Random Networks

Complex Networks
CSYS/MATH 303, Spring, 2011

Prof. Peter Dodds

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont

Outline

Basics
- Definitions
- How to build
- Some visual examples

Structure
- Clustering
- Degree distributions
- Configuration model
- Random friends are strange
- Largest component
- Simple, physically-motivated analysis

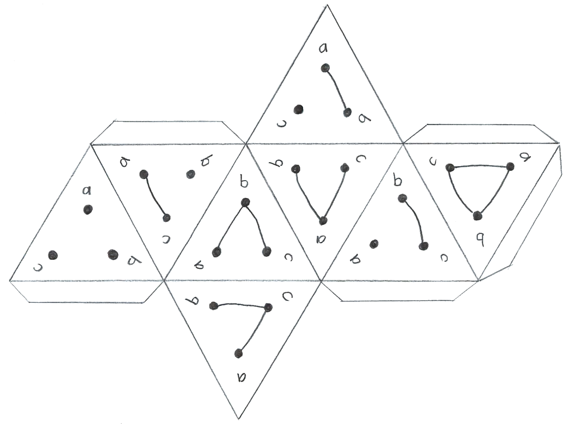
References

Random networks

Pure, abstract random networks:

- Consider set of all networks with N labelled nodes and m edges.
- Standard random network = one 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 network generator for $N = 3$:



- Get your own exciting generator here (⊞).
- As $N \nearrow$, our polyhedral die rapidly becomes a ball...

Random networks—basic features:

- Number of possible edges:

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

- Limit of $m = 0$: empty graph.
- Limit of $m = \binom{N}{2}$: complete or fully-connected graph.
- Number of possible networks with N labelled nodes:

$$2^{\binom{N}{2}} \sim e^{\frac{\ln 2}{2} N^2}.$$

- Given m edges, there are $\binom{\binom{N}{2}}{m}$ different possible networks.
- Crazy factorial explosion for $1 \ll m \ll \binom{N}{2}$.
- 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 relatively 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:

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

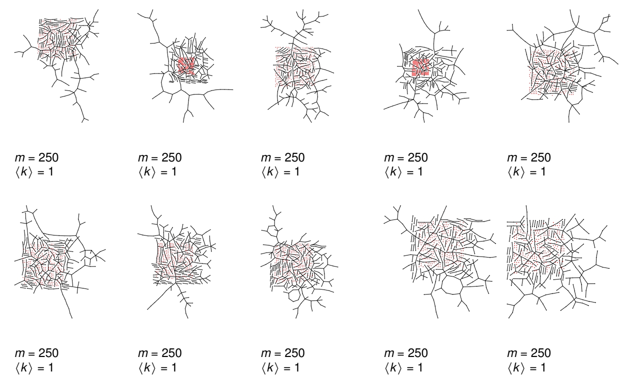
- So the expected or average degree is

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

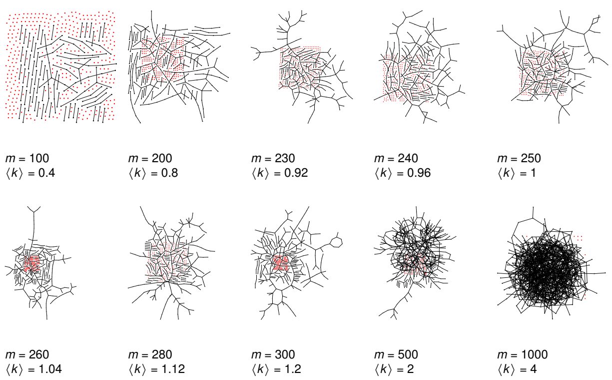
$$= \frac{2}{N}p\frac{1}{2}N(N-1) = \frac{\not{2}}{\not{N}}p\frac{1}{\not{2}}\not{N}(N-1) = p(N-1).$$

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

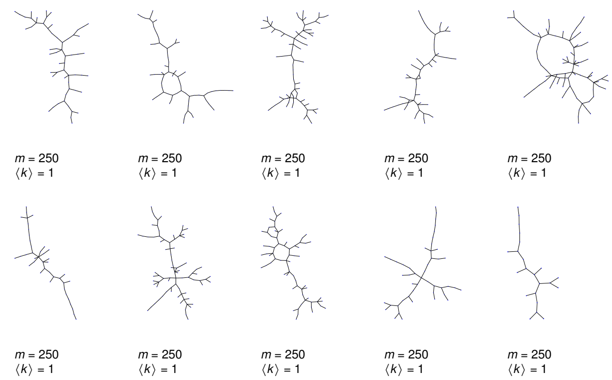
Random networks: examples for N=500



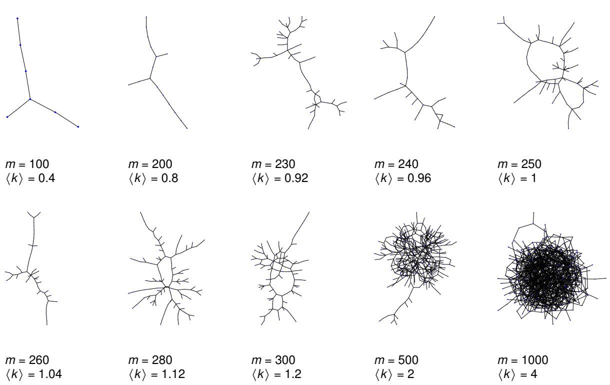
Random networks: examples for N=500



Random networks: largest components



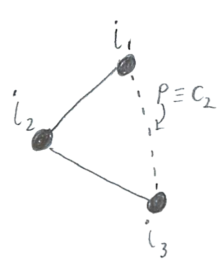
Random networks: largest components



Clustering in random networks:

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

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



- Recall: C_2 = probability that two friends of a node are also friends.
- Or: C_2 = probability that a triple is part of a triangle.
- For standard random networks, we have simply that

$$C_2 = p.$$

Other ways to compute clustering:

- Expected number of triples in entire network:
$$\frac{1}{2}N(N-1)(N-2)p^2$$
(Double counting dealt with by $\frac{1}{2}$.)
- Expected number of triangles in entire network:
$$\frac{1}{6}N(N-1)(N-2)p^3$$
(Over-counting dealt with by $\frac{1}{6}$.)
-
$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}} = \frac{3 \times \frac{1}{6}N(N-1)(N-2)p^3}{\frac{1}{2}N(N-1)(N-2)p^2} = p.$$

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

Other ways to compute clustering:

- Or: take any three nodes, call them a, b, and c.
- Triple a-b-c centered at b occurs with probability $p^2 \times (1-p) + p^2 \times p = p^2$.
- Triangle occurs with probability p^3.
- Therefore,
$$C_2 = \frac{p^3}{p^2} = p.$$

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

Clustering in random networks:

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

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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 $\langle k \rangle = p(N-1)$ = constant.

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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

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

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

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

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

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

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.)

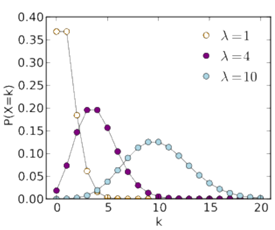
- Identifying $n = N-1$ and $x = -\langle k\rangle$:

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

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

Poisson basics:

$$P(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$$



- $\lambda > 0$
- $k = 0, 1, 2, 3, \ldots$
- Classic use: probability that an event occurs k times in a given time period, given an average rate of occurrence.
- e.g.: phone calls/minute, horse-kick deaths.
- 'Law of small numbers'

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

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

- Checking:

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

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

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

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

- Note: We'll get to a better and crazier 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...

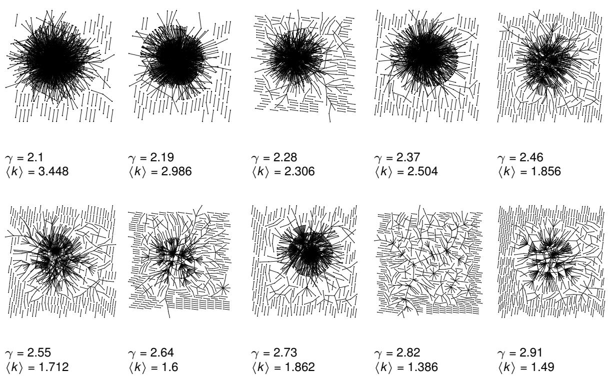
General random networks

- 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

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

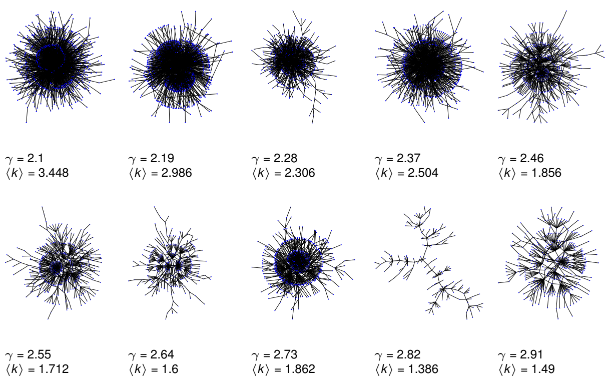
- But we'll be more interested in
  1. Randomly wiring up (and rewiring) already existing nodes with fixed degrees.
  2. Examining mechanisms that lead to networks with certain degree distributions.

Random networks: examples for N=1000



Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

Random networks: largest components



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):

$$\boxed{Q_k \propto kP_k}$$

- Normalized form:

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

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

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

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

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

(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 really are different from 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:

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

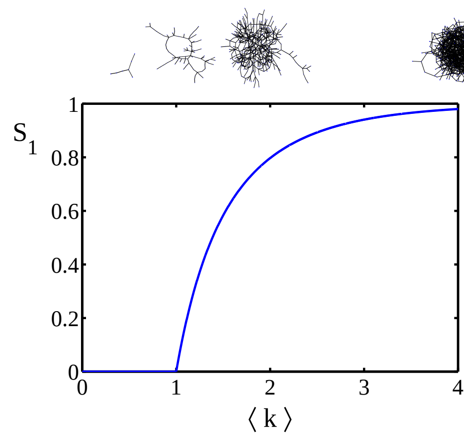
- So only if everyone has the same degree (variance$= \sigma^2 = 0$) can a node be the same as its friends.
- Intuition: for random networks, the more connected a node, the more likely it is to be chosen as a friend.

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 $N \rightarrow \infty$, 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 $N \rightarrow \infty$.
- Note: Component = Cluster

Giant component



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 $2 < \gamma < 3$, $k \geq 1$, then

$$\langle k^2 \rangle = c \sum_{k=1}^{\infty} k^2 k^{-\gamma}$$

$$\sim \int_{x=1}^{\infty} x^{2-\gamma} \mathrm{d}x$$

$$\propto x^{3-\gamma} \Big|_{x=1}^{\infty} = \infty \quad (\gg \langle k \rangle).$$

- 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 $\langle k \rangle_R$.
- How about $P_k = \delta_{kk_0}$?

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...

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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:

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

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

Giant component

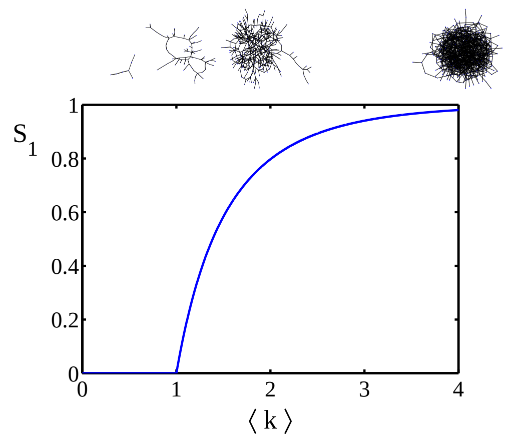
- We can figure out some limits and details for $S_1 = 1 - e^{-\langle k \rangle S_1}$.
- First, we can write $\langle k \rangle$ in terms of S_1:

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

- 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 $S_1 > 0$ when $\langle k \rangle > 1$.
- Really a transcritical bifurcation. [2]

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

Giant component



Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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 an edge leads to the giant (infinite) component.
- We can sort many things out with sensible probabilistic arguments...
- More detailed investigations will profit from a spot of Generatingfunctionology. [3]

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References

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 (⊞)

Random Networks
Basics
Definitions
How to build
Some visual examples
Structure
Clustering
Degree distributions
Configuration model
Random friends are strange
Largest component
Simple, physically-motivated analysis
References