Random Networks

Complex Networks, Course 303A, Spring, 2009

Prof. Peter Dodds

Department of Mathematics & Statistics
University of Vermont

Outline

Random networks

Pure, abstract random networks:

- Consider set of all networks with N labelled nodes and m edges.
- Standard random network = randomly chosen network from this set.
- To be clear: each network is equally probable.
- Sometimes equiprobability is a good assumption, but it is always an assumption.
- Known as Erdős-Rényi random networks or ER graphs.

Random networks

Some features:

- Number of possible edges:

$$0 \leq m \leq \binom{N}{2} = \frac{N(N-1)}{2}$$

- Given m edges, there are $\binom{\binom{N}{2}}{m}$ different possible networks.
- Crazy factorial explosion for $1 \ll m \ll \binom{N}{2}$.
- Limit of $m = 0$: empty graph.
- Limit of $m = \binom{N}{2}$: complete or fully-connected graph.
- Real world: links are usually costly so real networks are almost always sparse.

Random networks

How to build standard random networks:

- Given N and m.
- Two probablistic methods (we'll see a third later on)

1. Connect each of the $\binom{N}{2}$ pairs with appropriate probability p.
   - Useful for theoretical work.
2. Take N nodes and add exactly m links by selecting edges without replacement.
   - Algorithm: Randomly choose a pair of nodes i and j, $i \neq j$, and connect if unconnected; repeat until all m edges are allocated.
   - Best for adding small numbers of links (most cases).
   - 1 and 2 are effectively equivalent for large N.

Random networks

A few more things:

- For method 1, # links is probablistic:

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

- So the expected or average degree is

$$\langle k \rangle = \frac{2\,\langle m \rangle}{N}$$

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

- Which is what it should be...
- If we keep $\langle k \rangle$ constant then $p \propto 1/N \to 0$ as $N \to \infty$.

Random networks: examples

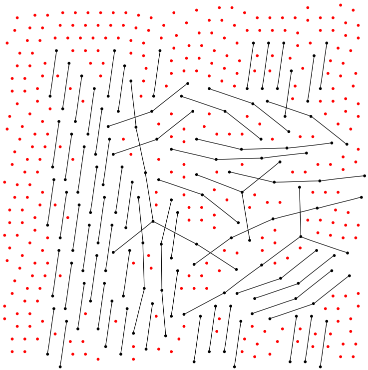
Next slides:

Example realizations of random networks

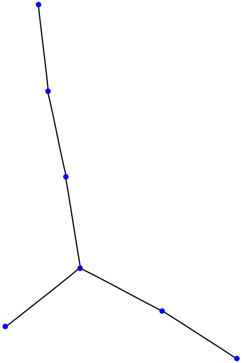
- $N = 500$
- Vary m, the number of edges from 100 to 1000.
- Average degree $\langle k \rangle$ runs from 0.4 to 4.
- Look at full network plus the largest component.

Random networks: examples

entire network:



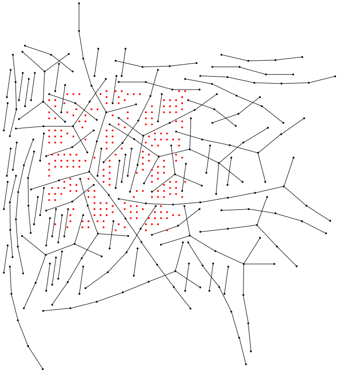
largest component:

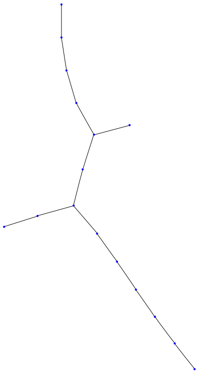


$N = 500$, number of edges $m = 100$
average degree $\langle k \rangle = 0.4$

Random networks: examples

entire network: largest component:

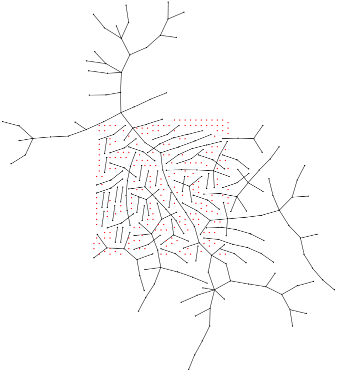




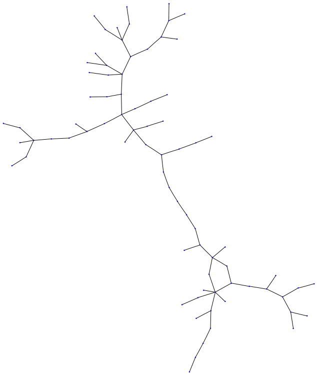
$N = 500$, number of edges $m = 200$
average degree $\langle k \rangle = 0.8$

Random networks: examples

entire network:



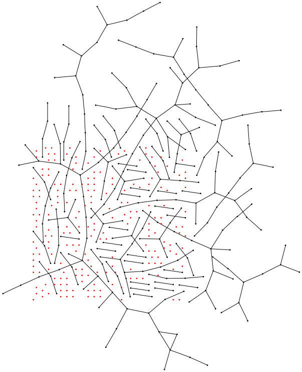
largest component:



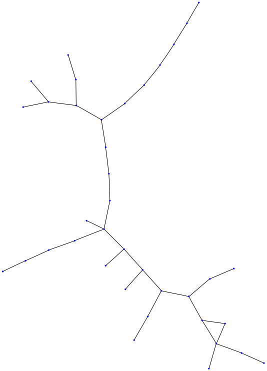
$N = 500$, number of edges $m = 230$
average degree $\langle k \rangle = 0.92$

Random networks: examples

entire network:



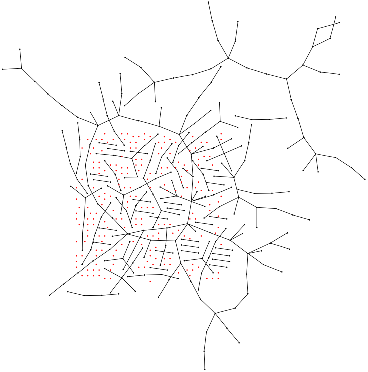
largest component:



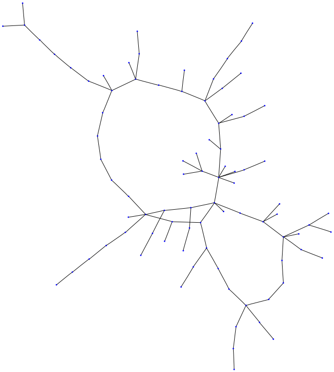
$N = 500$, number of edges $m = 240$
average degree $\langle k \rangle = 0.96$

Random networks: examples

entire network:



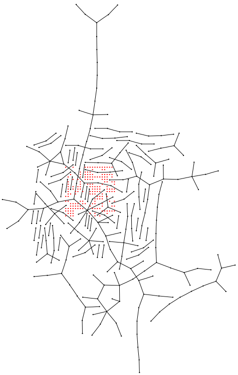
largest component:



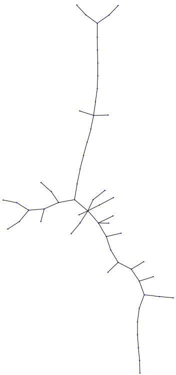
$N = 500$, number of edges $m = 250$
average degree $\langle k \rangle = 1$

Random networks: examples

entire network:



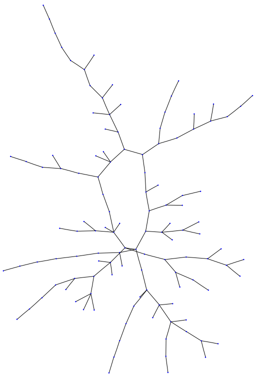
largest component:



$N = 500$, number of edges $m = 260$
average degree $\langle k \rangle = 1.04$

Random networks: examples

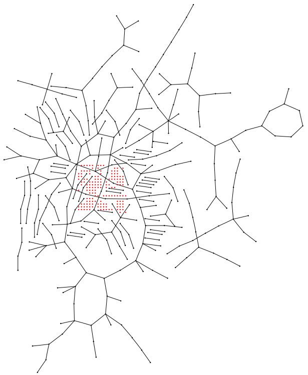
entire network: largest component:



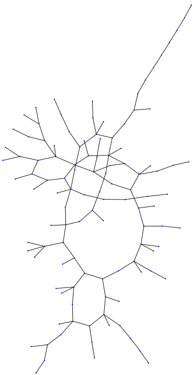
$N = 500$, number of edges $m = 280$
average degree $\langle k \rangle = 1.12$

Random networks: examples

entire network:



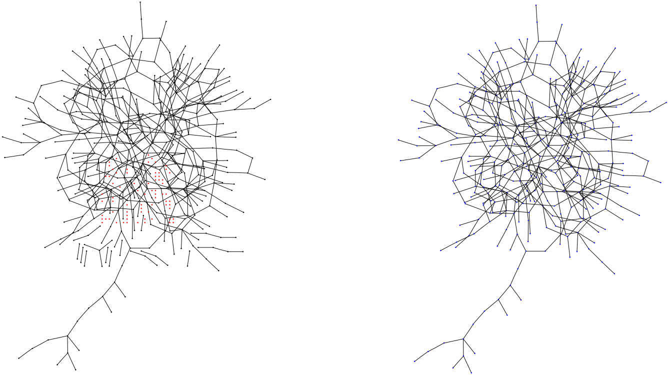
largest component:



$N = 500$, number of edges $m = 300$
average degree $\langle k \rangle = 1.2$

Random networks: examples

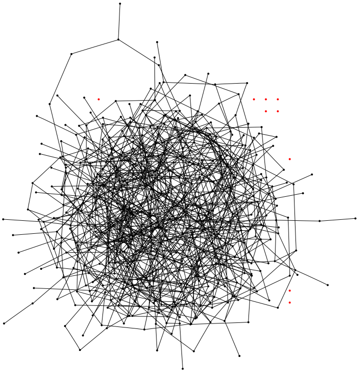
entire network: largest component:



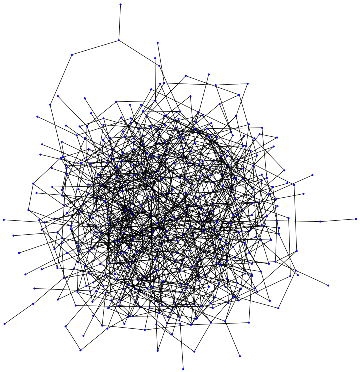
$N = 500$, number of edges $m = 500$
average degree $\langle k \rangle = 2$

Random networks: examples

entire network:



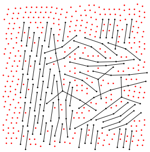
largest component:



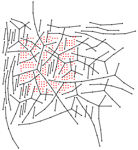
$N = 500$, number of edges $m = 1000$

average degree $\langle k \rangle = 4$

Random networks: examples for N=500

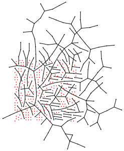


$m = 100$
$\langle k \rangle = 0.4$



$m = 200$
$\langle k \rangle = 0.8$

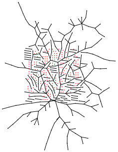
$m = 230$
$\langle k \rangle = 0.92$



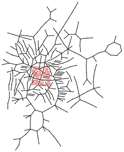
$m = 240$
$\langle k \rangle = 0.96$

$m = 250$
$\langle k \rangle = 1$

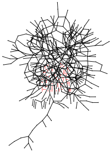
$m = 260$
$\langle k \rangle = 1.04$



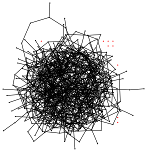
$m = 280$
$\langle k \rangle = 1.12$



$m = 300$
$\langle k \rangle = 1.2$

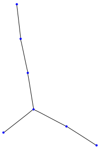


$m = 500$
$\langle k \rangle = 2$

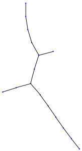


$m = 1000$
$\langle k \rangle = 4$

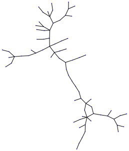
Random networks: largest components



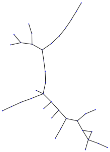
$m = 100$
$\langle k \rangle = 0.4$



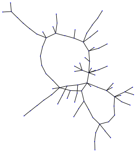
$m = 200$
$\langle k \rangle = 0.8$



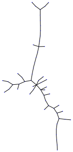
$m = 230$
$\langle k \rangle = 0.92$



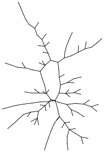
$m = 240$
$\langle k \rangle = 0.96$



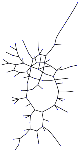
$m = 250$
$\langle k \rangle = 1$



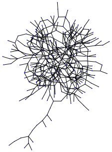
$m = 260$
$\langle k \rangle = 1.04$



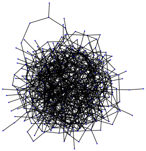
$m = 280$
$\langle k \rangle = 1.12$



$m = 300$
$\langle k \rangle = 1.2$

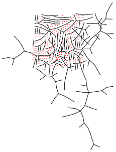


$m = 500$
$\langle k \rangle = 2$

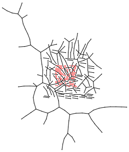


$m = 1000$
$\langle k \rangle = 4$

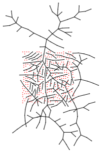
Random networks: examples for N=500



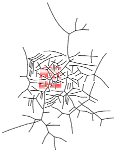
$m = 250$
$\langle k \rangle = 1$



$m = 250$
$\langle k \rangle = 1$

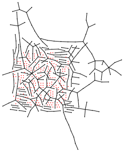


$m = 250$
$\langle k \rangle = 1$



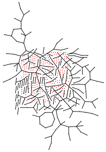
$m = 250$
$\langle k \rangle = 1$

$m = 250$
$\langle k \rangle = 1$

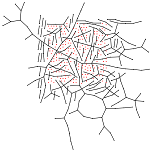


$m = 250$
$\langle k \rangle = 1$

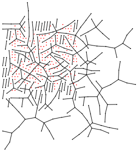
$m = 250$
$\langle k \rangle = 1$



$m = 250$
$\langle k \rangle = 1$

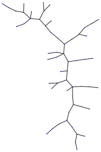


$m = 250$
$\langle k \rangle = 1$

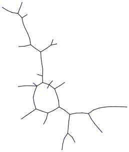


$m = 250$
$\langle k \rangle = 1$

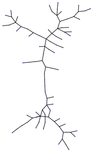
Random networks: largest components



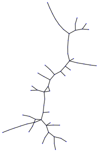
$m = 250$
$\langle k \rangle = 1$



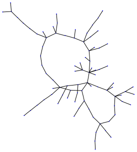
$m = 250$
$\langle k \rangle = 1$



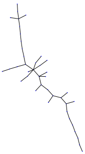
$m = 250$
$\langle k \rangle = 1$



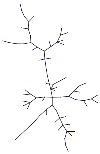
$m = 250$
$\langle k \rangle = 1$



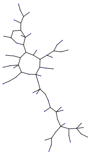
$m = 250$
$\langle k \rangle = 1$



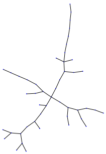
$m = 250$
$\langle k \rangle = 1$



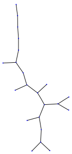
$m = 250$
$\langle k \rangle = 1$



$m = 250$
$\langle k \rangle = 1$



$m = 250$
$\langle k \rangle = 1$



$m = 250$
$\langle k \rangle = 1$

Random networks

Clustering:

- For method 1, what is the clustering coefficient for a finite network?
- Consider triangle/triple clustering coefficient (Newman [1]):

$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$$

- Recall: C_2 = probability that two nodes are connected given they have a friend in common.
- For standard random networks, we have simply that

$$C_2 = p.$$

Random networks

Clustering:

- So for large random networks ($N \to \infty$), clustering drops to zero.
- Key structural feature of random networks is that they locally look like branching networks (no loops).

Random networks

Degree distribution:

- Recall p_k = probability that a randomly selected node has degree k.
- Consider method 1 for constructing random networks: each possible link is realized with probability p.
- Now consider one node: there are '$N-1$ choose k' ways the node can be connected to k of the other $N-1$ nodes.
- Each connection occurs with probability p, each non-connection with probability $(1-p)$.
- Therefore have a binomial distribution:

$$P(k;p,N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

Random networks

Limiting form of $P(k; p, N)$:

- Our degree distribution: $P(k; p, N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$.
- What happens as $N \to \infty$?
- We must end up with the normal distribution right?
- If p is fixed, then we would end up with a Gaussian with average degree $\langle k \rangle \simeq pN \to \infty$.
- But we want to keep $\langle k \rangle$ fixed...
- So examine limit of $P(k; p, N)$ when $p \to 0$ and $N \to \infty$ with $\langle k \rangle = p(N-1) = \text{constant}$.

Limiting form of $P(k; p, N)$:

- Substitute $p = \frac{\langle k \rangle}{N-1}$ into $P(k; p, N)$ and hold k fixed:

$$P(k; p, N) = \binom{N-1}{k} \left(\frac{\langle k \rangle}{N-1}\right)^k \left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k}$$

$$= \frac{(N-1)!}{k!(N-1-k)!} \frac{\langle k \rangle^k}{(N-1)^k} \left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k}$$

$$= \frac{(N-1)(N-2)\cdots(N-k)}{k!} \frac{\langle k \rangle^k}{(N-1)^k} \left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k}$$

$$\simeq \frac{\cancel{N^k}(1-\frac{1}{\cancel{N}})\cdots(1-\frac{k}{\cancel{N}})}{k!\cancel{N^k}} \frac{\langle k \rangle^k}{(1-\frac{1}{\cancel{N}})^k} \left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k}$$

Limiting form of $P(k; p, N)$:

- We are now here:

$$P(k; p, N) \simeq \frac{\langle k \rangle^k}{k!}\left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k}$$

- Now use the excellent result:

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

(Use l'Hôpital's rule to prove.)

- Identifying $n = N - 1$ and $x = -\langle k \rangle$:

$$P(k; \langle k \rangle) \simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}\left(1 - \frac{\langle k \rangle}{N-1}\right)^{-k} \rightarrow \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

- This is a Poisson distribution (⊞) with mean $\langle k \rangle$.

General random networks

- So... standard random networks have a Poisson degree distribution
- Generalize to arbitrary degree distribution P_k.
- Also known as the configuration model [1].
- Can generalize construction method from ER random networks.
- Assign each node a weight w from some distribution P_w and form links with probability

$$P(\text{link between } i \text{ and } j) \propto w_i w_j.$$

- But we'll be more interested in
  1. Randomly wiring up (and rewiring) already existing nodes with fixed degrees.
  2. Examining mechanisms that lead to networks with certain degree distributions.

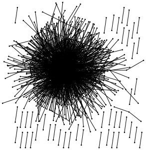
Random networks: examples

Coming up:

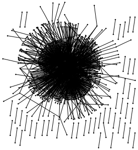
Example realizations of random networks with power law degree distributions:

- $N = 1000$.
- $P_k \propto k^{-\gamma}$ for $k \geq 1$.
- Set $P_0 = 0$ (no isolated nodes).
- Vary exponent γ between 2.10 and 2.91.
- Again, look at full network plus the largest component.
- Apart from degree distribution, wiring is random.

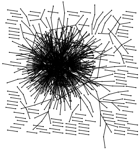
Random networks: examples for N=1000



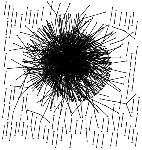
γ = 2.1
$\langle k \rangle$ = 3.448



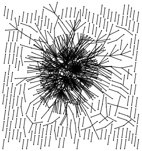
γ = 2.19
$\langle k \rangle$ = 2.986



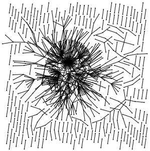
γ = 2.28
$\langle k \rangle$ = 2.306



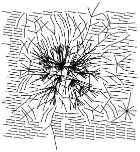
γ = 2.37
$\langle k \rangle$ = 2.504



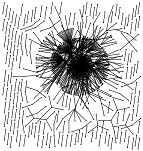
γ = 2.46
$\langle k \rangle$ = 1.856



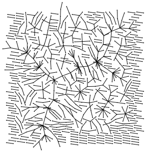
γ = 2.55
$\langle k \rangle$ = 1.712



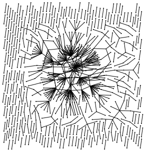
γ = 2.64
$\langle k \rangle$ = 1.6



γ = 2.73
$\langle k \rangle$ = 1.862

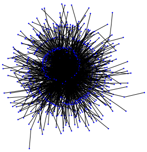


γ = 2.82
$\langle k \rangle$ = 1.386

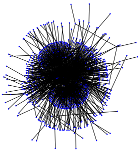


γ = 2.91
$\langle k \rangle$ = 1.49

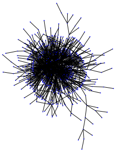
Random networks: largest components



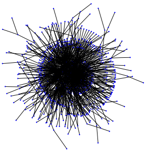
γ = 2.1
$\langle k \rangle$ = 3.448



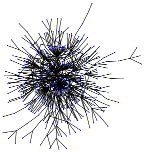
γ = 2.19
$\langle k \rangle$ = 2.986



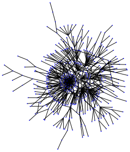
γ = 2.28
$\langle k \rangle$ = 2.306



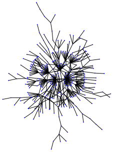
γ = 2.37
$\langle k \rangle$ = 2.504



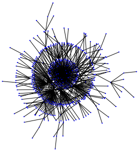
γ = 2.46
$\langle k \rangle$ = 1.856



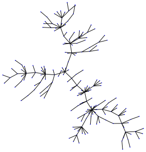
γ = 2.55
$\langle k \rangle$ = 1.712



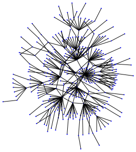
γ = 2.64
$\langle k \rangle$ = 1.6



γ = 2.73
$\langle k \rangle$ = 1.862



γ = 2.82
$\langle k \rangle$ = 1.386



γ = 2.91
$\langle k \rangle$ = 1.49

Poisson basics:

- Normalization: we must have

$$\sum_{k=0}^{\infty} P(k;\langle k\rangle) = 1$$

- Checking:

$$\sum_{k=0}^{\infty} P(k;\langle k\rangle) = \sum_{k=0}^{\infty} \frac{\langle k\rangle^k}{k!} e^{-\langle k\rangle}$$

$$= e^{-\langle k\rangle} \sum_{k=0}^{\infty} \frac{\langle k\rangle^k}{k!}$$

$$= e^{-\langle k\rangle} e^{\langle k\rangle} = 1 \checkmark$$

Poisson basics:

- Mean degree: we must have

$$\langle k \rangle = \sum_{k=0}^{\infty} k P(k; \langle k \rangle).$$

- Checking:

$$\sum_{k=0}^{\infty} k P(k; \langle k \rangle) = \sum_{k=0}^{\infty} k \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

$$= e^{-\langle k \rangle} \sum_{k=1}^{\infty} \frac{\langle k \rangle^k}{(k-1)!}$$

$$= \langle k \rangle e^{-\langle k \rangle} \sum_{k=1}^{\infty} \frac{\langle k \rangle^{k-1}}{(k-1)!}$$

$$= \langle k \rangle e^{-\langle k \rangle} \sum_{i=0}^{\infty} \frac{\langle k \rangle^i}{i!} = \langle k \rangle e^{-\langle k \rangle} e^{\langle k \rangle} = \langle k \rangle \checkmark$$

- We'll get to a better way of doing this...

Poisson basics:

- The variance of degree distributions for random networks turns out to be very important.
- Use calculation similar to one for finding $\langle k \rangle$ to find the second moment:

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle.$$

- Variance is then

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = \langle k \rangle^2 + \langle k \rangle - \langle k \rangle^2 = \langle k \rangle.$$

- So standard deviation σ is equal to $\sqrt{\langle k \rangle}$.
- Note: This is a special property of Poisson distribution and can trip us up...

The edge-degree distribution:

- The degree distribution P_k is fundamental for our description of many complex networks
- Again: P_k is the degree of randomly chosen node.
- A second very important distribution arises from choosing randomly on edges rather than on nodes.
- Define Q_k to be the probability the node at a random end of a randomly chosen edge has degree k.
- Now choosing nodes based on their degree (i.e., size):

$$\boxed{Q_k \propto k P_k}$$

- Normalized form:

$$Q_k = \frac{k P_k}{\sum_{k'=0}^{\infty} k' P_{k'}} = \frac{k P_k}{\langle k \rangle}.$$

The edge-degree distribution:

- For random networks, Q_k is also the probability that a friend (neighbor) of a random node has k friends.
- Useful variant on Q_k:

  R_k = probability that a friend of a random node has k other friends.

- $$R_k = \frac{(k+1)P_{k+1}}{\sum_{k'=0}(k'+1)P_{k'+1}} = \frac{(k+1)P_{k+1}}{\langle k \rangle}$$

- Equivalent to friend having degree $k+1$.
- Natural question: what's the expected number of other friends that one friend has?

The edge-degree distribution:

- Given R_k is the probability that a friend has k other friends, then the average number of friends' other friends is

$$\langle k \rangle_R = \sum_{k=0}^{\infty} k R_k = \sum_{k=0}^{\infty} k \frac{(k+1)P_{k+1}}{\langle k \rangle}$$

$$= \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k(k+1)P_{k+1}$$

$$= \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} \left((k+1)^2 - (k+1)\right) P_{k+1}$$

(where we have sneakily matched up indices)

$$= \frac{1}{\langle k \rangle} \sum_{j=0}^{\infty} (j^2 - j) P_j \quad \text{(using j = k+1)}$$

$$= \frac{1}{\langle k \rangle} \left(\langle k^2 \rangle - \langle k \rangle\right)$$

The edge-degree distribution:

- Note: our result, $\langle k \rangle_R = \frac{1}{\langle k \rangle}\left(\langle k^2 \rangle - \langle k \rangle\right)$, is true for all random networks, independent of degree distribution.
- For standard random networks, recall

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle.$$

- Therefore:

$$\langle k \rangle_R = \frac{1}{\langle k \rangle}\left(\langle k \rangle^2 + \langle k \rangle - \langle k \rangle\right) = \langle k \rangle$$

- Again, neatness of results is a special property of the Poisson distribution.
- So friends on average have $\langle k \rangle$ other friends, and $\langle k \rangle + 1$ total friends...

Two reasons why this matters

Reason #1:

- Average # friends of friends per node is

$$\langle k_2 \rangle = \langle k \rangle \times \langle k \rangle_R = \langle k \rangle \frac{1}{\langle k \rangle} \left(\langle k^2 \rangle - \langle k \rangle \right) = \langle k^2 \rangle - \langle k \rangle .$$

- Key: Average depends on the 1st and 2nd moments of P_k and not just the 1st moment.
- Three peculiarities:
  1. We might guess $\langle k_2 \rangle = \langle k \rangle (\langle k \rangle - 1)$ but it's actually $\langle k(k-1) \rangle$.
  2. If P_k has a large second moment, then $\langle k_2 \rangle$ will be big. (e.g., in the case of a power-law distribution)
  3. Your friends are different to you...

Two reasons why this matters

More on peculiarity #3:

- A node's average # of friends: $\langle k \rangle$
- Friend's average # of friends: $\frac{\langle k^2 \rangle}{\langle k \rangle}$
- Comparison:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = \langle k \rangle \frac{\langle k^2 \rangle}{\langle k \rangle^2} = \langle k \rangle \frac{\sigma^2 + \langle k \rangle^2}{\langle k \rangle^2} = \langle k \rangle \left(1 + \frac{\sigma^2}{\langle k \rangle^2}\right) \geq \langle k \rangle$$

- So only if everyone has the same degree (variance$= \sigma^2 = 0$) can a node be the same as its friends.
- Intuition: for random networks, the more connected a node, the more likely it is to be chosen as a friend.

Two reasons why this matters

(Big) Reason #2:

- $\langle k \rangle_R$ is key to understanding how well random networks are connected together.
- e.g., we'd like to know what's the size of the largest component within a network.
- As $N \to \infty$, does our network have a giant component?
- Defn: Component = connected subnetwork of nodes such that $\exists$ path between each pair of nodes in the subnetwork, and no node outside of the subnetwork is connected to it.
- Defn: Giant component = component that comprises a non-zero fraction of a network as $N \to \infty$.
- Note: Component = Cluster

Structure of random networks

Giant component:

- A giant component exists if when we follow a random edge, we are likely to hit a node with at least 1 other outgoing edge.
- Equivalently, expect exponential growth in node number as we move out from a random node.
- All of this is the same as requiring $\langle k \rangle_R > 1$.
- Giant component condition (or percolation condition):

$$\langle k \rangle_R = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1$$

- Again, see that the second moment is an essential part of the story.
- Equivalent statement: $\langle k^2 \rangle > 2\langle k \rangle$

Giant component

Standard random networks:

- Recall $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$.
- Condition for giant component:

$$\langle k \rangle_R = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} = \frac{\langle k \rangle^2 + \langle k \rangle - \langle k \rangle}{\langle k \rangle} = \langle k \rangle$$

- Therefore when $\langle k \rangle > 1$, standard random networks have a giant component.
- When $\langle k \rangle < 1$, all components are finite.
- Fine example of a continuous phase transition (⊞).
- We say $\langle k \rangle = 1$ marks the critical point of the system.

Giant component

Random networks with skewed P_k:

- e.g, if $P_k = ck^{-\gamma}$ with $2 < \gamma < 3$ then

$$\langle k^2 \rangle = c \sum_{k=0}^{\infty} k^2 k^{-\gamma}$$

$$\sim \int_{x=0}^{\infty} x^{2-\gamma} \mathrm{d}x$$

$$\propto x^{3-\gamma} \Big|_{x=0}^{\infty} = \infty \quad (> \langle k \rangle).$$

- So giant component always exists for these kinds of networks.
- Cutoff scaling is k^{-3}: if $\gamma > 3$ then we have to look harder at $\langle k \rangle_R$.

Giant component

And how big is the largest component?

- Define S_1 as the size of the largest component.
- Consider an infinite ER random network with average degree $\langle k \rangle$.
- Let's find S_1 with a back-of-the-envelope argument.
- Define δ as the probability that a randomly chosen node does not belong to the largest component.
- Simple connection: $\delta = 1 - S_1$.
- Dirty trick: If a randomly chosen node is not part of the largest component, then none of its neighbors are.
- So

$$\delta = \sum_{k=0}^{\infty} P_k \delta^k$$

- Substitute in Poisson distribution...

Giant component

- Carrying on:

$$\delta = \sum_{k=0}^{\infty} P_k \delta^k = \sum_{k=0}^{\infty} \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \delta^k$$

$$= e^{-\langle k \rangle} \sum_{k=0}^{\infty} \frac{(\langle k \rangle \delta)^k}{k!}$$

$$= e^{-\langle k \rangle} e^{\langle k \rangle \delta} = e^{-\langle k \rangle (1-\delta)}.$$

- Now substitute in $\delta = 1 - S_1$ and rearrange to obtain:

$$S_1 = 1 - e^{-\langle k \rangle S_1}.$$

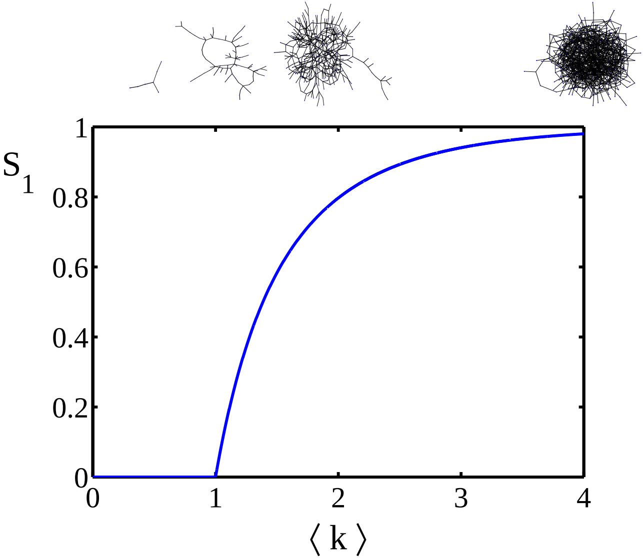
Giant component

- We can figure out some limits and details for $S_1 = 1 - e^{-\langle k \rangle S_1}$.
- First, we can write $\langle k \rangle$ in terms of S_1:

$$\langle k \rangle = \frac{1}{S_1} \ln \frac{1}{1 - S_1}.$$

- As $\langle k \rangle \to 0$, $S_1 \to 0$.
- As $\langle k \rangle \to \infty$, $S_1 \to 1$.
- Notice that at $\langle k \rangle = 1$, the critical point, $S_1 = 0$.
- Only solvable for $S > 0$ when $\langle k \rangle > 1$.
- Really a transcritical bifurcation [2].

Giant component



Giant component

Turns out we were lucky...

- Our dirty trick only works for ER random networks.
- The problem: We assumed that neighbors have the same probability δ of belonging to the largest component.
- But we know our friends are different from us...
- Works for ER random networks because $\langle k \rangle = \langle k \rangle_R$.
- We need a separate probability δ' for the chance that a node at the end of a random edge is part of the largest component.
- We can do this but we need to enhance our toolkit with Generatingfunctionology... [3]

Generating functions

- Idea: Given a sequence $a_0, a_1, a_2, \ldots,$ associate each element with a distinct function or other mathematical object.
- Well-chosen functions allow us to manipulate sequences and retrieve sequence elements.

Definition:

- The generating function (g.f.) for a sequence $\{a_n\}$ is

$$F(x) = \sum_{n=0}^{\infty} a_n x^n.$$

- Roughly: transforms a vector in R^∞ into a function defined on R^1.
- Related to Fourier, Laplace, Mellin, ...

Simple example

Rolling dice:

- $p_k^{(\square)} = \mathbf{Pr}(\text{throwing a } k) = 1/6$ where $k = 1, 2, \ldots, 6$.
- $$F^{(\square)}(x) = \sum_{k=1}^{6} p_k x^k = \frac{1}{6}(x + x^2 + x^3 + x^4 + x^5 + x^6).$$
- We'll come back to this simple example as we derive various delicious properties of generating functions.

Example

- Take a degree distribution with exponential decay:

$$P_k = ce^{-\lambda k}$$

  where $c = 1 - e^{-\lambda}$.
- The generating function for this distribution is

$$F(x) = \sum_{k=0}^{\infty} P_k x^k = \sum_{k=0}^{\infty} ce^{-\lambda k} x^k = \frac{c}{1 - xe^{-\lambda}}.$$

- Notice that $F(1) = c/(1 - e^{-\lambda}) = 1$.
- For probability distributions, we must always have $F(1) = 1$ since

$$F(1) = \sum_{k=0}^{\infty} P_k 1^k = \sum_{k=0}^{\infty} P_k = 1.$$

Properties of generating functions

- Average degree:

$$\langle k \rangle = \sum_{k=0}^{\infty} k P_k = \sum_{k=0}^{\infty} k P_k x^{k-1} \bigg|_{x=1}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x} F(x) \bigg|_{x=1} = F'(1)$$

- In general, many calculations become simple, if a little abstract.
- For our exponential example:

$$F'(x) = \frac{(1 - e^{-\lambda}) e^{-\lambda}}{(1 - x e^{-\lambda})^2}.$$

- So:

$$\langle k \rangle = F'(1) = \frac{e^{-\lambda}}{(1 - e^{-\lambda})}.$$

Properties of generating functions

Useful pieces for probability distributions:

- Normalization:

$$F(1) = 1$$

- First moment:

$$\langle k \rangle = F'(1)$$

- Higher moments:

$$\langle k^n \rangle = \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n F(x) \bigg|_{x=1}$$

- kth element of sequence (general):

$$P_k = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} F(x) \bigg|_{x=0}$$

Edge-degree distribution

- Recall our condition for a giant component:

$$\langle k \rangle_R = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1.$$

- Let's rëexpress our condition in terms of generating functions.
- We first need the g.f. for R_k.
- We'll now use this notation:

  $F_P(x)$ is the g.f. for P_k.
  $F_R(x)$ is the g.f. for R_k.

- Condition in terms of g.f. is:

$$\langle k \rangle_R = F_R'(1) > 1.$$

- Now find how F_R is related to F_P...

Edge-degree distribution

- We have

$$F_R(x) = \sum_{k=0}^{\infty} R_k x^k = \sum_{k=0}^{\infty} \frac{(k+1)P_{k+1}}{\langle k \rangle} x^k.$$

Shift index to $j = k+1$ and pull out $\frac{1}{\langle k \rangle}$:

$$F_R(x) = \frac{1}{\langle k \rangle} \sum_{j=1}^{\infty} j P_j x^{j-1} = \frac{1}{\langle k \rangle} \sum_{j=1}^{\infty} P_j \frac{\mathrm{d}}{\mathrm{d}x} x^j$$

$$= \frac{1}{\langle k \rangle} \frac{\mathrm{d}}{\mathrm{d}x} \sum_{j=1}^{\infty} P_j x^j = \frac{1}{\langle k \rangle} \frac{\mathrm{d}}{\mathrm{d}x} \left(F_P(x) - P_0 \right) = \frac{1}{\langle k \rangle} F_P'(x).$$

Finally, since $\langle k \rangle = F_P'(1)$,

$$\boxed{F_R(x) = \frac{F_P'(x)}{F_P'(1)}}$$

Edge-degree distribution

- Recall giant component condition is $\langle k \rangle_R = F_R'(1) > 1$.
- Since we have $F_R(x) = F_P'(x)/F_P'(1)$,

$$F_R'(x) = \frac{F_P''(x)}{F_P'(1)}.$$

- Setting $x = 1$, our condition becomes

$$\boxed{\frac{F_P''(1)}{F_P'(1)} > 1}$$

Size distributions

To figure out the size of the largest component (S_1), we need more resolution on component sizes.

Definitions:

- π_n = probability that a random node belongs to a finite component of size $n < \infty$.
- ρ_n = probability a random link leads to a finite subcomponent of size $n < \infty$.

Local-global connection:

$$P_k, R_k \Leftrightarrow \pi_n, \rho_n$$

neighbors $\Leftrightarrow$ components

Size distributions

G.f.'s for component size distributions:

- $$F_{\pi}(x) = \sum_{n=0}^{\infty} \pi_n x^n \text{ and } F_{\rho}(x) = \sum_{n=0}^{\infty} \rho_n x^n$$

The largest component:

- Subtle key: $F_{\pi}(1)$ is the probability that a node belongs to a finite component.
- Therefore: $S_1 = 1 - F_{\pi}(1)$.

Our mission, which we accept:

- Find the four generating functions

$$F_P, F_R, F_{\pi}, \text{ and } F_{\rho}.$$

Useful results we'll need for g.f.'s

Sneaky Result 1:

- Consider two random variables U and V whose values may be $0, 1, 2, \ldots$
- Write probability distributions as U_k and U_k and g.f.'s as F_U and F_V.
- SR1: If a third random variable is defined as

$$W = \sum_{i=1}^{V} U^{(i)} \text{ with each } U^{(i)} \stackrel{d}{=} U$$

then

$$\boxed{F_W(x) = F_V\left(F_U(x)\right)}$$

Proof of SR1:

Write probability that variable W has value k as W_k.

$$W_k = \sum_{j=0}^{\infty} V_j \times \Pr(\text{sum of } j \text{ draws of variable } U = k)$$

$$= \sum_{j=0}^{\infty} V_j \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} U_{i_1} U_{i_2} \cdots U_{i_j}$$

$$\therefore F_W(x) = \sum_{k=0}^{\infty} W_k x^k = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} V_j \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} U_{i_1} U_{i_2} \cdots U_{i_j} x^k$$

$$= \sum_{j=0}^{\infty} V_j \sum_{k=0}^{\infty} \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} U_{i_1} x^{i_1} U_{i_2} x^{i_2} \cdots U_{i_j} x^{i_j}$$

Proof of SR1:

With some concentration, observe:

$$F_W(x) = \sum_{j=0}^{\infty} V_j \underbrace{\sum_{k=0}^{\infty} \overbrace{\sum_{\substack{\{i_1,i_2,\ldots,i_j\}| \\ i_1+i_2+\ldots+i_j=k}} U_{i_1} x^{i_1} U_{i_2} x^{i_2} \cdots U_{i_j} x^{i_j}}^{x^k \text{ piece of } \left(\sum_{i'=0}^{\infty} U_{i'} x^{i'}\right)^j}}_{\left(\sum_{i'=0}^{\infty} U_{i'} x^{i'}\right)^j = (F_U(x))^j}$$

$$= \sum_{j=0}^{\infty} V_j \, (F_U(x))^j$$

$$= F_V\left(F_U(x)\right) \checkmark$$

Useful results we'll need for g.f.'s

Sneaky Result 2:

- Start with a random variable U with distribution U_k $(k = 0, 1, 2, \dots)$
- SR2: If a second random variable is defined as

$$V = U + 1 \quad \text{then} \quad \boxed{F_V(x) = xF_U(x)}$$

- Reason: $V_k = U_{k-1}$ for $k \geq 1$ and $V_0 = 0$.
-

$$\therefore F_V(x) = \sum_{k=0}^{\infty} V_k x^k = \sum_{k=1}^{\infty} U_{k-1} x^k$$

$$= x \sum_{j=0}^{\infty} U_j x^j = xF_U(x). \checkmark$$

Useful results we'll need for g.f.'s

Generalization of SR2:

- (1) If $V = U + i$ then

$$F_V(x) = x^i F_U(x).$$

- (2) If $V = U - i$ then

$$F_V(x) = x^{-i} F_U(x)$$

$$= x^{-i} \sum_{k=0}^{\infty} U_k x^k$$

Connecting generating functions

- Goal: figure out forms of the component generating functions, F_π and F_ρ.
- π_n = probability that a random node belongs to a finite component of size n

$$= \sum_{k=0}^{\infty} P_k \times \Pr \left(\begin{array}{l} \text{sum of sizes of subcomponents} \\ \text{at end of } k \text{ random links} = n - 1 \end{array} \right)$$

- Therefore: $$\boxed{F_\pi(x) = \underbrace{x}_{SR2} \underbrace{F_P\left(F_\rho(x)\right)}_{SR1}}$$

- Extra factor of x accounts for random node itself.

Connecting generating functions

- ρ_n = probability that a random link leads to a finite subcomponent of size n.
- Invoke one step of recursion: ρ_n = probability that a random node arrived along a random edge is part of a finite subcomponent of size n.

$$= \sum_{k=0}^{\infty} R_k \times \Pr\left(\begin{array}{l} \text{sum of sizes of subcomponents} \\ \text{at end of } k \text{ random links} = n-1 \end{array} \right)$$

- Therefore: $$\boxed{F_\rho(x) = \underbrace{x}_{SR2} \underbrace{F_R\left(F_\rho(x)\right)}_{SR1}}$$
- Again, extra factor of x accounts for random node itself.

Connecting generating functions

- We now have two functional equations connecting our generating functions:

$$F_\pi(x) = xF_P\left(F_\rho(x)\right) \text{ and } F_\rho(x) = xF_R\left(F_\rho(x)\right)$$

- Taking stock: We know $F_P(x)$ and $F_R(x) = F_P'(x)/F_P'(1)$.
- We first untangle the second equation to find F_ρ
- We can do this because it only involves F_ρ and F_R.
- The first equation then immediately gives us F_π in terms of F_ρ and F_R.

Component sizes

- Remembering vaguely what we are doing:

  Finding F_π to obtain the fractional size of the largest component $S_1 = 1 - F_\pi(1)$.
- Set $x = 1$ in our two equations:

$$F_\pi(1) = F_P\left(F_\rho(1)\right) \text{ and } F_\rho(1) = F_R\left(F_\rho(1)\right)$$

- Solve second equation numerically for $F_\rho(1)$.
- Plug $F_\rho(1)$ into first equation to obtain $F_\pi(1)$.

Component sizes

Example: Standard random graphs.

- We can show $F_P(x) = e^{-\langle k \rangle (1-x)}$

$$\therefore F_R(x) = F_P'(x)/F_P'(1) = e^{-\langle k \rangle (1-x)} / e^{-\langle k \rangle (1-x')}|_{x'=1}$$

$$= e^{-\langle k \rangle (1-x)} = F_P(x) \qquad \text{...aha!}$$

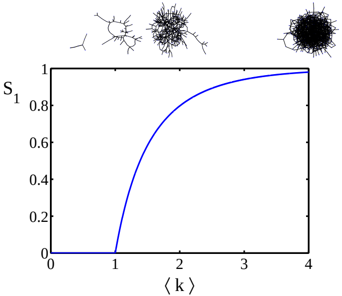
- RHS's of our two equations are the same.
- So $F_\pi(x) = F_\rho(x) = xF_R(F_\rho(x)) = xF_R(F_\pi(x))$
- Why our dirty (but wrong) trick worked earlier...

Component sizes

- We are down to $F_\pi(x) = xF_R(F_\pi(x))$ and $F_R(x) = e^{-\langle k \rangle (1-x)}$.
- $$\therefore F_\pi(x) = xe^{-\langle k \rangle (1 - F_\pi(x))}$$
- We're first after $S_1 = 1 - F_\pi(1)$ so set $x = 1$ and replace $F_\pi(1)$ by $1 - S_1$:

$$1 - S_1 = e^{-\langle k \rangle S_1}$$

$$\text{Or: } \langle k \rangle = \frac{1}{S_1} \ln \frac{1}{1 - S_1}$$



- Just as we found with our dirty trick ...
- Again, we (usually) have to resort to numerics ...

Average component size

- Next: find average size of finite components $\langle n \rangle$.
- Using standard G.F. result: $\langle n \rangle = F'_\pi(1)$.
- Try to avoid finding $F_\pi(x)$...
- Starting from $F_\pi(x) = xF_P(F_\rho(x))$, we differentiate:

$$F'_\pi(x) = F_P(F_\rho(x)) + xF'_\rho(x)F'_P(F_\rho(x))$$

- While $F_\rho(x) = xF_R(F_\rho(x))$ gives

$$F'_\rho(x) = F_R(F_\rho(x)) + xF'_\rho(x)F'_R(F_\rho(x))$$

- Now set $x = 1$ in both equations.
- We solve the second equation for $F'_\rho(1)$ (we must already have $F_\rho(1)$).
- Plug $F'_\rho(1)$ and $F_\rho(1)$ into first equation to find $F'_\pi(1)$.

Average component size

Example: Standard random graphs.

- Use fact that $F_P = F_R$ and $F_\pi = F_\rho$.
- Two differentiated equations reduce to only one:

$$F'_\pi(x) = F_P\left(F_\pi(x)\right) + xF'_\pi(x)F'_P\left(F_\pi(x)\right)$$

$$\text{Rearrange:} \quad F'_\pi(x) = \frac{F_P\left(F_\pi(x)\right)}{1 - xF'_P\left(F_\pi(x)\right)}$$

- Simplify denominator using $F'_P(x) = \langle k \rangle F_P(x)$
- Replace $F_P(F_\pi(x))$ using $F_\pi(x) = xF_P(F_\pi(x))$.
- Set $x = 1$ and replace $F_\pi(1)$ with $1 - S_1$.

$$\text{End result: } \langle n \rangle = F'_\pi(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

Average component size

- Our result for standard random networks:

$$\langle n \rangle = F'_{\pi}(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

- Recall that $\langle k \rangle = 1$ is the critical value of average degree for standard random networks.
- Look at what happens when we increase $\langle k \rangle$ to 1 from below.
- We have $S_1 = 0$ for all $\langle k \rangle < 1$ so

$$\langle n \rangle = \frac{1}{1 - \langle k \rangle}$$

- This blows up as $\langle k \rangle \to 1$.
- Reason: we have a power law distribution of component sizes at $\langle k \rangle = 1$.
- Typical critical point behavior....

Average component size

- Limits of $\langle k \rangle = 0$ and ∞ make sense for

$$\langle n \rangle = F'_{\pi}(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

- As $\langle k \rangle \to 0$, $S_1 = 0$, and $\langle n \rangle \to 1$.
- All nodes are isolated.
- As $\langle k \rangle \to \infty$, $S_1 \to 1$ and $\langle n \rangle \to 0$.
- No nodes are outside of the giant component.

Extra on largest component size:

- For $\langle k \rangle = 1$, $S_1 \sim N^{2/3}$.
- For $\langle k \rangle < 1$, $S_1 \sim \log N$.

References I

[1] M. E. J. Newman.
The structure and function of complex networks.
SIAM Review, 45(2):167–256, 2003. pdf (⊞)

[2] S. H. Strogatz.
Nonlinear Dynamics and Chaos.
Addison Wesley, Reading, Massachusetts, 1994.

[3] H. S. Wilf.
Generatingfunctionology.
A K Peters, Natick, MA, 3rd edition, 2006. pdf (⊞)