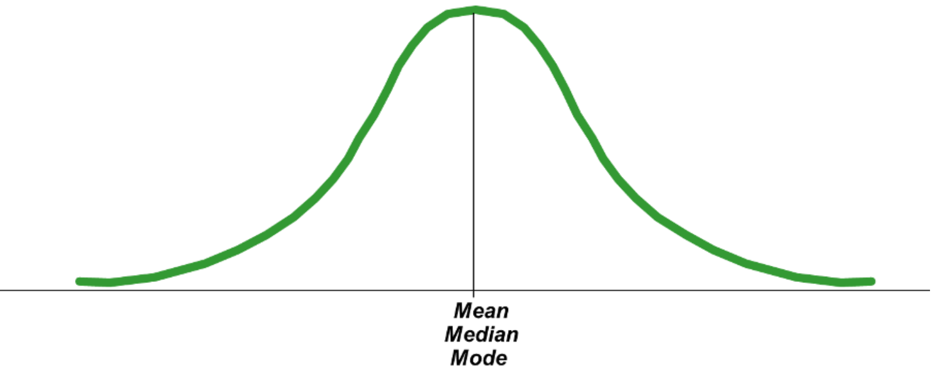
Statistical Graph Model

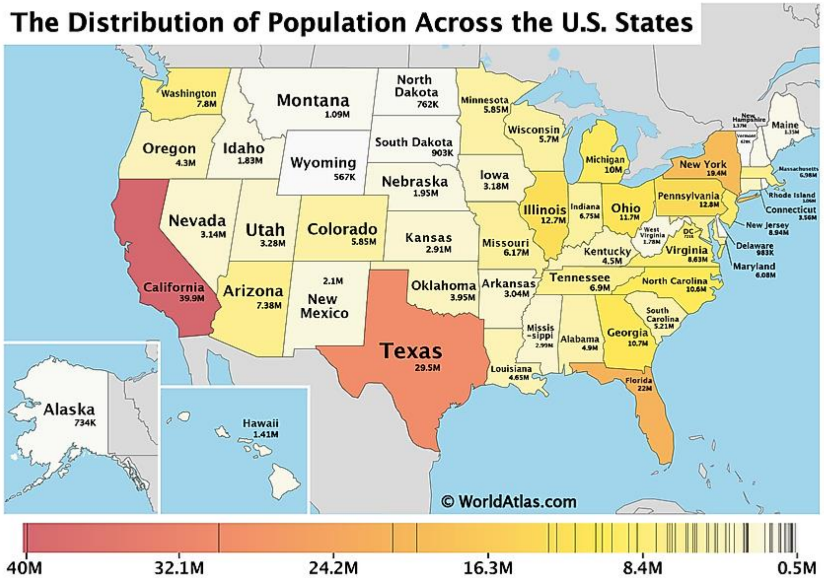
Random Graph, Small World Model

Yu Wang, Ph.D.
Assistant Professor
Computer and Information Science
University of Oregon
CS 410/510 - Fall 2024

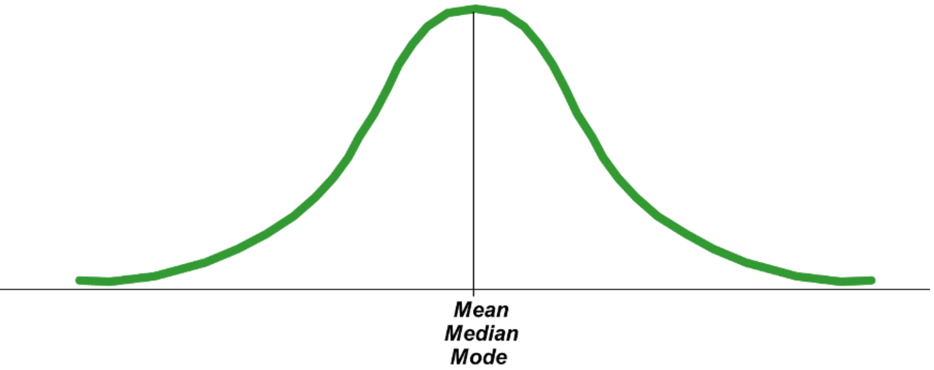
Why Statistical Analysis?



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

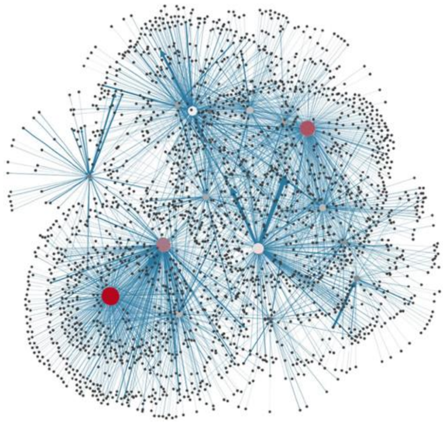


Why Statistical Analysis?



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Why Statistical Network Analysis?



What is the average degree?

How much percentage of users have degree less than 1?

Statistical Graph Model

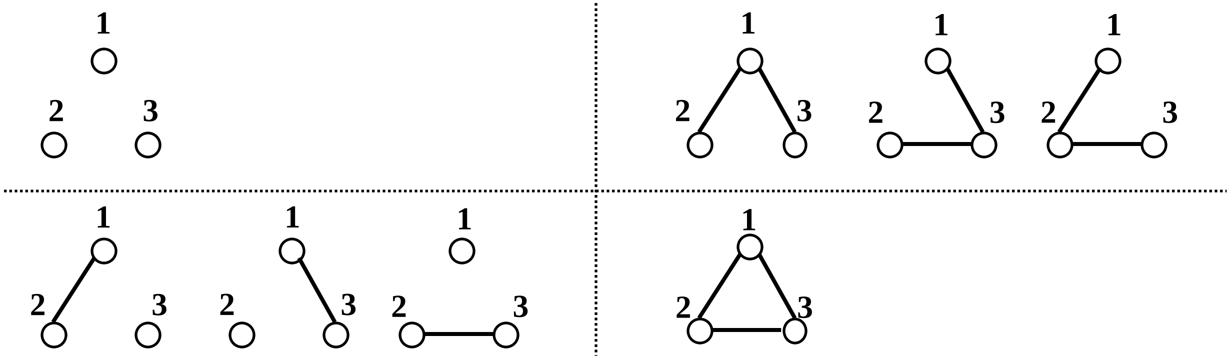
- Random Graph Model
- Small World Model
- Barabasi-Albert Model
- Other Models

Statistical Graph Model

- **Random Graph Model**
- **Small World Model**
- Barabasi-Albert Model
- Other Models

In $G(n,m)$, a graph is chosen uniformly at random from the collection of all graphs which have n nodes and m edges

What does graph with n nodes look like?



How many edges would the graph with n nodes potentially have?

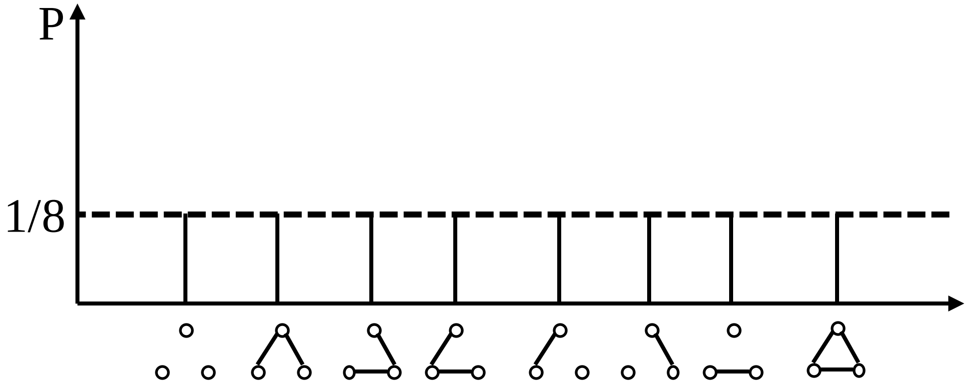
$$0, 1, 2, \ldots C_n^2$$

If the graph with n nodes also have m edges, how many possibilities are there?

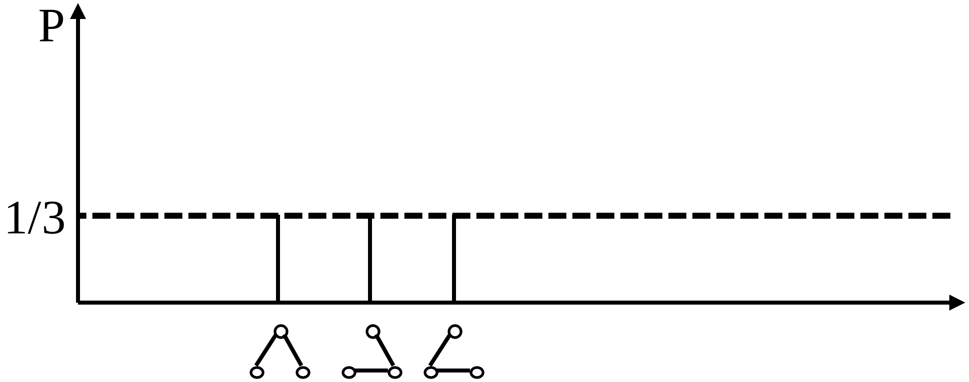
$$C_{C_n^2}^{m}$$

Statistical Graph Model – Random Graph Models

What is the distribution of graph with 3 nodes

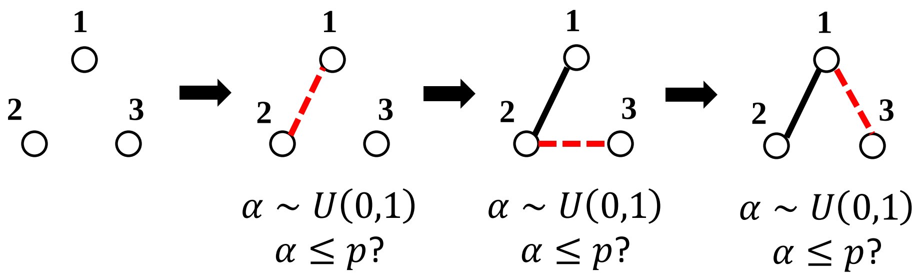


What is the distribution of graph with 3 nodes and 2 edges



Statistical Graph Model – Random Graph Models – $G(n,p)$

In $G(n,p)$, a graph is constructed by connecting labeled nodes randomly. Each edge is included in the graph with probability p.



How many steps should we consider?

$$C_n^2 = \frac{n(n-1)}{2}$$

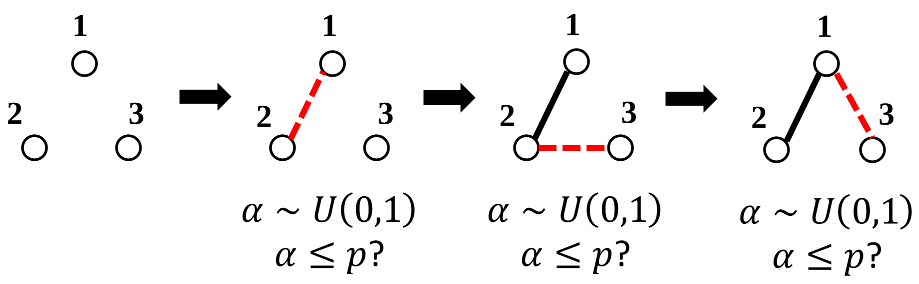
$G(n,p)$: a random graph with totally n nodes and among each pair of nodes, the edge is added with the probability of p

What is the expectation of the number of edges?

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

Statistical Graph Model – Random Graph Models – $G(n,p)$

In $G(n,p)$, a graph is constructed by connecting labeled nodes randomly. Each edge is included in the graph with probability p.



How many steps should we consider?

$$C_n^2 = \frac{n(n-1)}{2}$$

$G(n,p)$: a random graph with totally n nodes and among each pair of nodes, the edge is added with the probability of p

What is the expectation of the average degree?

$$\bar{k} = \frac{2\bar{m}}{n} = 2 * \frac{p\frac{n(n-1)}{2}}{n} = p(n-1)$$

Statistical Graph Model – Random Graph Models – $G(n, p)$

What is the probability that a node i has a degree $d_i = k$?

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

Bernoulli distribution

p^k - probability that connects k nodes (i.e., has k-edges)

$(1 - p)^{(n-1-k)}$ - probability that does not connect to any other node of the n-1 other nodes (i.e., since no self loops)

C_{n-1}^{k} - number of ways to select k nodes from the other n-1

- Limiting case of Bernoulli distribution, when $n \to \infty$ if we fix <k> = pn= λ

Proof

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

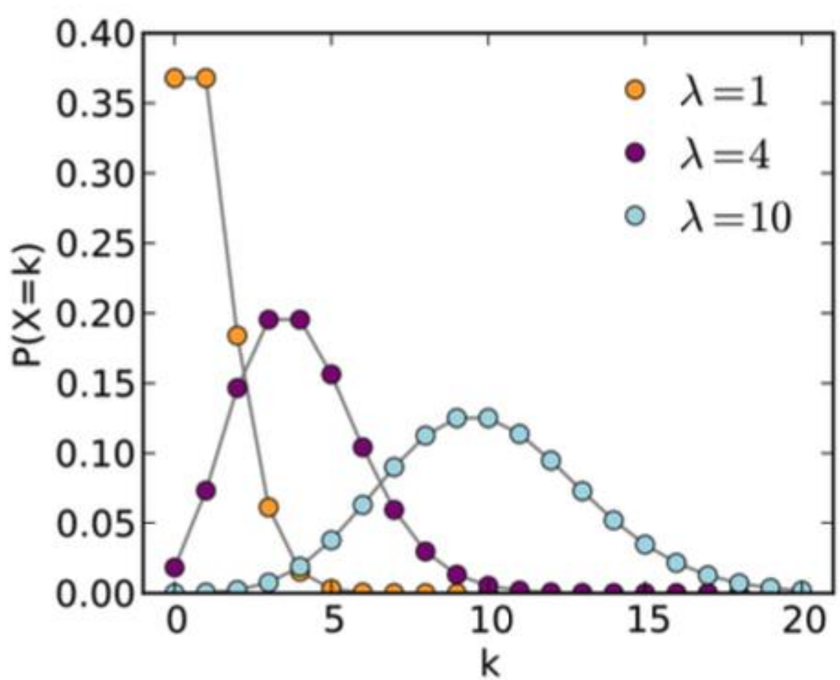
Poisson distribution from Binomial Distribution

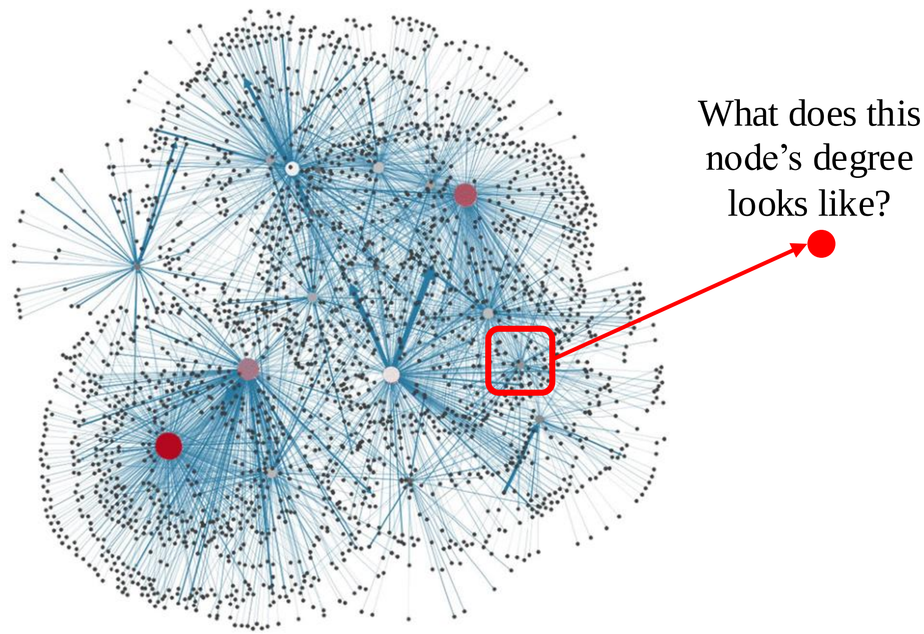
Statistical Graph Model – Random Graph Models – $G(n, p)$

- Limiting case of Bernoulli distribution, when $n \to \infty$ if we fix <k> = pn= λ

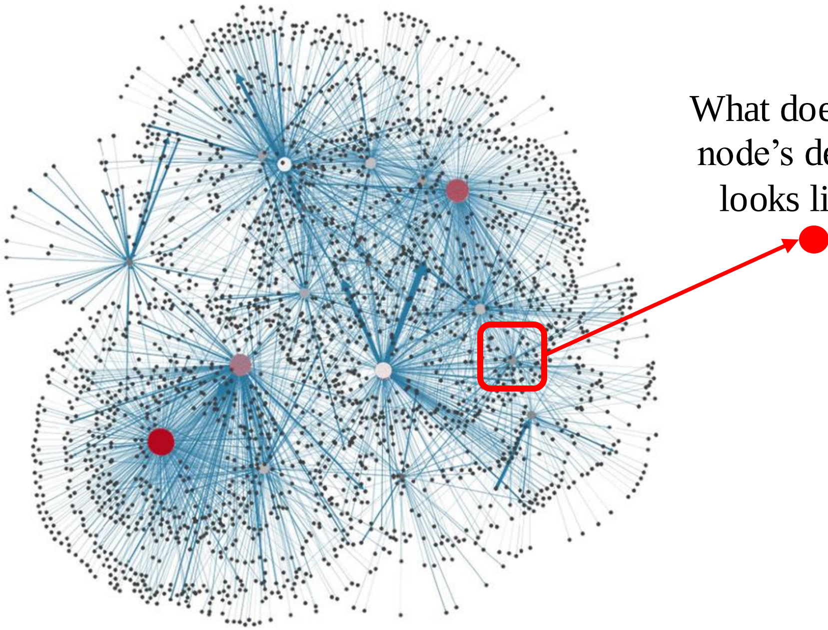
[Proof]

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





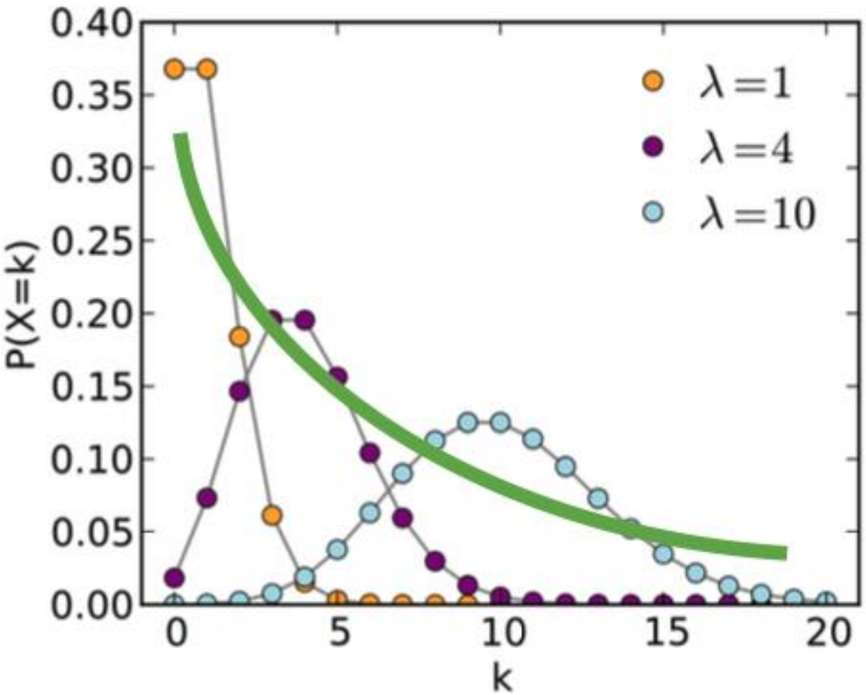
Statistical Graph Model – Random Graph Models – $G(n,p)$



What does this node's degree looks like?

Each node would have a degree

$$d_1, d_2, d_3, \dots, d_{|\mathcal{V}|}$$



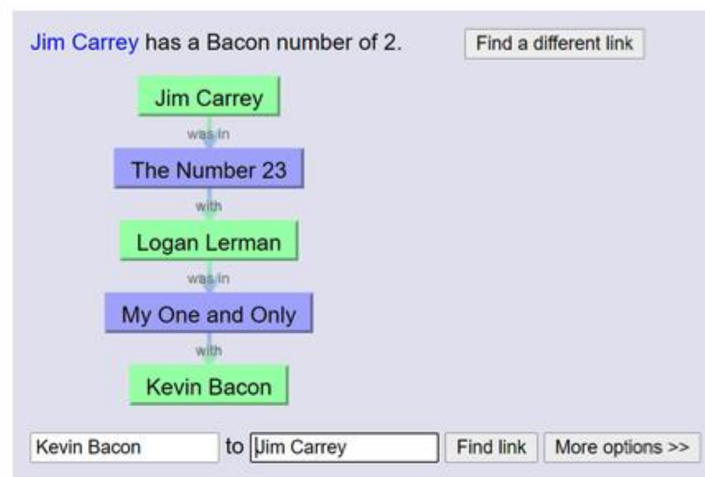
Empirically, power-law distribution!

Statistical Graph Model – Small World

Six Degrees of Kevin Bacon

Six Degrees of Kevin Bacon. **Six Degrees of Kevin Bacon** is a parlour game based on the "six degrees of separation" concept, which posits that any two people on Earth are six or fewer acquaintance links apart.

- Nodes: Hollywood actors
- Edges: Co-appearance in a movie
- Bacon# = # of steps from Kevin Bacon



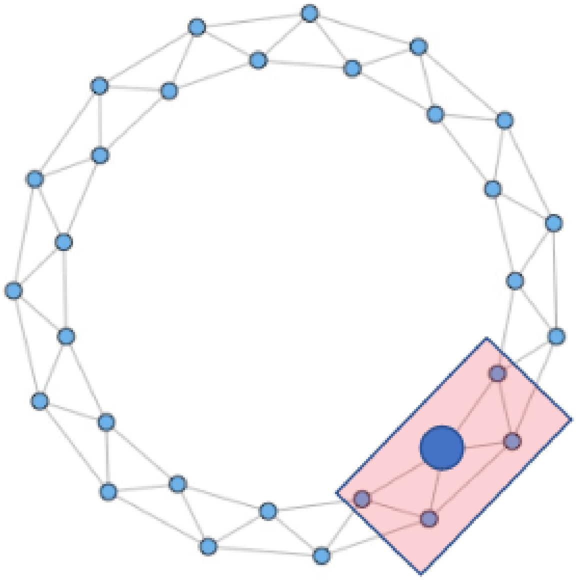
oracleofbacon.org

Small world experiment

- Stanley Milgram 1967
  - 300 randomly selected people
  - Asked them all to get a letter to a stockbroker in Boston by passing the letter through people they know
  - How many steps did it take?
    - 6.2 steps on average

Motivation:

- maintain high clustering and small diameter



Example:

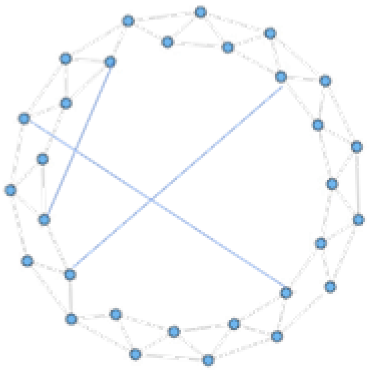
Clustering coefficient C=1/2

Graph diameter = 8

Statistical Graph Model – Small World

Idea:

- Randomly reconnect some links

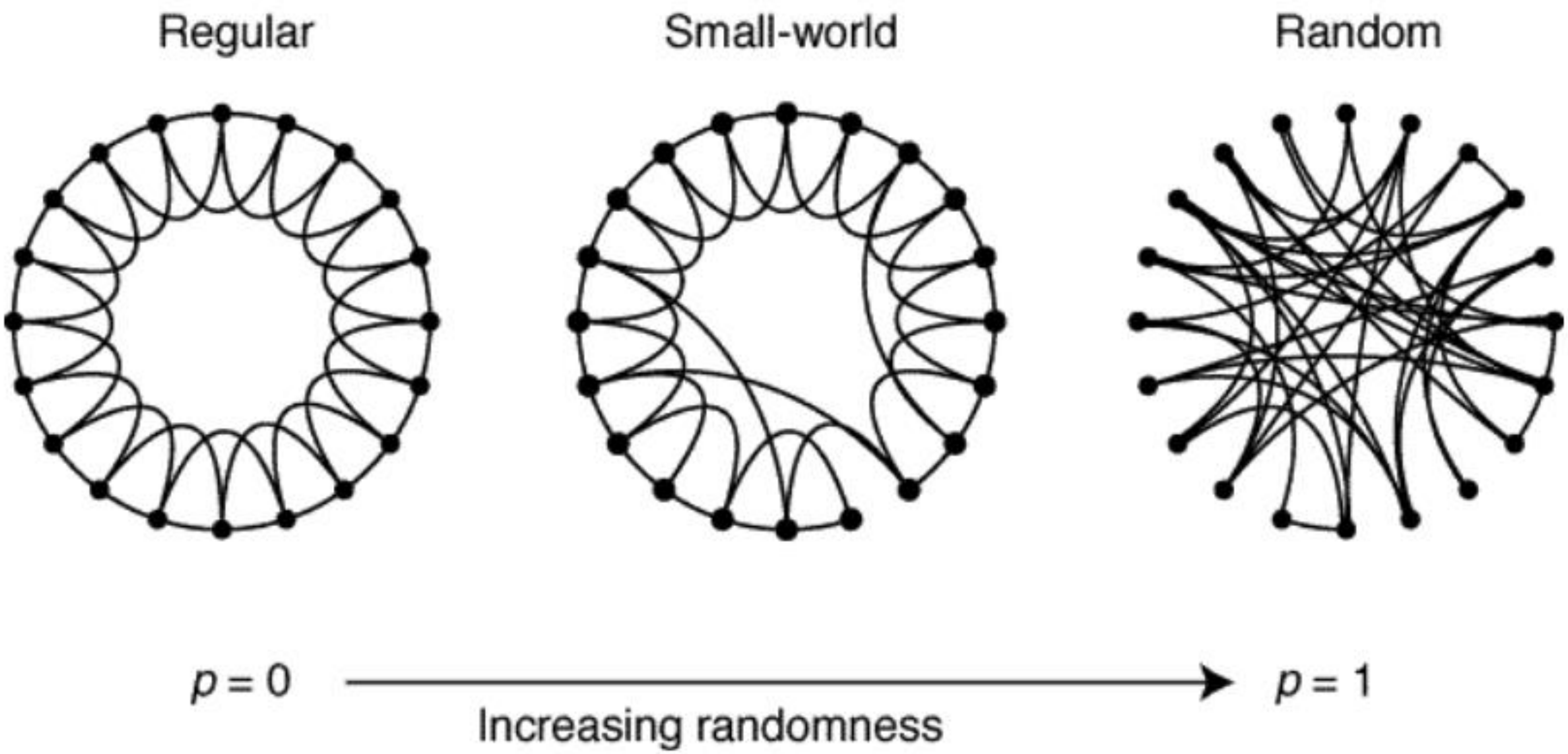


Watts and Strogatz 1998.

Single parameter model

- Go between regular lattice and random graph
- Start with regular lattice of n nodes, k edges per vertex where k<<n
- Randomly reconnect with other nodes with probability p
  - Creates pnk/2 “long distance” connections
- p=0 regular lattice and p=1 random graph

Statistical Graph Model – Small World



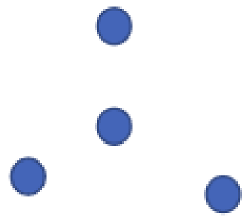
Statistical Graph Model – Dynamic Graph

- What does a real graph look like?
- What is normal/abnormal?
- Are real graphs random?

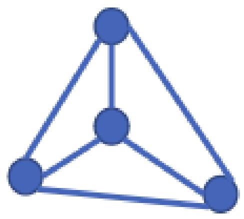
- Real Networks are growing!
  - Evolving with time
    - New nodes and edges
  - Citation/collaboration networks
  - Web
  - Social networks

Statistical Graph Model – Random Graph Models – $G(n, p)$

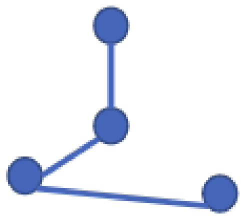
- Consider $G_{n,p}$ as a function of p
  - p=0, empty graph



  - p=1, complete (full) graph

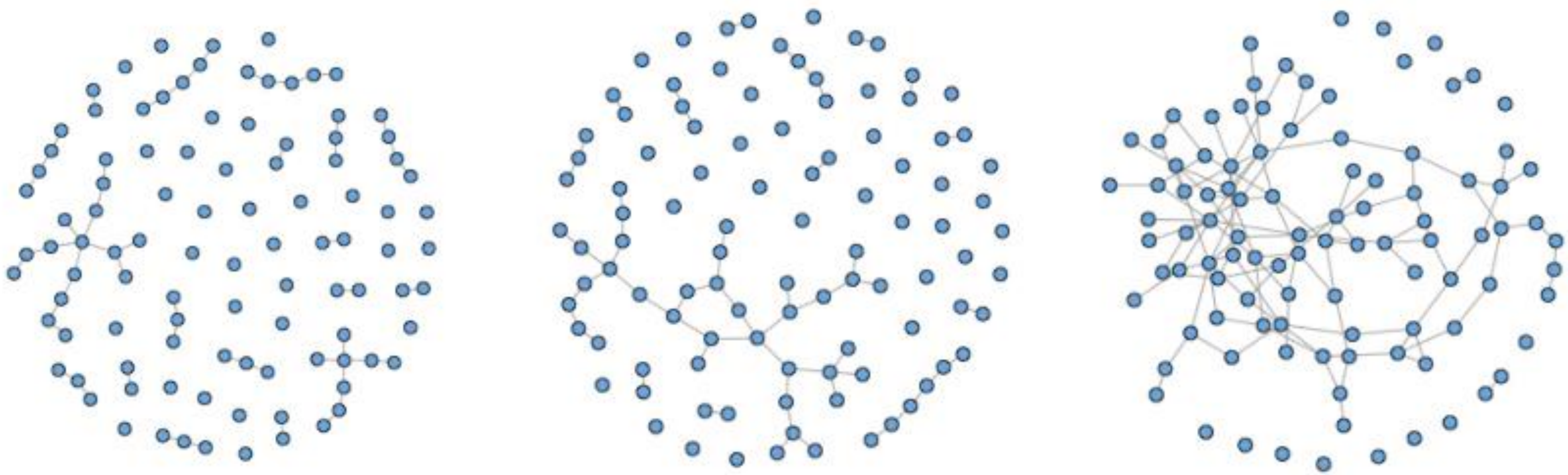


  - There exist a critical p_c, structural changes from $p < p_c$ to $p > p_c$



Gigantic connected component appears at $p > p_c$

Statistical Graph Model – Random Graph Models – $G(n, p)$



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

At the critical value nodes start becoming connected

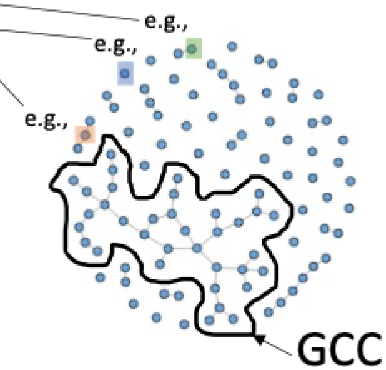
Note:

Just the number of nodes not in GCC divided by total number of nodes in the graph $u = \frac{n - n_{GCC}}{n}$

Phase Transition

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

$$u = \frac{n - n_{GCC}}{n} = P(k=0) + P(k=1) * u + P(k=2) * u^2 + P(k=3) * u^3 + \cdots$$



n_{GCC}=number of nodes in GCC

- If the node has 0 neighbors
- If the node has 1 neighbor, then that node should not be connected to GCC
- If the node has 2 neighbors, then those two nodes should not be connected to GCC

Phase Transition

Substitute

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

Note: $\lambda = pn = \langle k \rangle$

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

$$u = P(k=1) * u + P(k=2) * u^2 + P(k=3) * u^3 \ldots =$$

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

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

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Note: Pull $e^{-\lambda}$ out of the sum

Note: Combine u^k with λ^k (Taylor expansion →)

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

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

Let s – fraction of nodes belonging to GCC (size of GCC)

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

High density: What if $\lambda \to \infty$, then $s \to 1$

Low density: What if $\lambda \to 0$, then $s \to 0$

Note: $\lambda = pn = \langle k \rangle$

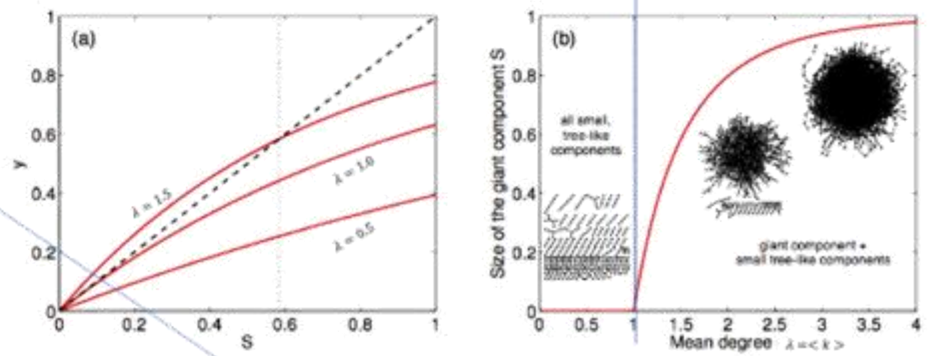
Phase Transition

When these intersect that is our solution

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

$y_1 = s$

$y_2 = 1 - e^{-\lambda s}$



Take the derivative

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

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

Only can happen then y_2 starts above the dotted line

critical value:

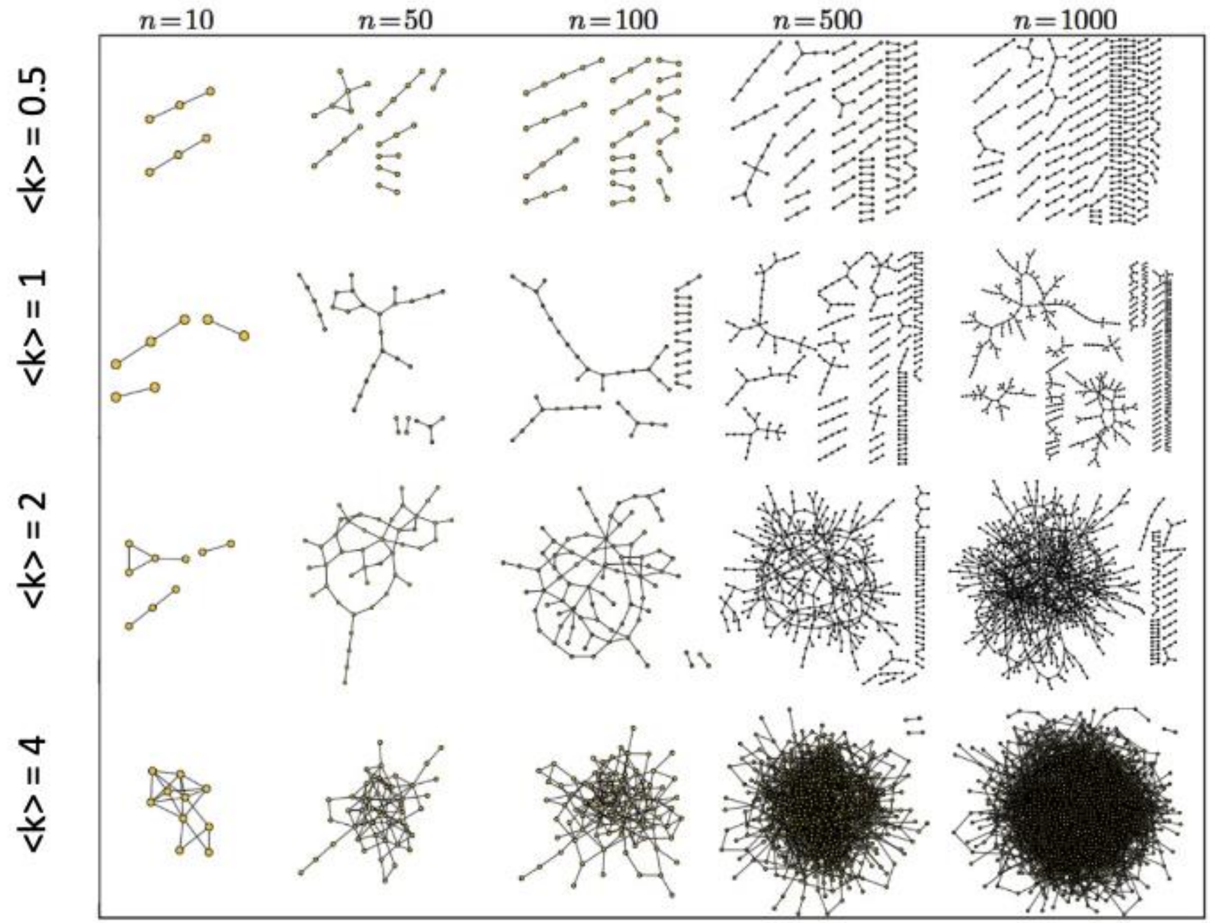
$$\lambda_c = 1$$

Interpretation: If having on average more than one neighbor then we have GCC

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

Statistical Graph Model – Random Graph Models – $G(n, p)$

Phase Transition



Any Question?