P_n be the matrix s.t. its i,j th term counts the total number of non-backtracking walks of length K from i to j within the network.

- Show that we must have

$$P_0 = I, \quad P_1 = A, \quad P_2 = A^2 - D, \quad \text{and} \quad P_r = AP_{r-1} - (D-I)P_{r-2}, \; r \geq 3$$

AP_{r-1} ~~...~~ is a matrix s.t.

$(AP_{r-1})_{ij}$ = number of paths from i to j that are non-backtracking up to the $r-1$ th step.

$\Rightarrow$ ~~...~~ AP_{r-1} includes some backtracking walks.

For each NBTW between i and j of length $r-2$, counted in P_{r-2}, starting from vertex i and going through a 3rd vertex K, AP_{r-1} will count the walks like

$v_i v_l v_i v_k \dots v_j$.

We can correct this by subtracting the $-(D-I)P_{r-2}$ term.

- Solve the above difference eq. for Q. Show that

$$Q = (1-\alpha^2)\left(I - \alpha A + \alpha^2(D-I)\right)^{-1}$$

Remember $Q = \sum_r \alpha^r P_r$.

We start from

$$P_r = AP_{r-1} - (D-I)P_{r-2}$$

Multiplying by α^r

$$\alpha^r P_r = \alpha A \alpha^{r-1} P_{r-1} - \alpha^2 (D-I)\alpha^{r-2} P_{r-2}, \quad r \geq 3$$

Summing over $r \geq 3$ we have

$$Q - \alpha^2 P_2 - \alpha P_1 - P_0 = \alpha A\left(Q - \alpha P_1 - P_0\right) - \alpha^2 (D-I)(Q-P_0)$$

$$\Rightarrow \left(I - \alpha A + \alpha^2(D-I)\right) Q = \alpha^2 P_2 + \alpha P_1 + P_0 - \alpha A(\alpha P_1 + P_0) + \alpha^2 (D-I) P_0$$

$$= \dots \Rightarrow Q = (1-\alpha^2)\left(I - \alpha A + \alpha^2(D-I)\right)^{-1}$$