Network models: random graphs

Yury Dvorkin

February 11, 2020

Empirical network features:

- Power-law (heavy-tailed) degree distribution
- Small average distance (graph diameter)
- Large clustering coeffecient (transitivity)
- Giant connected component, hierarchical structure,etc

Generative models:

- Random graph model (Erdos & Renyi, 1959)
- "Small world" model (Watts & Strogatz, 1998)
- Preferential attachement model (Barabasi & Albert, 1999)

Well-known property:

$$\text{If } a \circ b \;\&\; b \circ c, \text{ then } a \circ c,$$

where $\circ$ is a transitive operator. Example:

$$\text{If } a = b \;\&\; b = c, \text{ then } a = c$$

Graph extension is for edges:

$$\text{If nodes } a \circ b \;\&\; \text{nodes } b \circ c, \text{ then nodes } a \circ c,$$

where $\circ$ is a transitive operator (e.g. connection) Graph example of transitivity:

- Full: a friend of my friend is my friend
- Partial: a friend of my friend is not necessarily my friend

Well-known property:

$$\text{If } a \circ b \;\&\; b \circ c, \text{ then } a \circ c,$$

where $\circ$ is a transitive operator. Example:

$$\text{If } a = b \;\&\; b = c, \text{ then } a = c$$

Graph extension is for edges:

$$\text{If nodes } a \circ b \;\&\; \text{nodes } b \circ c, \text{ then nodes } a \circ c,$$

where $\circ$ is a transitive operator (e.g. connection) Graph example of transitivity:

- Full: a friend of my friend is my friend
- Partial: a friend of my friend is not necessarily my friend

Well-known property:

$$\text{If } a \circ b \ \& \ b \circ c, \text{ then } a \circ c,$$

where $\circ$ is a transitive operator. Example:

$$\text{If } a = b \ \& \ b = c, \text{ then } a = c$$

Graph extension is for edges:

$$\text{If nodes } a \circ b \ \& \text{ nodes } b \circ c, \text{ then nodes } a \circ c,$$

where $\circ$ is a transitive operator (e.g. connection) Graph example of transitivity:

- Full: a friend of my friend is my friend
- Partial: a friend of my friend is not necessarily my friend

Well-known property:

$$\text{If } a \circ b \ \&\ b \circ c, \text{ then } a \circ c,$$

where $\circ$ is a transitive operator. Example:

$$\text{If } a = b \ \&\ b = c, \text{ then } a = c$$

Graph extension is for edges:

$$\text{If nodes } a \circ b \ \&\ \text{nodes } b \circ c, \text{ then nodes } a \circ c,$$

where $\circ$ is a transitive operator (e.g. connection) Graph example of transitivity:

- Full: a friend of my friend is my friend
- Partial: a friend of my friend is not necessarily my friend

How can transitivity be measured?

- Only relative to all two-edge paths within a given graph
- Using *closed* two-edge paths: closed path forms a loop of three nodes
- Clustering coefficient:

$$C = \frac{\text{N of closed two-edge paths}}{\text{N of all two-edge paths}}$$

- Interpretations:
  - C = 0: full transitivity
  - C = 1: radial/tree topology

How can transitivity be measured?

- Only relative to all two-edge paths within a given graph
- Using *closed* two-edge paths: closed path forms a loop of three nodes
- Clustering coefficient:

$$C = \frac{\text{N of closed two-edge paths}}{\text{N of all two-edge paths}}$$

- Interpretations:
  - C = 0: full transitivity
  - C = 1: radial/tree topology

How can transitivity be measured?

- Only relative to all two-edge paths within a given graph
- Using *closed* two-edge paths: closed path forms a loop of three nodes
- Clustering coefficient:

$$C = \frac{\text{N of closed two-edge paths}}{\text{N of all two-edge paths}}$$

- Interpretations:
  - $C = 0$: full transitivity
  - $C = 1$: radial/tree topology

Why generative models?

- Move from analyzing statistical properties to analyzing network behavior?
- Need to simulate behavior given a property (e.g. power laws)
- Hence, random grpahs as a starting point

Simple example: Given n nodes and m edges, but edges are random

- Can we select a subgraph?
- Can we represent a graph as an ensemble of subgraphs?
- How to obtain a probabilistic representation of G in terms of subgraphs?
  - Properties of the random graph = *statistical* properties of the ensemble of subgraphs

Simple example: Given n nodes and m edges, but edges are random

- Can we select a subgraph?
- Can we represent a graph as an ensemble of subgraphs?
- How to obtain a probabilistic representation of G in terms of subgraphs?
  - Properties of the random graph = *statistical* properties of the ensemble of subgraphs

Simple example: Given n nodes and m edges, but edges are random

- Can we select a subgraph?
- Can we represent a graph as an ensemble of subgraphs?
- How to obtain a probabilistic representation of G in terms of subgraphs?
  - Properties of the random graph = *statistical* properties of the ensemble of subgraphs

Simple example: Given n nodes and m edges, but edges are random

- Can we select a subgraph?
- Can we represent a graph as an ensemble of subgraphs?
- How to obtain a probabilistic representation of G in terms of subgraphs?
  - Properties of the random graph = *statistical* properties of the ensemble of subgraphs

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution:

$$P(G) = \frac{1}{\binom{\binom{n}{2}}{m}}$$

  - Diameter of the graph:

$$< d >= \sum_G P(G)d(G), \text{ where } d(G) \text{ is given}$$

  - Assume that instead of m, we parameterize the probability of edge p

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution:

$$P(G) = \frac{1}{(\frac{(\frac{n}{2})}{m})}$$

  - Diameter of the graph:

$$< d >= \sum_G P(G)d(G), \text{ where } d(G) \text{ is given}$$

  - Assume that instead of m, we parameterize the probability of edge p

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution:

$$P(G) = \frac{1}{(\frac{(\frac{n}{2})}{m})}$$

  - Diameter of the graph:

$$< d >= \sum_G P(G)d(G), \text{ where } d(G) \text{ is given}$$

  - Assume that instead of m, we parameterize the probability of edge p

Graph $G\{E,V\}$, nodes $n=|V|$, edges $m=|E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N=\frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N=\frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution:

$$P(G)=\frac{1}{\binom{\binom{n}{2}}{m}}$$

  - Diameter of the graph:

$$<d>=\sum_G P(G)d(G), \text{ where } d(G) \text{ is given}$$

  - Assume that instead of m, we parameterize the probability of edge p

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution with parameterized m:

$$P(G) = \frac{1}{\binom{\binom{n}{2}}{m}}$$

  - Probability distribution with parameterized p: **(Erdos-Renyi random graph)**:

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

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution with parameterized m:

$$P(G) = \frac{1}{\binom{\binom{n}{2}}{m}}$$

  - Probability distribution with parameterized p: **(Erdos-Renyi random graph)**:

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

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$, Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
  - Probability distribution with parameterized m:

$$P(G) = \frac{1}{\binom{\binom{n}{2}}{m}}$$

  - Probability distribution with parameterized p: **(Erdos-Renyi random graph)**:

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

Graph $G\{E,V\}$, nodes $n=|V|$, edges $m=|E|$ Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N=\frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N=\frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
- What properties should a random graph have?
- Number of edges (m) and of degrees (k), as well as density (ρ)

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

$$\langle k\rangle = \frac{1}{n}\sum_i k_i = \frac{2\langle m\rangle}{n} = p(n-1)\approx pn$$

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

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$ Erdos and Renyi, 1959. Random graph models

- $G_{n,m}$, a randomly selected graph from the set of G_N^mgraphs;$N = \frac{n(n-1)}{2}$,with n nodes and m edges
- Random graph = ensemble of subgraphs
- $G_{n,q}$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number
- What properties should a random graph have?
- Number of edges (m) and of degrees (k), as well as density (ρ)

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

$$\langle k \rangle = \frac{1}{n}\sum_i k_i = \frac{2\langle m \rangle}{n} = p(n-1) \approx pn$$

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

- Probability that i-th node has a degree is connected to k other nodes is $p^k(1-p)^{n-1-k}$
- Numbers of way to choose k out of n nodes is $\frac{n-1}{k}$
- What is the probability that node i is connected to k nodes?

$$p_k \frac{n-1}{k} p^k (1-p)^{n-1-k}$$

(binominal distribution, i.e a Bernoulli distribution)

- Probability that i-th node has a degree is connected to k other nodes is $p^k(1-p)^{n-1-k}$
- Numbers of way to choose k out of n nodes is $\frac{n-1}{k}$
- What is the probability that node i is connected to k nodes?

$$p_k \frac{n-1}{k} p^k (1-p)^{n-1-k}$$

(binominal distribution, i.e a Bernoulli distribution)

- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

(Bernoulli distribution)
p^k probability that connects to k nodes (has k-edges)
$(1-p)^{n-1-k}$ probability that does not connect to any other node
C_{n-1}^{k} number of ways to select k nodes out of all to connect

- Interesting property of $ln\Big[(1-p)^{n-1-k}\Big]$:

$$ln\Big[(1-p)^{n-1-k}\Big] = n-1-k\,ln\Big[(1-p)\Big]$$

Recall that $p = \frac{c}{n-1}$ and hence $ln\Big[(1-p)\Big] = ln\Big[1 - \frac{c}{n-1}\Big]$
Use the Taylor's approximation:

$$ln\Big[(1-p)^{n-1-k}\Big] \approx -(n-1-k)\frac{c}{n-1}$$

- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

(Bernoulli distribution)
p^k probability that connects to k nodes (has k-edges)
$(1-p)^{n-1-k}$ probability that does not connect to any other node
C_{n-1}^{k} number of ways to select k nodes out of all to connect

- Interesting property of $ln\Big[(1-p)^{n-1-k}\Big]$:

$$ln\Big[(1-p)^{n-1-k}\Big] = n - 1 - k ln\Big[(1-p)\Big]$$

Recall that $p = \frac{c}{n-1}$ and hence $ln\Big[(1-p)\Big] = ln\Big[1 - \frac{c}{n-1}\Big]$
Use the Taylor's approximation:

$$ln\Big[(1-p)^{n-1-k}\Big] \approx -(n-1-k)\frac{c}{n-1}$$

- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

(Bernoulli distribution)
p^k probability that connects to k nodes (has k-edges)
$(1-p)^{n-1-k}$ probability that does not connect to any other node
C_{n-1}^{k} number of ways to select k nodes out of all to connect

- Interesting property of $ln\Big[(1-p)^{n-1-k}\Big]$:

$$ln\Big[(1-p)^{n-1-k}\Big] = n - 1 - k\,ln\Big[(1-p)\Big]$$

Recall that $p = \frac{c}{n-1}$ and hence $ln\Big[(1-p)\Big] = ln\Big[1 - \frac{c}{n-1}\Big]$
Use the Taylor's approximation:

$$ln\Big[(1-p)^{n-1-k}\Big] \approx -(n-1-k)\frac{c}{n-1}$$

How will

$$ln\Big[(1-p)^{n-1-k}\Big] \approx -(n-1-k)\frac{c}{n-1}$$

behave if $n \to \infty$?

- No general rule!
- But if we fix k, the exact equality follows $ln\Big[(1-p)^{n-1-k}\Big] = -(n-1-k)\frac{c}{n-1} = -c$

Now we have a new assumption ($n \to \infty$?). How will it affect C_{n-1}^{k}?

- $C_{n-1}^{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \Big\{n \to \infty\Big\} \approx \frac{(n-1)^k}{k!}$

How will

$$ln\Big[(1-p)^{n-1-k}\Big] \approx -(n-1-k)\frac{c}{n-1}$$

behave if $n \to \infty$?

- No general rule!
- But if we fix k, the exact equality follows $ln\Big[(1-p)^{n-1-k}\Big] = -(n-1-k)\frac{c}{n-1} = -c$

Now we have a new assumption ($n \to \infty$?). How will it affect C_{n-1}^{k}?

- $C_{n-1}^{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \big\{n \to \infty\big\} \approx \frac{(n-1)^k}{k!}$

How will

$$ln\left[(1-p)^{n-1-k}\right] \approx -(n-1-k)\frac{c}{n-1}$$

behave if $n \to \infty$?

- No general rule!
- But if we fix k, the exact equality follows $ln\left[(1-p)^{n-1-k}\right] = -(n-1-k)\frac{c}{n-1} = -c$

Now we have a new assumption ($n \to \infty$?). How will it affect C_{n-1}^{k}?

- $C_{n-1}^{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \left\{n \to \infty\right\} \approx \frac{(n-1)^k}{k!}$

- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

(Bernoulli distribution)
p^k probability that connects to k nodes (has k-edges)
$(1-p)^{n-1-k}$ probability that does not connect to any other node
C_{n-1}^{k} number of ways to select k nodes out of all to connect

- Limiting case of Bernoulli distribution, when $n \to \infty$ at fixed $\langle k \rangle = pn = \lambda$

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

(Poisson distribution)

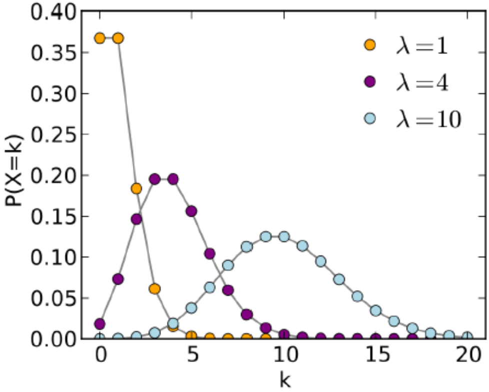
- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

  (Bernoulli distribution)
  p^k probability that connects to k nodes (has k-edges)
  $(1-p)^{n-1-k}$ probability that does not connect to any other node
  C_{n-1}^{k} number of ways to select k nodes out of all to connect
- Limiting case of Bernoulli distribution, when $n \to \infty$ at fixed $\langle k \rangle = pn = \lambda$

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

  (Poisson distribution)



$$P(k_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = pn$$

- For $n \to \infty$ at fixed $\langle k \rangle = pn = \lambda$, we have the following Poisson distribution:

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

- Useful for the clustering coefficient:

$$C = \frac{\text{N of closed two-edge paths}}{\text{N of all two-edge paths}}$$

- Recall that $p = \frac{c}{n-1}$, which leads to

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

- Important: when $n \to \infty$ we have $C \to 0$ (unlike in many real networks)

- For $n \to \infty$ at fixed $\langle k \rangle = pn = \lambda$, we have the following Poisson distribution:

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

- Useful for the clustering coefficient:

$$C = \frac{\text{N of closed two-edge paths}}{\text{N of all two-edge paths}}$$

- Recall that $p = \frac{c}{n-1}$, which leads to

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

- Important: when $n \to \infty$ we have $C \to 0$ (unlike in many real networks)

Consider $G_{n,p}$ as a function of p:

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Consider $G_{n,p}$ as a function of p:

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Consider $G_{n,p}$ as a function of p:

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Consider $G_{n,p}$ as a function of p:

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Consider $G_{n,p}$ as a function of p:

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Consider $G_{n,p}$ as a function of p:

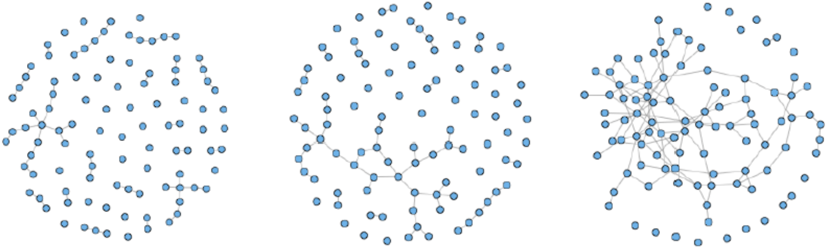
$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\langle k \rangle = pn = \lambda$

- $p = 0$, empty graph (means no edges and you have n separate components)
  - Each component has cardinality of 1.
- $p = 1$, complete (full) graph (means nodes are interconnected and you have one giant component)
  - The giant component has cardinality of n.
- There exist critical p_c , structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

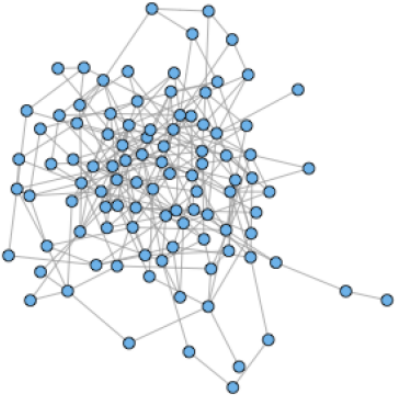
A more formal definition of the giant component:

- Component of the network
- This component grows in size in proportion to n



$p < p_c$ $p = p_c$ $p > p_c$

Any real life examples come to mind?



$$p \gg p_c$$

Let u be a fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

$$u = P(k=0) + P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \cdots =$$

$$= \sum_{k=0}^{\infty} P(k)u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

(Recall that $\langle k \rangle = pn = \lambda$)

Let s be a fraction of nodes belonging to GCC (size of GCC)

$$s = 1 - u$$

$$1 - s = u = e^{\lambda(u-1)} = e^{-\lambda s}$$

when $\lambda \to \infty, s \to 1$

when $\lambda \to 0, s \to 0$

Let u be a fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

$$u = P(k=0) + P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \cdots =$$

$$= \sum_{k=0}^{\infty} P(k) u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

(Recall that $\langle k \rangle = pn = \lambda$)

Let s be a fraction of nodes belonging to GCC (size of GCC)

$$s = 1 - u$$

$$1 - s = u = e^{\lambda(u-1)} = e^{-\lambda s}$$

when $\lambda \to \infty, s \to 1$

when $\lambda \to 0, s \to 0$

Let u be a fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

$$u = P(k=0) + P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \cdots =$$

$$= \sum_{k=0}^{\infty} P(k)u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

(Recall that $\langle k \rangle = pn = \lambda$)

Let s be a fraction of nodes belonging to GCC (size of GCC)

$$s = 1 - u$$

$$1 - s = u = e^{\lambda(u-1)} = e^{-\lambda s}$$

when $\lambda \to \infty, s \to 1$

when $\lambda \to 0, s \to 0$

Let u be a fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

$$u = P(k=0) + P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \cdots =$$

$$= \sum_{k=0}^{\infty} P(k)u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

(Recall that $\langle k \rangle = pn = \lambda$)

Let s be a fraction of nodes belonging to GCC (size of GCC)

$$s = 1 - u$$

$$1 - s = u = e^{\lambda(u-1)} = e^{-\lambda s}$$

when $\lambda \to \infty, s \to 1$

when $\lambda \to 0, s \to 0$

Let u be a fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

$u = P(k=0) + P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \cdots =$

$$= \sum_{k=0}^{\infty} P(k) u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

(Recall that $\langle k \rangle = pn = \lambda$)

Let s be a fraction of nodes belonging to GCC (size of GCC)

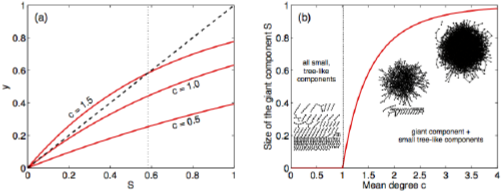
$$s = 1 - u$$

$$1 - s = u = e^{\lambda(u-1)} = e^{-\lambda s}$$

when $\lambda \to \infty, s \to 1$

when $\lambda \to 0, s \to 0$

$$s = 1 - e^{-\lambda s}$$



non-zero solution exists when $(at\ s = 0)$:

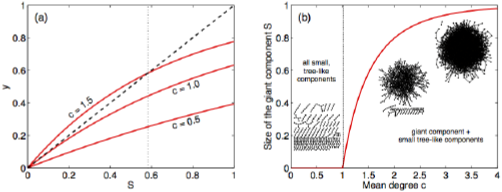
$$\lambda e^{-\lambda s} > 1$$

critical value:

$$\lambda_c = 1$$

$$\lambda_c = \langle k \rangle = p_c n = 1, p_c = \frac{1}{n}$$

$$s = 1 - e^{-\lambda s}$$



non-zero solution exists when $(at\ s = 0)$:

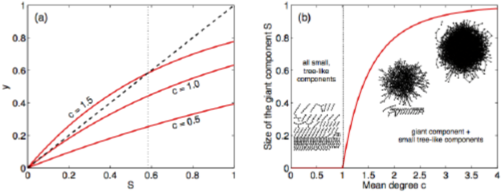
$$\lambda e^{-\lambda s} > 1$$

critical value:

$$\lambda_c = 1$$

$$\lambda_c = \langle k \rangle = p_c n = 1, p_c = \frac{1}{n}$$

$$s = 1 - e^{-\lambda s}$$



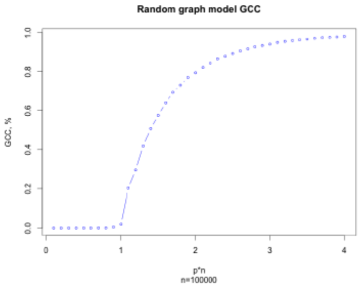
non-zero solution exists when $(at\ s = 0)$:

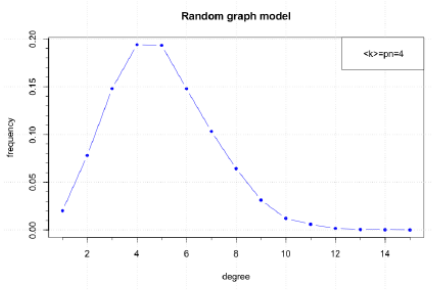
$$\lambda e^{-\lambda s} > 1$$

critical value:

$$\lambda_c = 1$$

$$\lambda_c = \langle k \rangle = p_c n = 1, p_c = \frac{1}{n}$$





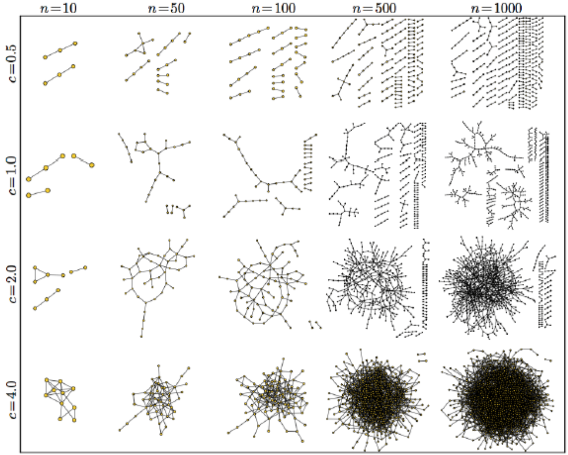
$$\langle k \rangle = pn$$

Graph $G(n, p)$, for $n \to \infty$, critical value $p_c = \frac{1}{n}$

- when $p < p_c$, ($\langle k \rangle < 1$) there is no components with more than $O(\ln n)$ nodes, largest component is a tree
- when $p = p_c$, ($\langle k \rangle = 1$) the largest component has $O(n^{2/3})$ nodes
- when $p > p_c$, ($\langle k \rangle > 1$) gigantic component has all $O(n)$ nodes

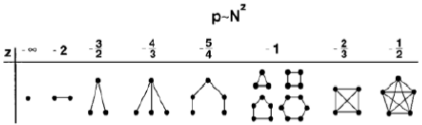
Critical value: $\langle k \rangle = p_c n = 1$- on average one neighbor for a node

Another interpretation: *core* vs *periphery*



Clauset, 2014

Graph $G(n, p)$
Threshold probabilities when different subgraphs of k-nodes and l-edges appear in a random graph $p_c \sim n^{-k/l}$



When $p > p_c$:

- $p_c \sim n^{-k/(k-1)}$,having a tree with k nodes
- $p_c \sim n^{-1}$, having a cycle with k nodes
- $p_c \sim n^{-2/(k-1)}$, complete subgraph with k nodes

Barabasi, 2002

- Clustering coefficient (probability that two neighbors link to each other):

$$C_i k = \frac{\#of\ links\ between\ NN}{\#max\ number\ of\ links\ NN} = \frac{pk(k-1)/2}{k(k-1)/2} = p$$

$$C = p = \frac{\langle k \rangle}{n}$$

- when $n \to \infty,\ C \to 0$

- $G(n, p)$ is locally tree-like (GCC) (no loops; low clustering coefficient)

- on average, the number of nodes d steps away from a node

$$n = 1 + \langle k \rangle + \langle k \rangle^2 + \cdots \langle k \rangle^D = \frac{\langle k \rangle^{D+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^D$$

- in GCC, around p_c, $\langle k \rangle^D \sim n$,

$$D \sim \frac{\ln n}{\ln \langle k \rangle}$$

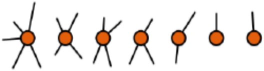
G(n,p) model has many problems:

- Node degree distribution - Poisson, not a power law distribution:
$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = pn$$
- Clustering coefficient - small (no transitivity, unlike in many real networks):
$$C = p$$
- Graph diameter - small (no correlation among connected nodes):
$$D \sim \ln n$$

- Key idea: replace a degree distribution with a degree sequence
- In other words: specify an exact degree of each node, rather than a degree distribution
- Random graph with n nodes with a given degree sequence: $D = \{k_1, k_2, k_3 \cdots k_n\}$, and $m = 1/2 \sum_i k_i$ edges is fixed
- Configuration model is somewhere between $G(n, m)$ and $G(n, p)$, but more accurate than the latter

- Key idea: replace a degree distribution with a degree sequence
- In other words: specify an exact degree of each node, rather than a degree distribution
- Random graph with n nodes with a given degree sequence: $D = \{k_1, k_2, k_3 \cdots k_n\}$, and $m = 1/2 \sum_i k_i$ edges is fixed
- Configuration model is somewhere between $G(n, m)$ and $G(n, p)$, but more accurate than the latter

- Key idea: replace a degree distribution with a degree sequence
- In other words: specify an exact degree of each node, rather than a degree distribution
- Random graph with n nodes with a given degree sequence: $D = \{k_1, k_2, k_3 \cdots k_n\}$, and $m = 1/2 \sum_i k_i$ edges
- Construct subgraphs by randomly matching two stubs and connecting them by an edge.



- Can contain self loops and multiple edges
- Probability that two nodes i and j are connected

$$p_{ij} = \frac{k_i k_j}{2m - 1}$$

- Will be a simple graph for special "graphical degree sequence"

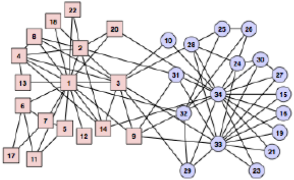
- Probability that two nodes i and j are connected

$$p_{ij} = \frac{k_i k_j}{2m - 1}$$

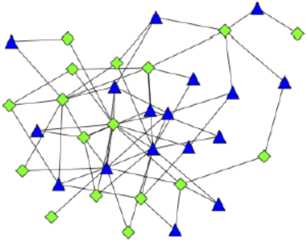
- When we have a large number of edges, i.e. $m \to \infty$:

$$p_{ij} = \frac{k_i k_j}{2m}$$

Can be used as a "null model" for comparative network analysis



karate club



configuration model

Clauset, 2014

- On random graphs I, P. Erdos and A. Renyi, Publicationes Mathematicae 6, 290297 (1959).
- On the evolution of random graphs, P. Erdos and A. Renyi, Publicaton of the Mathematical Institute of the Hungarian Academy of Sciences, 17-61 (1960)