Configuration model

SYSM 6302

CLASS 10

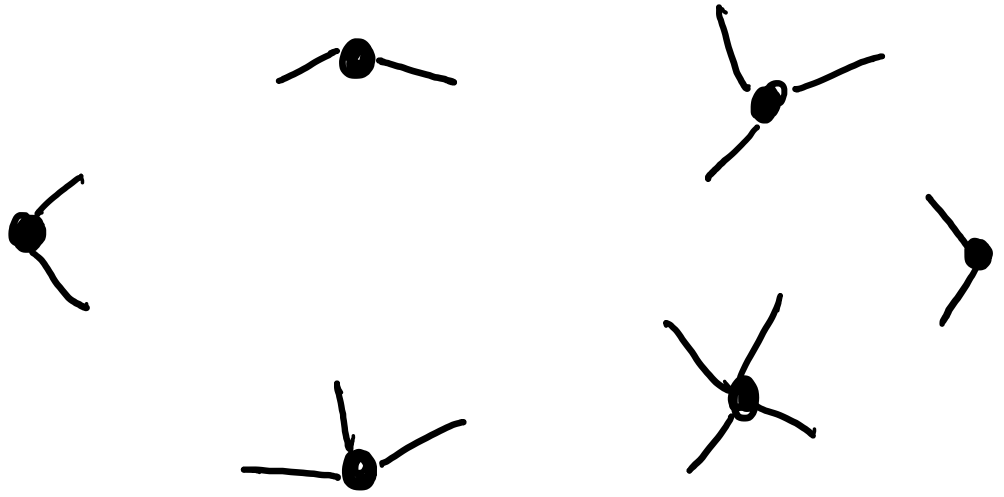
Configuration Model

→ Adapt the idea of random graphs to mimic realistic degree distributions

→ Still maintain some of the analytic tractability of the Erdos-Renyi random network.

→ Takes as input the desired degree sequence.

4, 3, 3, 2, 2, 2



① Degree sequence $\{k_i\}_{i=1}^{n}$

② Add n nodes and give the i^{th} node k_i "stubs" of edges

③ Pick two stubs uniformly randomly & connect them as an edge

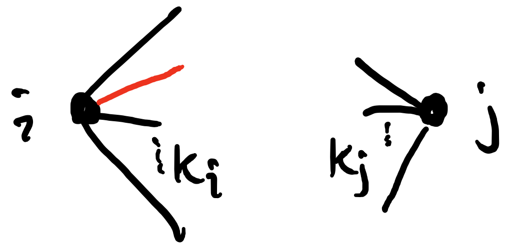
→ Like in ER the uniform probability is the property that provides analytic tractability

→ Note that the # of stubs must be even, i.e., k_i such that $\sum_{i=1}^{n} k_i = 2m$

i.e., degree sequence is graphical

must be even

What is the probability that two vertices are connected?



k_i stubs of i $\times$ $\dfrac{k_j \text{ stubs on } j}{2m-1 \text{ possible stubs}}$ $\Rightarrow$ $p_{ij} \approx \dfrac{k_i k_j}{2m}$

$\uparrow$ aside from the one attached to i

$\uparrow$ in the large m limit

→ Notice that self-edges and multi-edges are allowed. How many?

↳ i.e., the probability of having two edges between a pair of nodes:

$$\text{for } i \neq j \quad \frac{k_i k_j}{2m} \times \frac{(k_i - 1)(k_j - 1)}{2m} \Rightarrow \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{k_i k_j (k_i - 1)(k_j - 1)}{(2m)^2} = \frac{1}{2 \langle k \rangle^2 n^2} \underbrace{\sum_i k_i (k_i - 1)}_{\sum_i (k_i^2 - k_i)} \sum_j k_j (k_j - 1)$$

$\langle k \rangle = \frac{1}{n} \sum_i k_i$

$$= \frac{1}{2} \left[\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right]^2$$

$\langle k^2 \rangle = \frac{1}{n} \sum_i k_i^2$

Self-edges?

$$p_{ii} = \frac{\binom{k_i}{2}}{2m} = \frac{\frac{k_i(k_i-1)}{2}}{2m} = \frac{k_i(k_i-1)}{4m}$$

$$\sum_i p_{ii} = \sum_i \frac{k_i(k_i-1)}{4m} = \frac{\langle k^2\rangle - \langle k\rangle}{2\langle k\rangle}$$

If $\langle k^2\rangle = \frac{1}{n}\sum_i k_i^2$ (the second moment of the degree distribution) is constant or finite — which occurs in most cases — the quantities of self-edges and multiedges is fixed and in fact decays as $n \to \infty$

showing the details of this involves more probability than we use in this course.

Another tractable question: what is the expected number of neighbors a node's neighbor has?

(how many Friends do your friends have?)

→ what is the probability that a neighbor has degree k?

$$\underbrace{\frac{k}{2m-1}}_{\text{probability that an edge will end at a specific node with degree } k} \times \underbrace{np_k}_{\text{number of nodes with degree } k} \approx \frac{k}{2m} \cdot np_k = \frac{k}{n\langle k\rangle} np_k = \frac{kp_k}{\langle k\rangle}$$

probability of an edge attaching to any vertex with degree k.

$$\text{average degree of a neighbor} = \sum_k k\left(\frac{kp_k}{\langle k\rangle}\right) = \frac{\langle k^2\rangle}{\langle k\rangle}$$

Friendship Paradox – "$\underbrace{\text{your friends have more friends}}_{\frac{\langle k^2\rangle}{\langle k\rangle}}$ than $\underbrace{\text{you do}}_{=\langle k\rangle}$"

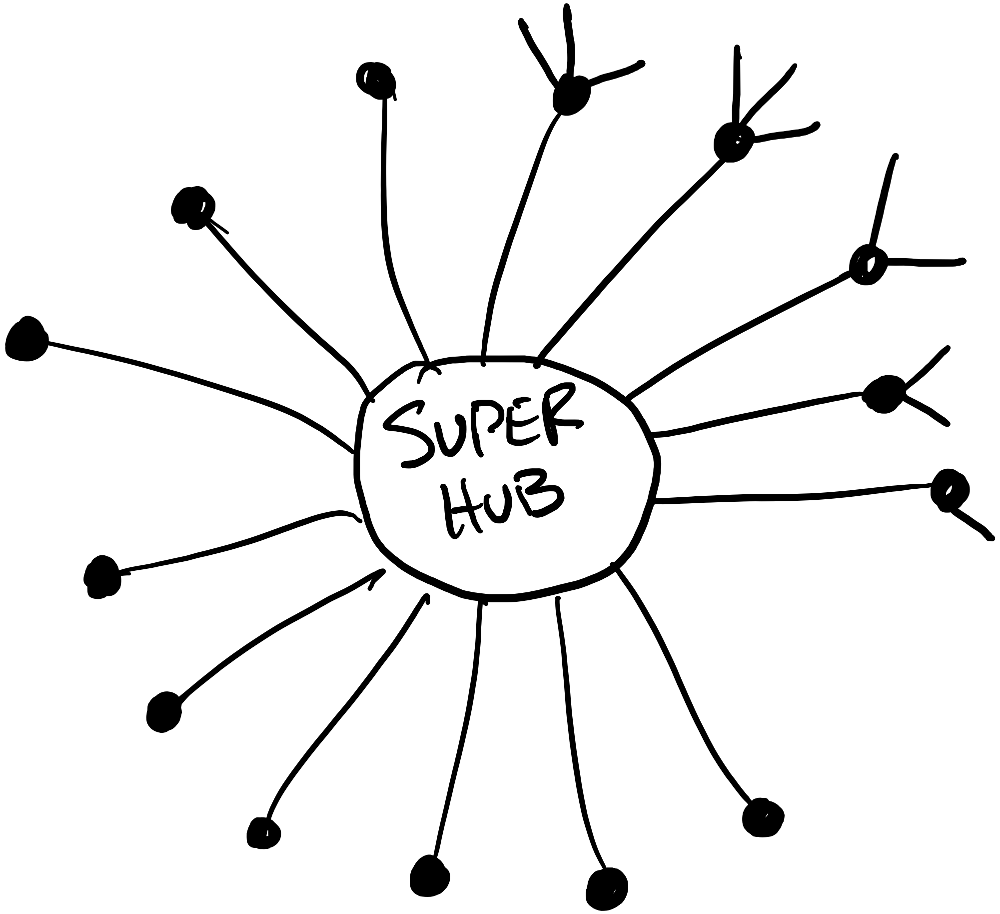
$$\frac{\langle k^2\rangle}{\langle k\rangle} - \langle k\rangle = \underbrace{\frac{1}{\langle k\rangle}}_{>0}\underbrace{\left(\langle k^2\rangle - \langle k\rangle^2\right)}_{\text{variance of the degree distribution}} > 0$$

> 0 assuming the graph has any edges

variance of the degree distribution ← = square of standard deviation

> 0 as long as all degrees are not the same

$$\Rightarrow \quad \frac{\langle k^2\rangle}{\langle k\rangle} > \langle k\rangle$$



high degree nodes are counted more frequently (overrepresented) in the friends-of-friends calculation because they have many edges

Similarly, nodes with low degree e.g., zero degree are underrepresented in the calculation

Clustering

$$C = \frac{1}{n} \frac{\left[\langle k^2 \rangle - \langle k \rangle\right]^2}{\langle k \rangle^3}$$

(we won't prove this)

→ Does not, in general, lead to significant clustering.