Random Networks Nutshell

Last updated: 2023/08/22, 11:48:21 EDT

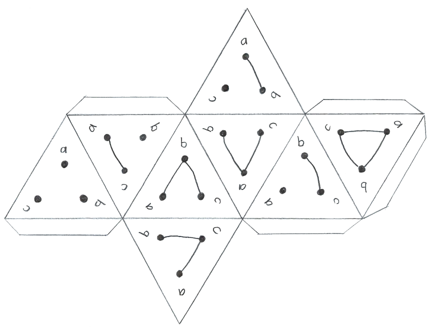
Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number,
2023–2024 | @pocsvox

Prof. Peter Sheridan Dodds | @peterdodds

Computational Story Lab | Vermont Complex Systems Center
Santa Fe Institute | University of Vermont

Outline

Random network generator for $N = 3$:



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

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 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:
$$2^{\binom{N}{2}} \sim e^{\frac{\ln 2}{2} N(N-1)}.$$
- 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:

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

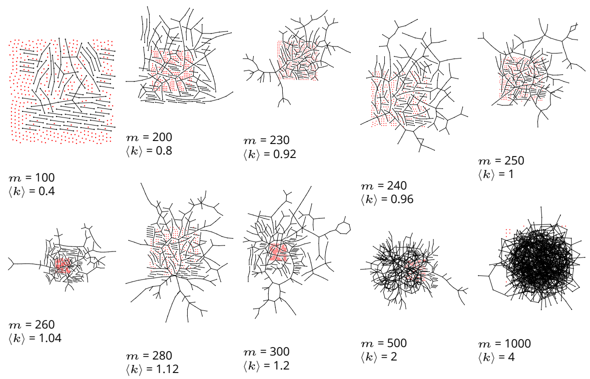
- So the expected or average degree is

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

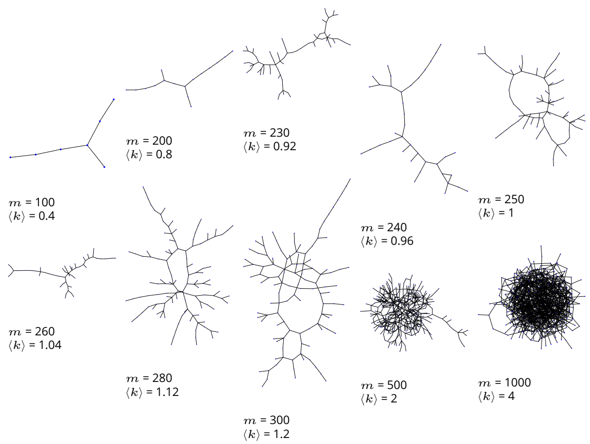
$$= \frac{2}{N}p\frac{1}{2}N(N-1) = \frac{\not 2}{\not N}p\frac{1}{\not 2}\not N(N-1) = p(N-1).$$

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

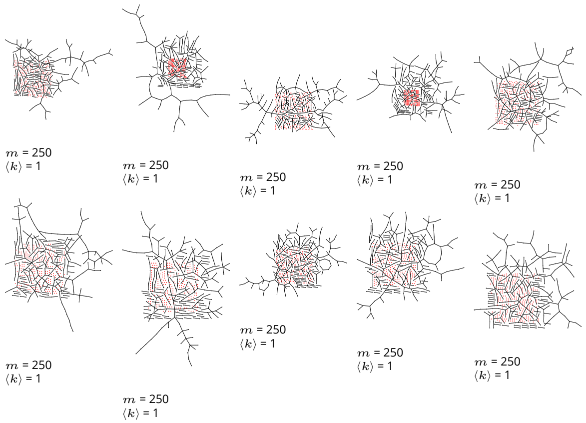
Random networks: examples for N=500



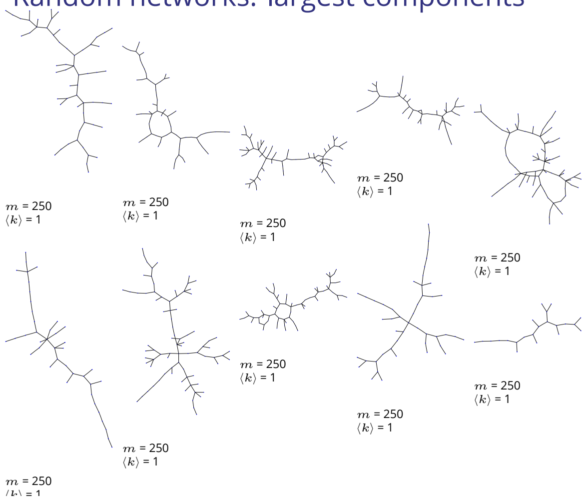
Random networks: largest components



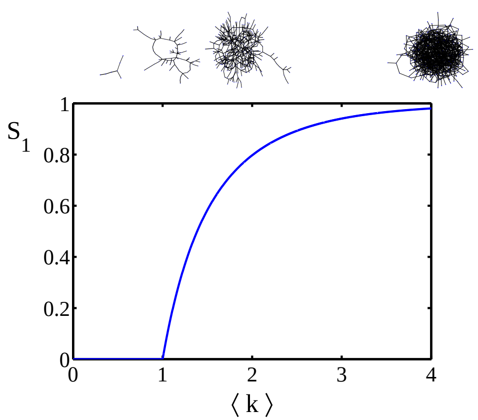
Random networks: examples for N=500



Random networks: largest components



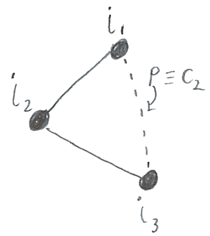
Giant component



Clustering in random networks:

- For construction method 1, what is the clustering coefficient for a finite network?
- Consider triangle/triple clustering coefficient: [6]

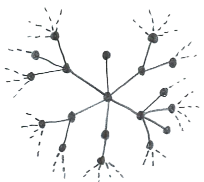
$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$$



- 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

$$C_2 = p.$$

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.

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:

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

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

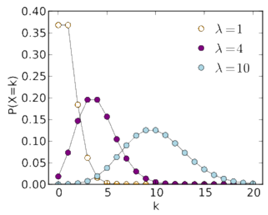
- Our degree distribution: $P(k;p,N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$.
- What happens as $N \to \infty$?
- We must end up with the normal distribution right?
- If p is fixed, then we would end up with a Gaussian with average degree $\langle k \rangle \simeq pN \to \infty$.
- But we want to keep $\langle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

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

- 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:

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

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

- 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...

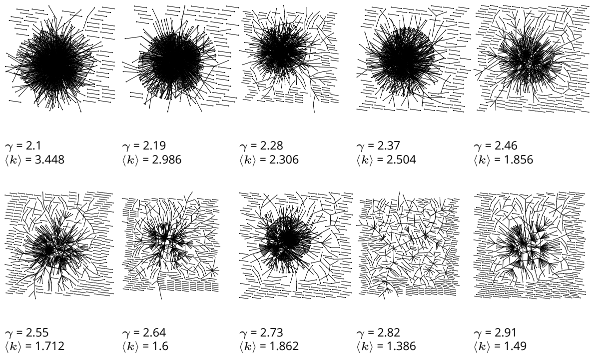
General random networks

- So... standard random networks have a Poisson degree distribution
- Generalize to arbitrary degree distribution P_k.
- Also known as the configuration model. [6]
- Can generalize construction method from ER random networks.
- Assign each node a weight w from some distribution P_w and form links with probability

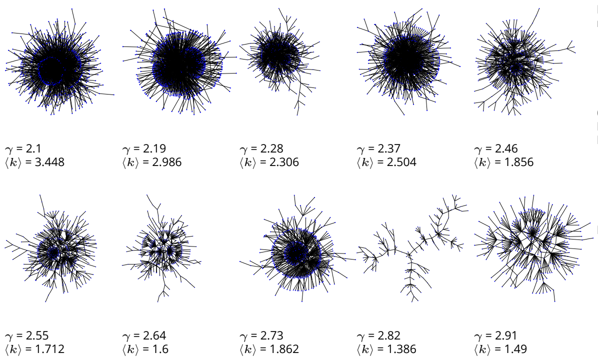
$$P(\text{link between } i \text{ and } j) \propto w_i w_j.$$

- 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: largest components



Models

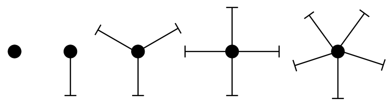
Generalized random networks:

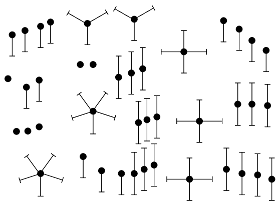
- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Building random networks: Stubs

Phase 1:

- Idea: start with a soup of unconnected nodes with stubs (half-edges):



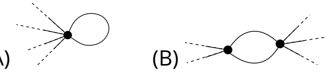


- Randomly select stubs (not nodes!) and connect them.
- Must have an even number of stubs.
- Initially allow self- and repeat connections.

Building random networks: First rewiring

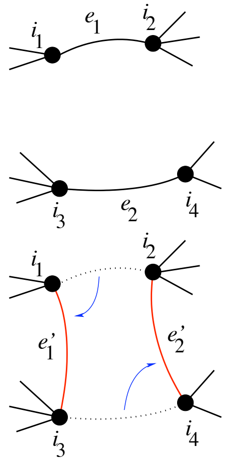
Phase 2:

- Now find any (A) self-loops and (B) repeat edges and randomly rewire them.



- Being careful: we can't change the degree of any node, so we can't simply move links around.
- Simplest solution: randomly rewire two edges at a time.

General random rewiring algorithm



- Randomly choose two edges. (Or choose problem edge and a random edge)
- Check to make sure edges are disjoint.
- Rewire one end of each edge.
- Node degrees do not change.
- Works if e_1 is a self-loop or repeated edge.
- Same as finding on/off/on/off 4-cycles. and rotating them.

Sampling random networks

Phase 2:

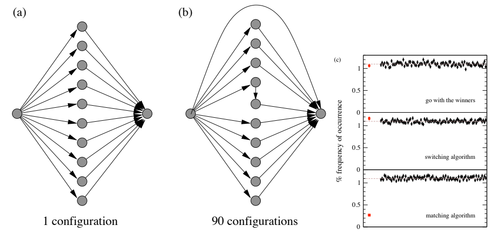
- Use rewiring algorithm to remove all self and repeat loops.

Phase 3:

- Randomize network wiring by applying rewiring algorithm liberally.
- Rule of thumb: # Rewirings $\simeq 10 \times$ # edges [4].

Random sampling

- Problem with only joining up stubs is failure to randomly sample from all possible networks.
- Example from Milo et al. (2003) [4]:



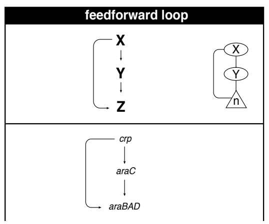
Sampling random networks

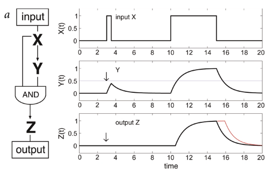
- What if we have P_k instead of N_k?
- Must now create nodes before start of the construction algorithm.
- Generate N nodes by sampling from degree distribution P_k.
- Easy to do exactly numerically since k is discrete.
- Note: not all P_k will always give nodes that can be wired together.

Network motifs

- Idea of motifs [7] introduced by Shen-Orr, Alon et al. in 2002.
- Looked at gene expression within full context of transcriptional regulation networks.
- Specific example of Escherichia coli.
- Directed network with 577 interactions (edges) and 424 operons (nodes).
- Used network randomization to produce ensemble of alternate networks with same degree frequency N_k.
- Looked for certain subnetworks (motifs) that appeared more or less often than expected

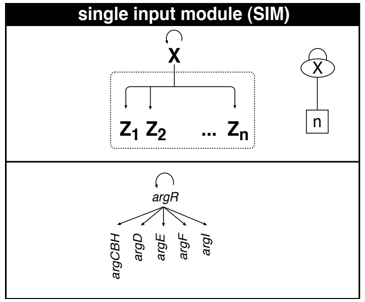
Network motifs





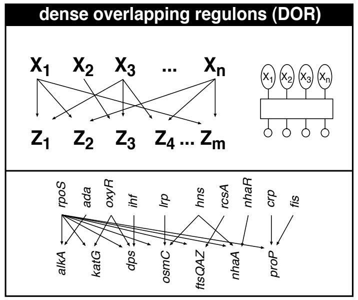
- Z only turns on in response to sustained activity in X.
- Turning off X rapidly turns off Z.
- Analogy to elevator doors.

Network motifs



- Master switch.

Network motifs



Network motifs

- Note: selection of motifs to test is reasonable but nevertheless ad-hoc.
- For more, see work carried out by Wiggins *et al.* at Columbia.

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:

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

- Big deal: Rich-get-richer mechanism is built into this selection process.

- Probability of randomly selecting a node of degree k by choosing from nodes: $P_1 = 3/7$, $P_2 = 2/7$, $P_3 = 1/7$, $P_6 = 1/7$.
- Probability of landing on a node of degree k after randomly selecting an edge and then randomly choosing one direction to travel: $Q_1 = 3/16$, $Q_2 = 4/16$, $Q_3 = 3/16$, $Q_6 = 6/16$.
- Probability of finding # outgoing edges = k after randomly selecting an edge and then randomly choosing one direction to travel: $R_0 = 3/16$ $R_1 = 4/16$, $R_2 = 3/16$, $R_5 = 6/16$.

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...

The edge-degree distribution:

- In fact, R_k is rather special for pure random networks ...
- Substituting

$$P_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

into

$$R_k = \frac{(k+1)P_{k+1}}{\langle k \rangle}$$

we have

$$R_k = \frac{(k+1)}{\langle k \rangle} \frac{\langle k \rangle^{(k+1)}}{(k+1)!} e^{-\langle k \rangle} = \frac{\cancel{(k+1)}}{\cancel{\langle k \rangle}} \frac{\langle k \rangle^{(k+\cancel{1})}}{\cancel{(k+1)}k!} e^{-\langle k \rangle}$$

$$= \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \equiv P_k.$$

- #samesies.

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... [3, 5]
  4. See also: class size paradoxes (nod to: Gelman)

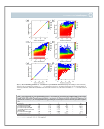
Two reasons why this matters

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.



"Generalized friendship paradox in complex networks: The case of scientific collaboration"
Eom and Jo,
Nature Scientific Reports, **4**, 4603, 2014. [2]

Your friends really are ~~monsters~~ #winners:[1]

- Go on, hurt me: Friends have more coauthors, citations, and publications.
- Other horrific studies: your connections on Twitter have more followers than you, your sexual partners more partners than you, ...
- The hope: Maybe they have more enemies and diseases too.

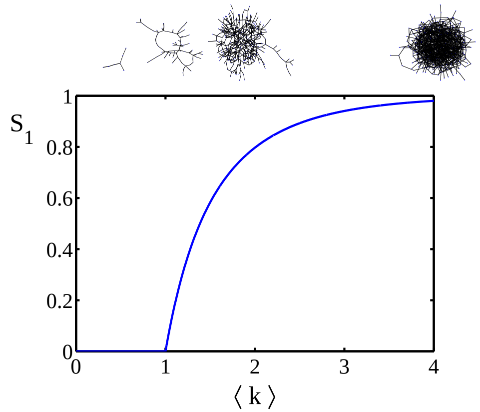
[1] Some press here [MIT Tech Review].

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

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):

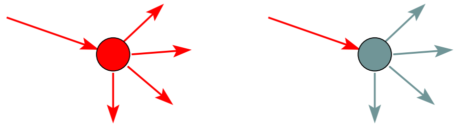
$$\langle k \rangle_R = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1$$

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

Spreading on Random Networks

- For random networks, we know local structure is pure branching.
- Successful spreading is ∴ contingent on single edges infecting nodes.

Success Failure:



- Focus on binary case with edges and nodes either infected or not.
- First big question: for a given network and contagion process, can global spreading from a single seed occur?

Global spreading condition

- We need to find: [1]
  R = the average # of infected edges that one random infected edge brings about.
- Call **R** the gain ratio.
- Define B_{k1} as the probability that a node of degree k is infected by a single infected edge.
-

$$\mathbf{R} = \sum_{k=0}^{\infty} \underbrace{\frac{kP_k}{\langle k \rangle}}_{\substack{\text{prob. of} \\ \text{connecting to} \\ \text{a degree } k \text{ node}}} \bullet \underbrace{(k-1)}_{\substack{\text{\# outgoing} \\ \text{infected} \\ \text{edges}}} \bullet \underbrace{B_{k1}}_{\substack{\text{Prob. of} \\ \text{infection}}}$$

$$+ \sum_{k=0}^{\infty} \overbrace{\frac{kP_k}{\langle k \rangle}} \bullet \underbrace{0}_{\substack{\text{\# outgoing} \\ \text{infected} \\ \text{edges}}} \bullet \underbrace{(1 - B_{k1})}_{\substack{\text{Prob. of} \\ \text{no infection}}}$$

Global spreading condition

- Our global spreading condition is then:

$$\boxed{\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) \bullet B_{k1} > 1.}$$

- Case 1–Rampant spreading: If $B_{k1} = 1$ then

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) = \frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1.$$

- Good: This is just our giant component condition again.

Global spreading condition

- Case 2—Simple disease-like: If $B_{k1} = \beta < 1$ then

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) \bullet \beta > 1.$$

- A fraction (1-β) of edges do not transmit infection.
- Analogous phase transition to giant component case but critical value of $\langle k \rangle$ is increased.
- Aka bond percolation.
- Resulting degree distribution $\tilde{P}_k$:

$$\tilde{P}_k = \beta^k \sum_{i=k}^{\infty} \binom{i}{k} (1-\beta)^{i-k} P_i.$$

Giant component for standard random networks:

- Recall $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$.
- Determine 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.

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...

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:

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

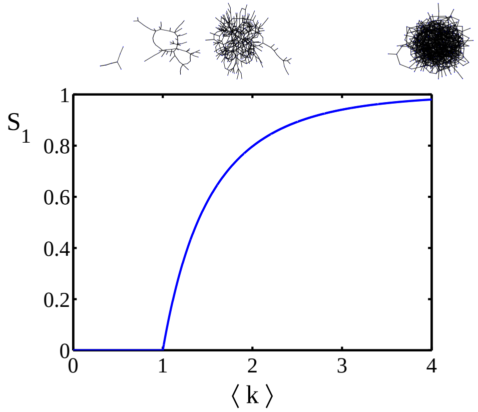
Giant component

- 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 $S_1 > 0$ when $\langle k \rangle > 1$.
- Really a transcritical bifurcation. [8]

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 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. [9]

References I

[1] P. S. Dodds, K. D. Harris, and J. L. Payne.
Direct, phyisically motivated derivation of the contagion condition for spreading processes on generalized random networks.
Phys. Rev. E, 83:056122, 2011. pdf

[2] Y.-H. Eom and H.-H. Jo.
Generalized friendship paradox in complex networks: The case of scientific collaboration.
Nature Scientific Reports, 4:4603, 2014. pdf

[3] S. L. Feld.
Why your friends have more friends than you do.
Am. J. of Sociol., 96:1464–1477, 1991. pdf

References II

[4] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon.
On the uniform generation of random graphs with prescribed degree sequences, 2003. pdf

[5] M. E. J. Newman.
Ego-centered networks and the ripple effect,.
Social Networks, 25:83–95, 2003. pdf

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

[7] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon.
Network motifs in the transcriptional regulation network of *Escherichia coli*.
Nature Genetics, 31:64–68, 2002. pdf

References III

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

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