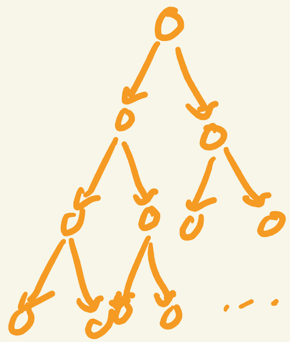
Week 3   Sept 9, 2020

Random Graph Models.

Are patterns that I observe "interesting?"



Two classes of models:

1. Mechanistic
2. Generative

Hint: there is overlap

Hypothesis about the mechanism that formed a network. A set of mathematical rules that can be applied to update, grow, or otherwise create a network.

Generative models are "weaker" recipes for network formation.

AKA

Stochastic Generative Models

↑

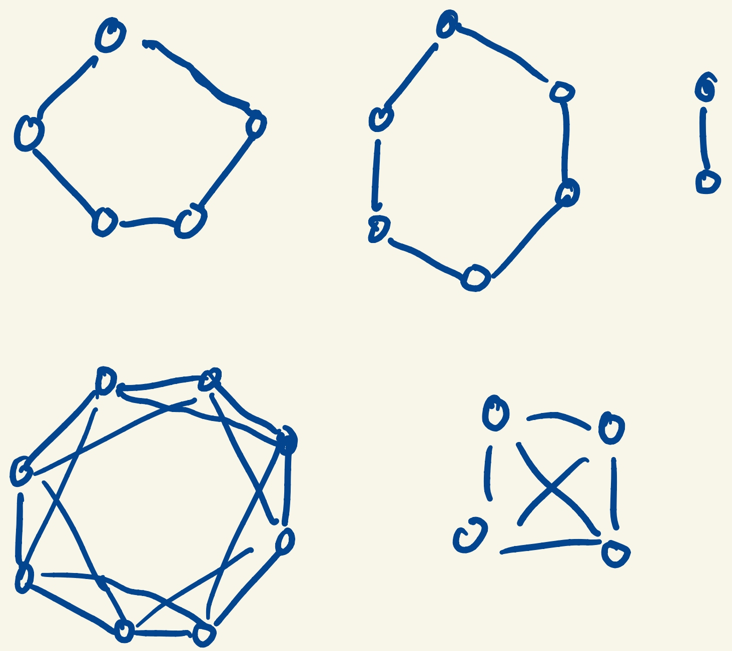
Randomness!

Recipe:

1. choose parameters
2. make the network.

Ex: (no randomness)

Ring networks



Parameters?

nodes (size of ring) n

edges per node k, even

$n > k$

Stochastic Generative Models are also recipes!

Choose params, draw a network.

1 model and parameters

$\downarrow$

set of outcomes.

Random graph model (generative models) are sets of networks, in which each element of the set occurs with some probability.

$\boxed{\text{Gen. Model = distribution over networks.}}$

Why do we like Gen. Models?

① easy to compute or estimate properties!

$$\langle x \rangle = \sum_G x(G) \Pr(G)$$

Prob of observing graph G, under the model.

typical value of x under a model.

exact value of x in a member G of the model's set

Hint: this is a weighted avg.
OR: expectation taken over a probability distribution.
↓
G.M.

Imagine if in our real network data we have some x_{data}, and we have a hypothesized model, with $\langle x \rangle_{model}$, we can compare x_{data}, $\langle x \rangle_{model}$ to see whether our data could have plaus

ER Random Graphs

$G(n,p)$

no self loops

undirected

binary ⇒ multigraphs $\emptyset$, weighted $\emptyset$

$$\forall\, i>j \qquad A_{ij} = A_{ji} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

vs $G(n,m)$

\# nodes

exact # of edges

NB: the total # of edges in $G(n,p)$ is not fixed.

Expected # of edges?

Recall that if I flip N coins, each coming up heads IID w.p. p, then $E[\text{Heads}] = N \cdot p$.

$$\langle m \rangle = \binom{n}{2} p = \frac{n(n-1)}{2} p$$

What is the Prob that I draw a graph with exactly m edges?

$$\Pr(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$$

binomial distribution over edge counts in the ER graph ensemble

Mean degree?

In a network w/ m edges and n nodes, $\langle k \rangle = \frac{2m}{n}$

$$\langle k \rangle = \frac{2m}{n}$$

random variable under $G(n,p)$!

$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} \times Pr(m)$$

$$= \frac{2}{n} \binom{n}{2} p$$

$$= \frac{\cancel{2}}{\cancel{n}} \frac{\cancel{n}(n-1)}{\cancel{2}} p$$

$$= (n-1)p.$$

Mean degree, under $G(n,p)$ in $= (n-1)p$.

Intuitive?

Each node can connect to up to $n-1$ possible neighbors ... but does so IID w.p. p.

"n-1 coins, w.p. p" $\rightarrow E[k] = (n-1)p$

Degree Distribution?

$$Pr(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$\Rightarrow$ degree distribution of an ER graph is binomial.

MND:

sum of degs of nbrs of u $= \sum_{v=1}^{n} k_v A_{uv}$

$$\frac{1}{2m}\sum_{u=1}^{n}\left(\sum_{v=1}^{n} k_v A_{uv}\right) = MND$$

Hint: swap order of summation

$$\frac{1}{2m}\sum_{v=1}^{n}\sum_{u=1}^{n} k_v A_{uv}$$

$$\frac{1}{2m}\sum_{v=1}^{n} k_v \underbrace{\sum_{u=1} A_{uv}}_{k_v}$$

$$\frac{1}{2m}\sum_{v=1}^{n} k_v^2 \ldots$$

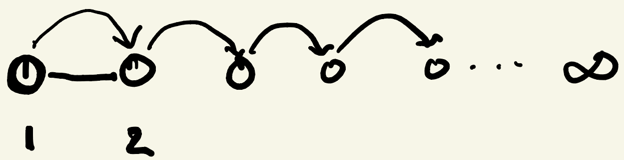
Goal: $\langle k^2 \rangle$, $\langle k \rangle$

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

Induction for #2?

Try induction on d.

1. show n=1 ✓
2. show n=2 ✓
3. show that, if I assume n case is true, that this implies n+1 case is true.



What if p is v. small?

$$\Pr(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$p = \frac{c}{n-1}$, $c = \Theta(1)$ as n grow

"sparse network"

$\Rightarrow p = \Theta\left(\frac{1}{n}\right)$

network, mean degree typically remains constant.

$$(1-p)^{n-1-k}$$

$$\log\left((1-p)^{n-1-k}\right) = (n-1-k)\log(1-p)$$

$$= (n-1-k)\log\left(1 - \frac{c}{n-1}\right)$$

$$= (n-1-k)\frac{-c}{n-1}$$

$$\frac{n-1-k}{n-1} \to 1 \text{ as } n \to \infty$$

$$\approx -c$$

$$(1-p)^{n-1-k} \approx e^{-c}$$

$\log(1+x) \approx x$ if x is very small

$\Rightarrow$

$$\Pr(k) = \binom{n-1}{k} p^k e^{-c}$$

$$= \frac{(n-1)!}{(n-1-k)!\,k!} p^k e^{-c}$$

fun at home!

$$\approx \frac{(n-1)^k}{k!} p^k e^{-c}$$

$$= \frac{(n-1)^k}{k!} p^k e^{-c}$$

$$= \frac{\cancel{(n-1)^k}}{k!} \left(\frac{c}{\cancel{n-1}}\right)^k e^{-c}$$

$$\boxed{P(k) = \frac{c^k}{k!} e^{-c} \quad \text{when } p \text{ is very small}}$$

Poisson!

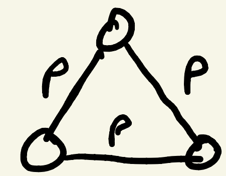
Why?

$$\log P(k) = \log \frac{c^k}{k!} e^{-c}$$

$$= k \log c - \underset{\substack{\downarrow \\ \text{Stirling's Approx}}}{\log k!} - c$$

What about clustering?

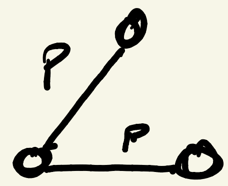
$$C = \frac{\# \Delta}{\# \text{ connected triples}}$$



what's prob that all are connected?

$\Rightarrow p^3$

$$\propto \frac{\binom{n}{3} p^3}{\binom{n}{3} p^2}$$



what's the prob that connect triple?

p^2

$$= p$$

$$= \frac{c}{n-1}$$

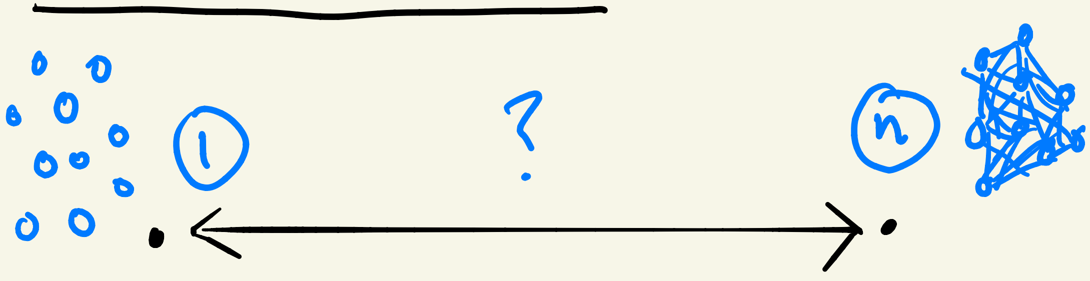
$$\Rightarrow \lim_{n \to \infty} C = 0$$

Sparse network, under ER, have zero clustering.

$\Rightarrow$ sparse, $G(n,p)$ are "locally tree-like"

Real Soc. nets. Do NOT have this property!

Phase Transition in Connectedness



size of largest conn. component.

intensive
indep. of system's size

extensive
depends on system's size.

p governs the transition from size of largest conn. component. intensive $\to$ extensive

- Let u be the avg. fraction of vertices in $G(n,p)$ that do not belong to the giant component (G.C.)

e.g. $p=0$, $u=1$
(no G.C.)

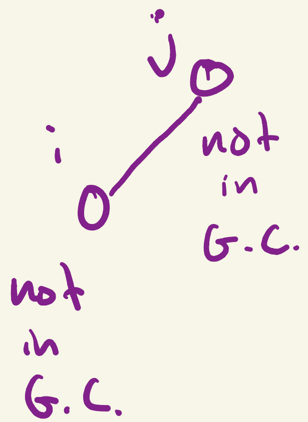
Let u be the prob. that a vertex chosen u.a.r. is not in the G.C.

$\in$ \in
$\forall$ \forall

If $i \notin G.C.$

Then $\forall j$ either: $i \not\leftrightarrow j$ $1-p$

or: $i \leftrightarrow j$ $p u$
but $j \notin G.C.$



$$u = (1-p + p \cdot u)^{n-1}$$

$$u = (1-p + p\cdot u)^{n-1}$$

$$p = \frac{c}{n-1}$$

$$u = \left(1 - p(1-u)\right)^{n-1}$$

$$u = \left(1 - \frac{c(1-u)}{n-1}\right)^{n-1}$$

recall that

$$\lim_{n\to\infty}\left(1 - \frac{x}{n}\right)^n = e^{-x}$$

$$u = e^{-c(1-u)}$$

for v. large networks.

Let S = size of G.C. $= 1-u$

$$1 - S = e^{-cS} \quad \text{or}$$

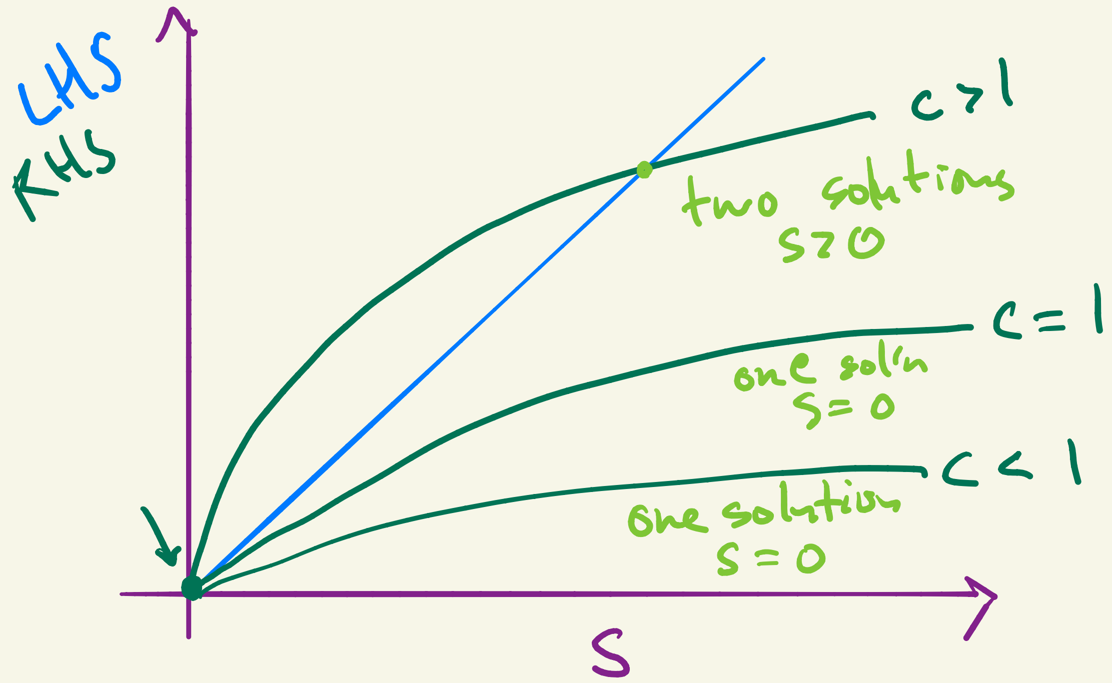
$$\boxed{S = 1 - e^{-cS}}$$

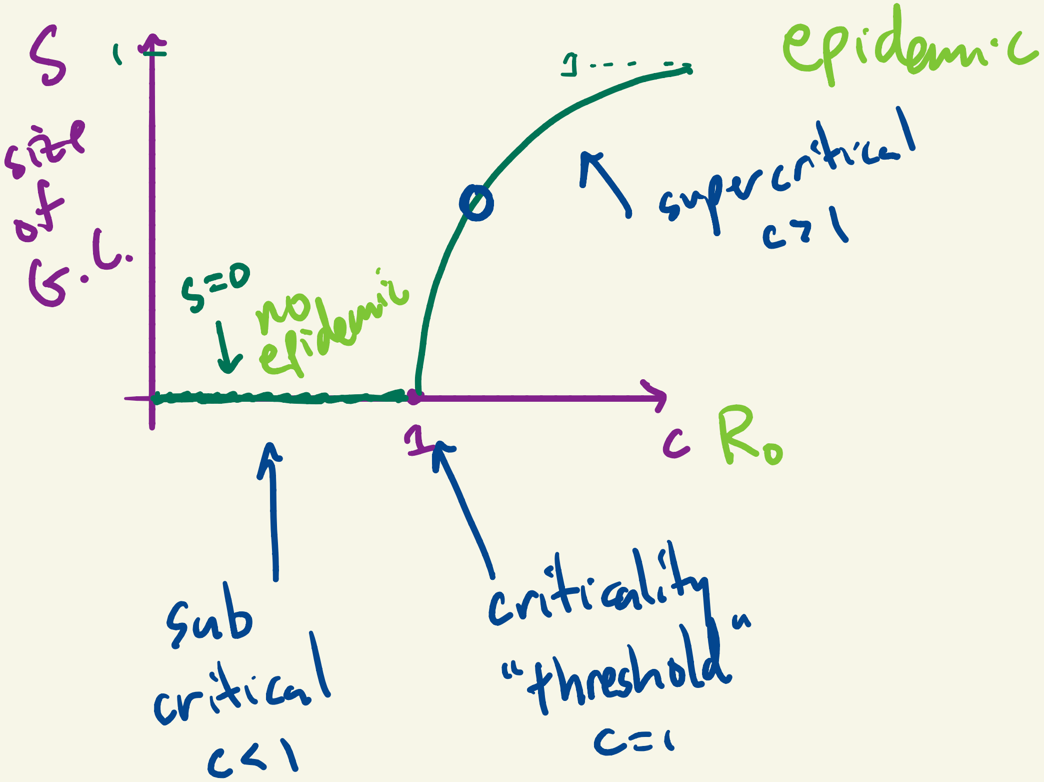
$c = 20$

$S = \frac{1}{2}$

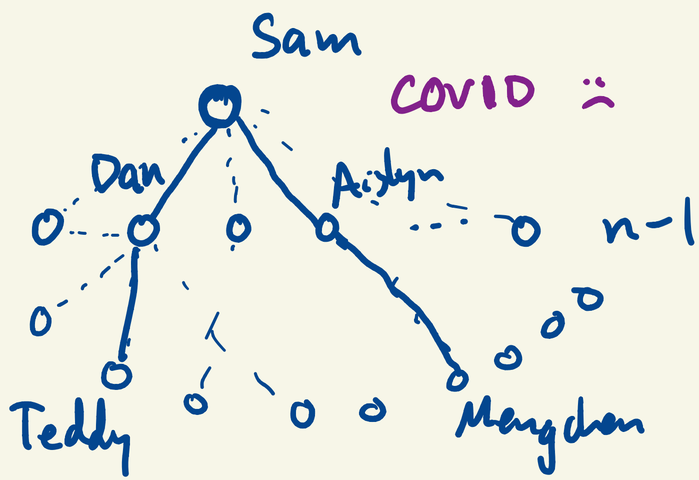
$1 - e^{-10}$

- S is the fraction of network's nodes that are in the G.C. $S \in [0,1]$
- c is the mean degree. It's a # not a probability.
- equation is "transcendental" $\Rightarrow$ no closed form soln.





Imagine one node



Dynamics.

Branching Process?

o at this point $G(n,p)$

$G(n, \frac{c}{n-1})$

If, on average, $c>1$

$\Rightarrow$ epidemic size (G.C.)

$\Theta(n)$ extensive!

if $c<1$

$\Rightarrow$ epidemic size <u>intensive</u>!

fraction of ppl infected $\rightarrow 0$