Machine Learning with Graphs (MLG)

# Network Generative Models

## Explain Why Graphs Exhibit the Three Properties

Cheng-Te Li (李政德)

Institute of Data Science

National Cheng Kung University

chengte@mail.ncku.edu.tw

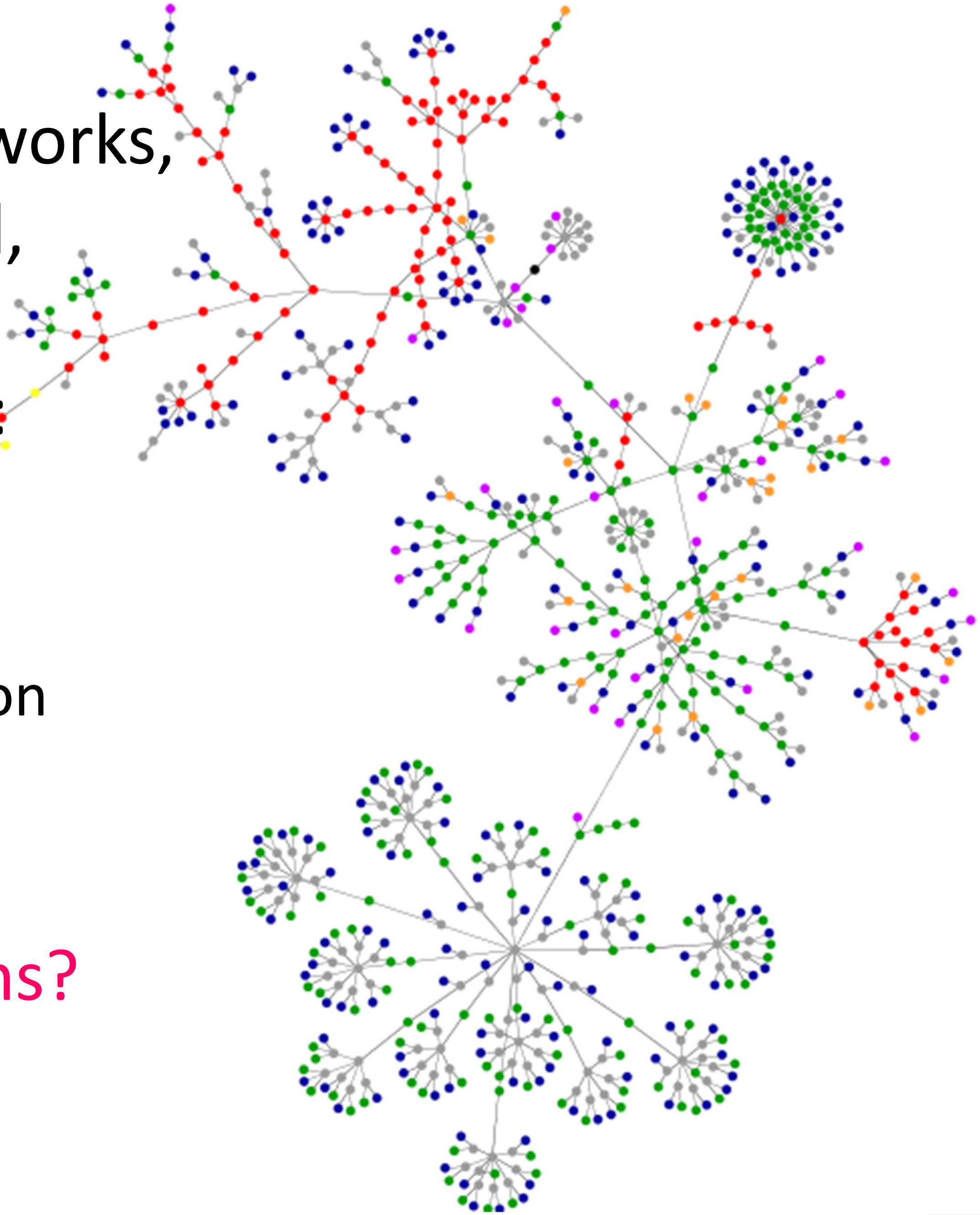
# 3 Essential Network Properties

- **Short** Average Path Length
  - Small-world Effect (小世界現象)
- **High** Clustering Coefficient
  - Friends of Friends are Friends (群聚現象)
- **Power-law** Degree Distribution
  - Long-tail Effect (長尾效應)

# Network Generative Model

- Social and Information Networks, WWW, computer, biological, social networks, etc. exhibit common properties:
  - High Clustering Coefficient
  - Low Average Path Length
  - Power-law Degree Distribution

How can we produce synthetic but realistic graphs?

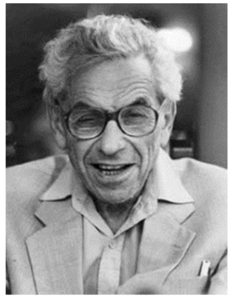


# Network Generative Model

- Why do we need synthetic graphs?

1) Simulation
2) Sampling & Extrapolation
3) Summarization & Compression
4) Motivation to understand properties' generating processes
5) Learn how human society forms over time

# 3 Network Generative Models



1960

**Random Graph**

Erdos-Renyi Model

數學家

On the Evolution of Random Graphs

citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.153.5943 ▾ 翻譯這個網頁

由 P Erdős 著作 - 1960 - 被引用 5441 次 - 相關文章 **6888**

On the Evolution of Random Graphs (1960) ... Correct Errors · Monitor Changes. by P. Erdős , A Rényi ... author = {P. Erdős and A Rényi}, title = {On the Evolution ...

Cite

1999

**Scale-free Network**

Barabasi-Albert Model

物理學家

Science

Emergence of scaling in random networks

AL Barabási, R Albert - science, 1999 - science.sciencemag.org

Abstract Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be

被引用 26318 次 相關文章 全部共 72 個版本 引用 儲存

**35860**

Cite

1998

**Small-world Network**

Watts-Strogatz Model

社會學家

Nature

Collective dynamics of /'small-world/' networks : Article : Nature

www.nature.com › Journal home › Archive › Letters to Nature - 翻譯這個網頁

由 DJ Watts 著作 - 1998 - 被引用 31894 次 - 相關文章 **41906**

Nature 393, 440-442 (4 June 1998) | doi :10.1038/30918; Received 27 ... In particular, infectious diseases spread more easily in small-world networks than in ...

# Erdos-Renyi Model

## Random Graph

- (**5446** cites) P. Erdos and A. Renyi. “On the evolution of random graphs.” The Mathematical Institute of the Hungarian Academy of Sciences, 1960.
- (4955 cites) P. Erdos and A. Renyi. “On Random Graphs I.” Publicationes Mathematicae, 1959.
- (3234 cites) M. E. J Newman, S. H. Strogatz, and D. J. Watts. “Random graphs with arbitrary degree distributions and their applications.” Physical Review E., 2001.
- (897 cites) E. N. Gilbert. “Random Graphs.” Annals of Mathematical Statistics, 1959.

[Erdos-Renyi'60]

# Erdos-Renyi Random Graphs

(ER Model)

- Consider a graph with $n$ nodes
- Let $m$ denote the total number of possible edges

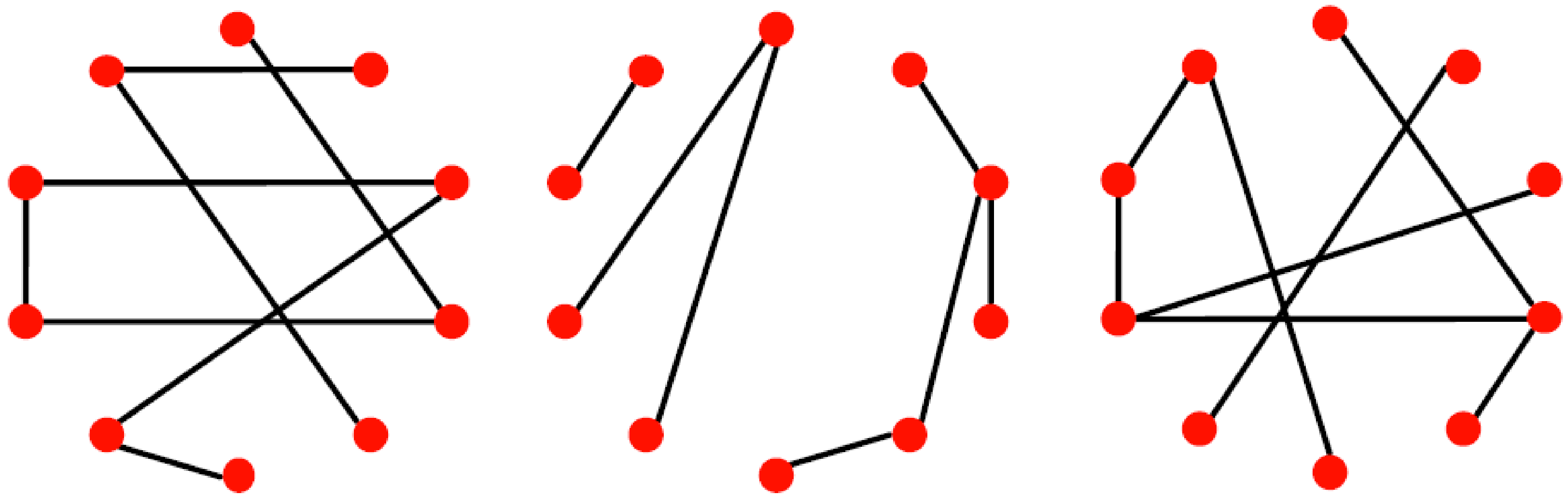
$$m = C_2^n = \frac{n(n-1)}{2}$$

- $\boldsymbol{G_{np}}$: undirected graph on $n$ nodes and each edge $(u, v)$ appears I.I.D. with probability $p$

What kinds of networks does such model produce?

# Random Graph Model

- $n$ and $p$ do not uniquely determine the graph!
  - The graph is a result of a random process
- We can have many different realizations given the same $n$ and $p$



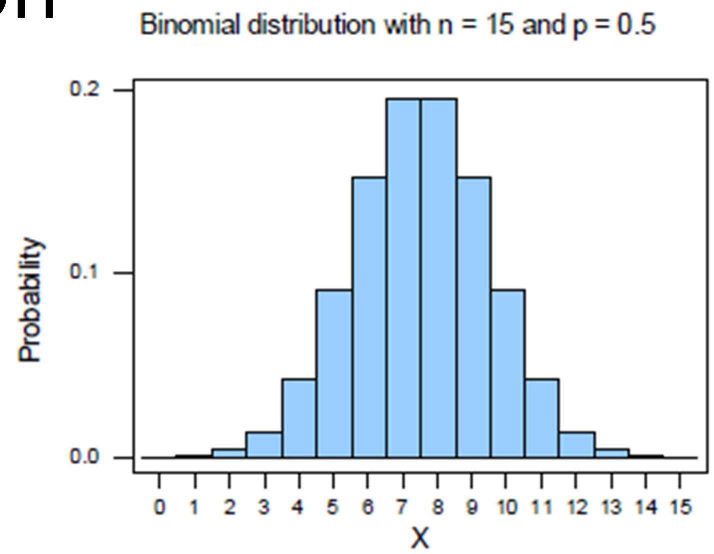
n = 10
p= 1/6

# Random Graph Model: Edges

- How likely is a graph on $m$ edges?
- $P(m)$: the probability that a given $G_{np}$ generates a graph on exactly $m$ edges

$$P(m) = C_m^{C_2^n} p^m (1-p)^{C_2^n - m}$$

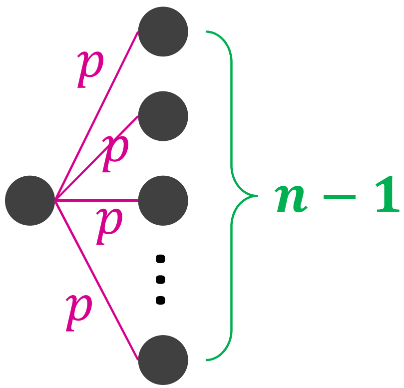
- $P(m)$ is the Binomial distribution
  - Number of successes in a sequence of $C_2^n$ independent yes/no experiments



# Random Graph Model: Node Degree

- What is the expected degree of a node?
  - Let $X_v$ be a random variable measuring the degree of node $v$
- Expected Degree
  - Decompose $X_v$ into $X_v = X_{v,1} + X_{v,2} + \cdots + X_{v,n-1}$
    - $X_{v,u}$ is a $\{0,1\}$-random variable which tells if edge $(v,u)$ exists or not

$$E[X_v] = \sum_{u=1}^{n-1} E[X_{vu}] = (n-1)p$$



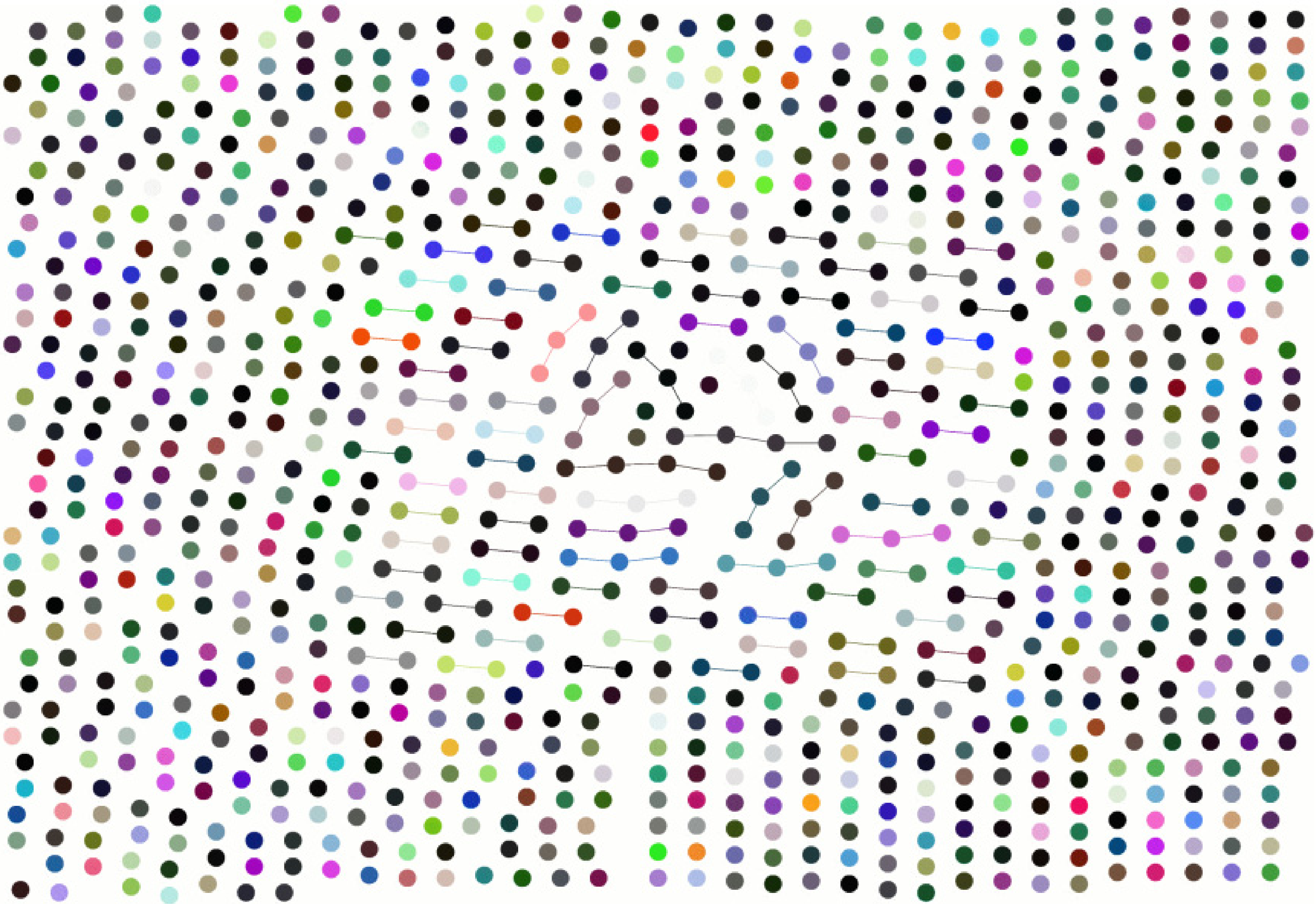

Prob. of node $u$ linking to node $v$ is $p$
$u$ can link (flips a coin) to all other $(n-1)$ nodes
Thus, the expected degree of node $u$ is: $p(n-1)$

# Evolution of Random Graphs

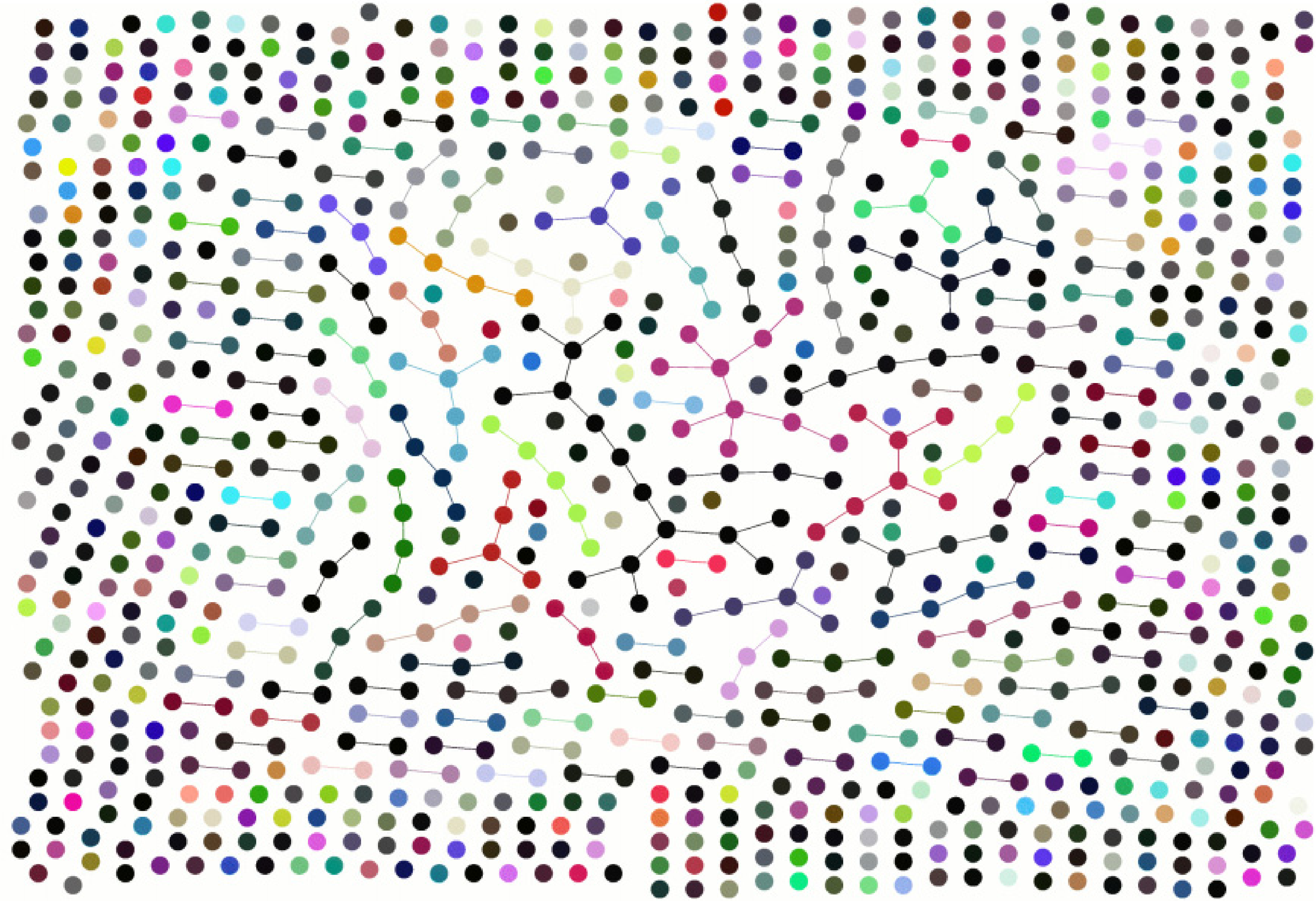
- In random graphs, as we increase $p$, a large fraction of nodes start getting connected
    - i.e., we gradually have a path between any pair of nodes
  - This large fraction forms a connected component:
  - $C_{max}(p)$: the number of nodes of the largest connected component, a.k.a. **Giant Component**, in $G_{np}$
- In random graphs:
  - $p = 0$ : the size of the giant component is $0$
  - $p = 1$ : the size of the giant component is $n$
- Apparently, $C_{max}(p)$ depends on $p$
  - What happens when $p$ grows from 0 to 1 ?

For small $p$, few edges on the graph.

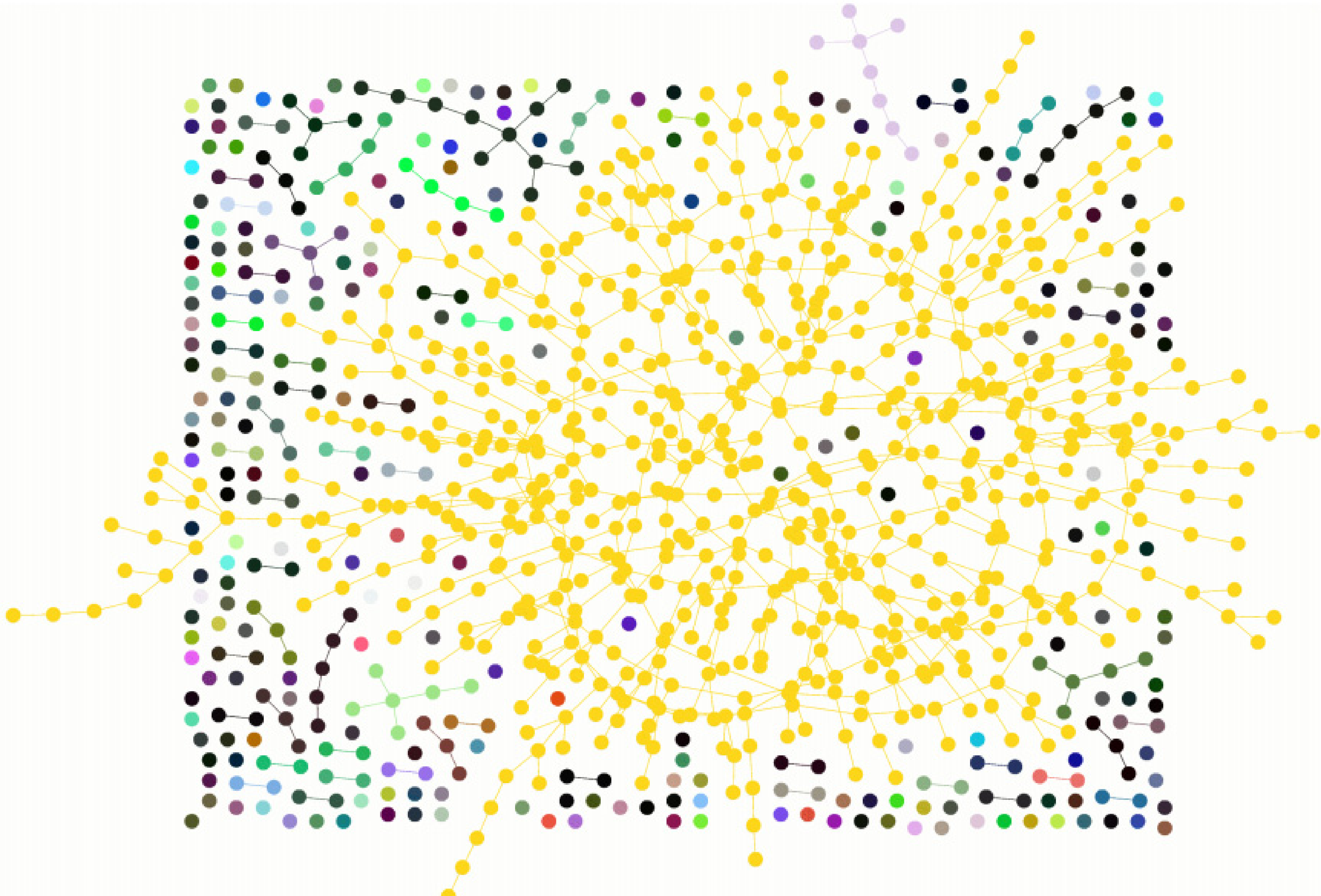
Almost all vertices disconnected. The component sizes are small.



# Keep increasing $p$ …

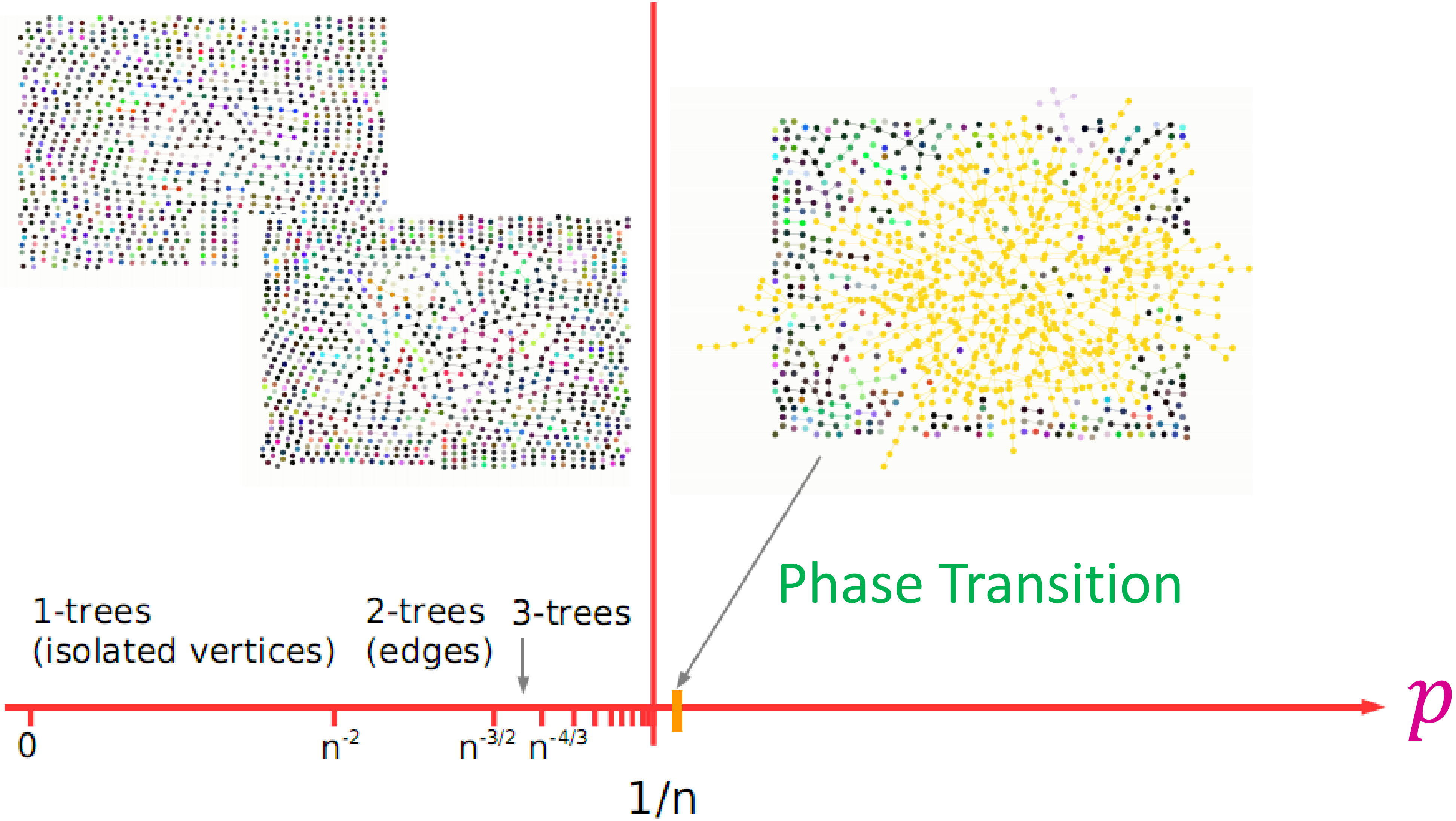


# Keep increasing $p$ …



# Emergence of the Giant Component

- After $p = \frac{1}{n}$
  - Suddenly the largest component starts emerging



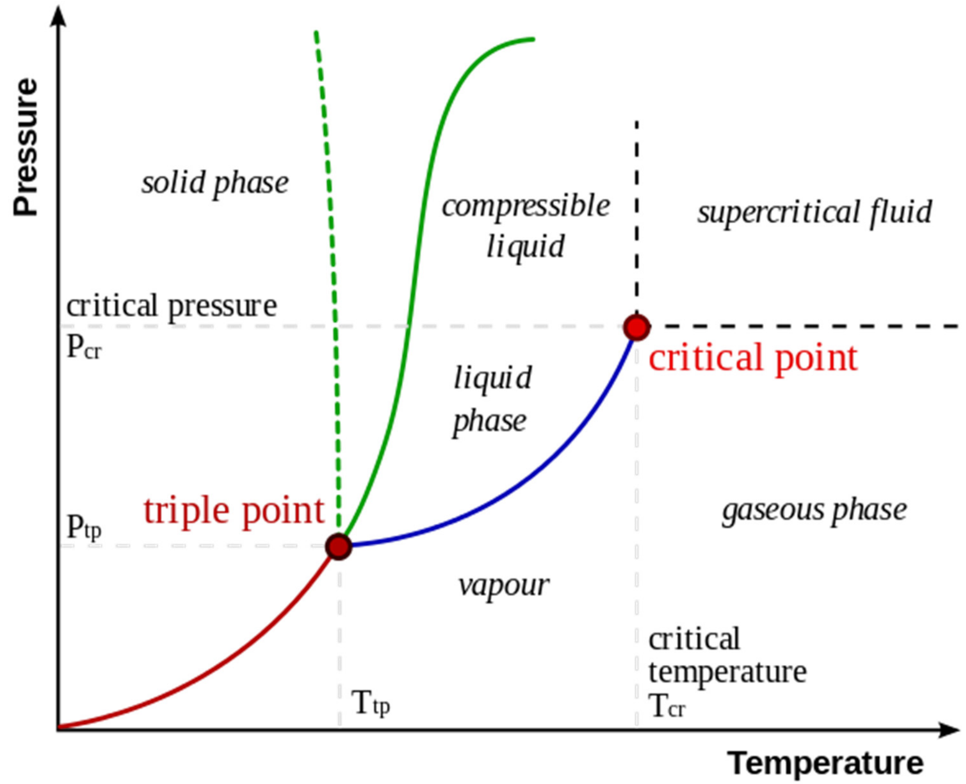
# Phase Transition (相變)

- An abrupt sudden change in one or more physical properties, resulting from a small change in a external control parameter
- It occurs in many physical systems
  - Liquid/Gas
  - Superconductivity

材料可以在在特定溫度以下，
呈現電阻為零的導體

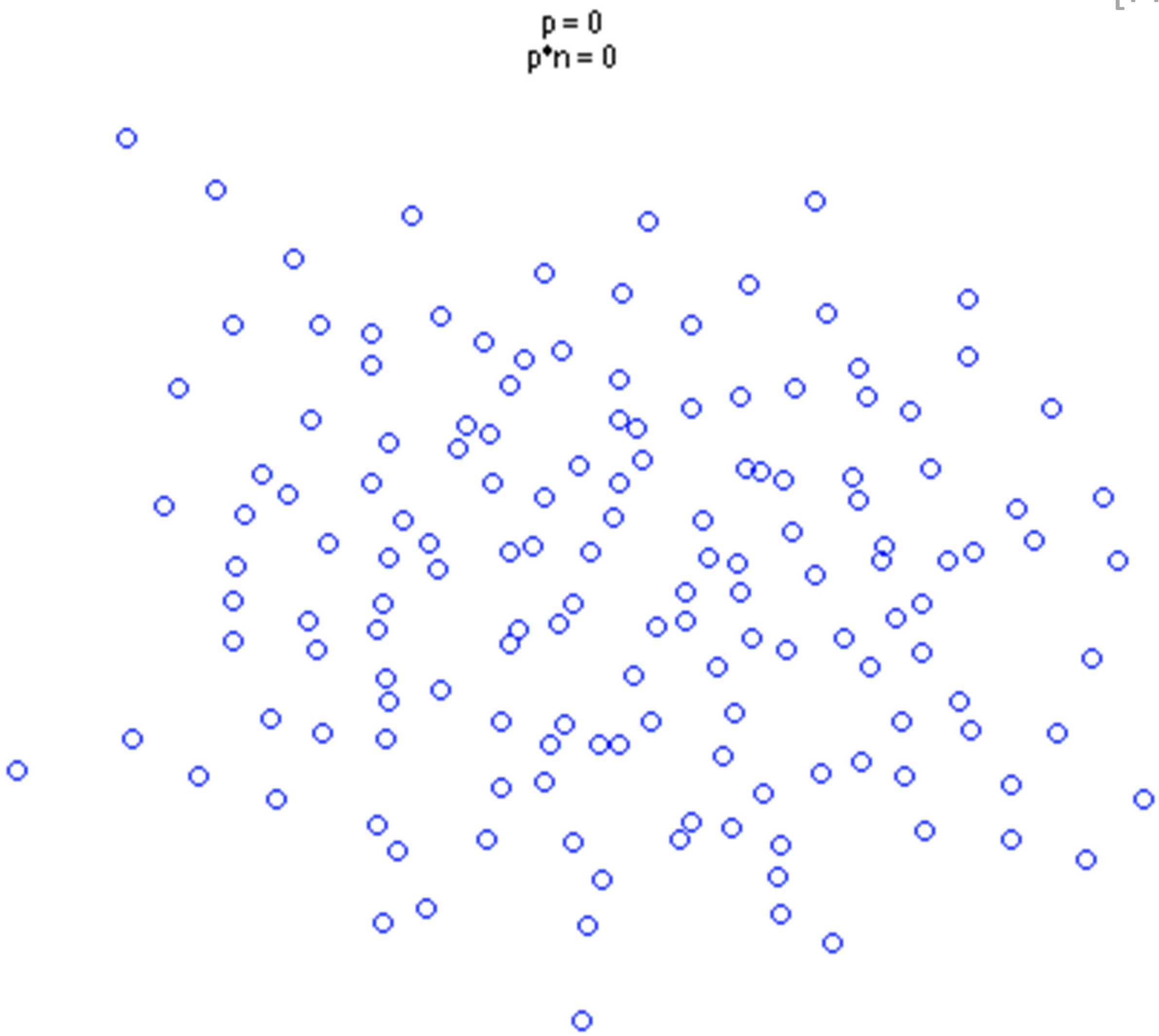
  - Magnetization

加熱一塊磁鐵，磁鐵的鐵磁性忽然消失



# [DEMO] Giant Component ($n = 150$)

[From David Gleich]




https://www.cs.purdue.edu/homes/dgleich/demos/erdos_renyi/er-150.gif

# Phase Transition in Connectivity

- At phase transition, $p = 1/n$
  - $p < 1/n$: only small disconnected components
  - $p > 1/n$: one large component, which quickly gain more mass
- Average node degree $z = E[X_v]$

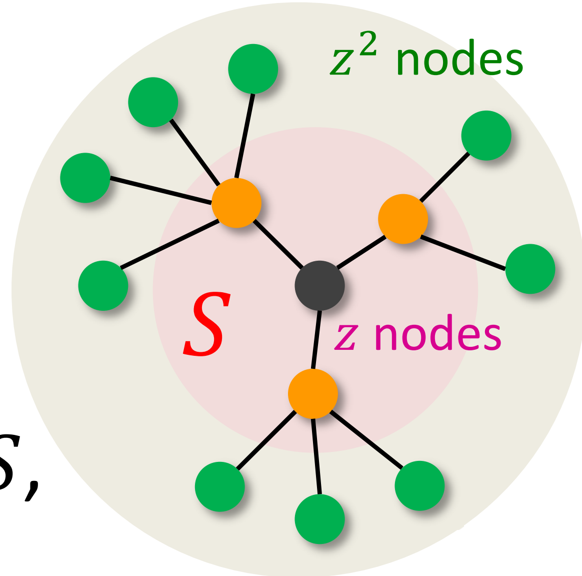
$$z = \frac{2 \times m}{n} = \frac{2 \times p \times C_2^n}{n} = \frac{pn(n-1)}{n} = (n-1)p \approx np$$

- $z = np = n\frac{1}{n} = 1$
  - Phase transition at average node degree = 1

# The Intuition behind $z = 1$

Consider a random graph with expected node degree $z$

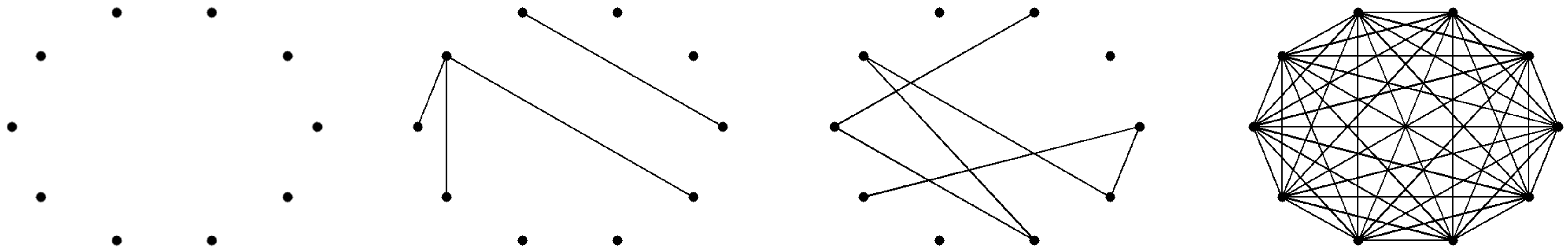
- For any connected set of nodes $S$
  - Let $S' = V - S$ denote the complement set
  - Assume $|S| << |S'|$




- If moving 1 step away from any node $v$ in $S$, we visit approximately $z$ nodes
  - If moving 1 step away from $S$, we visit approximately $|S|z$ nodes
- In the limit, if we want this connected component to be the largest one, then after traveling $l$ steps, its size must grow and we must have

$$z^l \geq 1 \text{ or equivalently } z \geq 1$$

# The Giant Component



| Probability ($p$) | 0.0 | 0.055 | 0.11 | 1.0 |
|---|---|---|---|---|
| Average Node Degree ($z$) | 0.0 | 0.8 | $\approx 1$ | $n - 1 = 9$ |
| Diameter | 0 | 2 | 6 | 1 |
| Giant Component Size | 0 | 4 | 7 | 10 |
| Average Path Length | 0.0 | 1.5 | 2.66 | 1.0 |

# Evolution of Random Graphs

If $z < 1$:

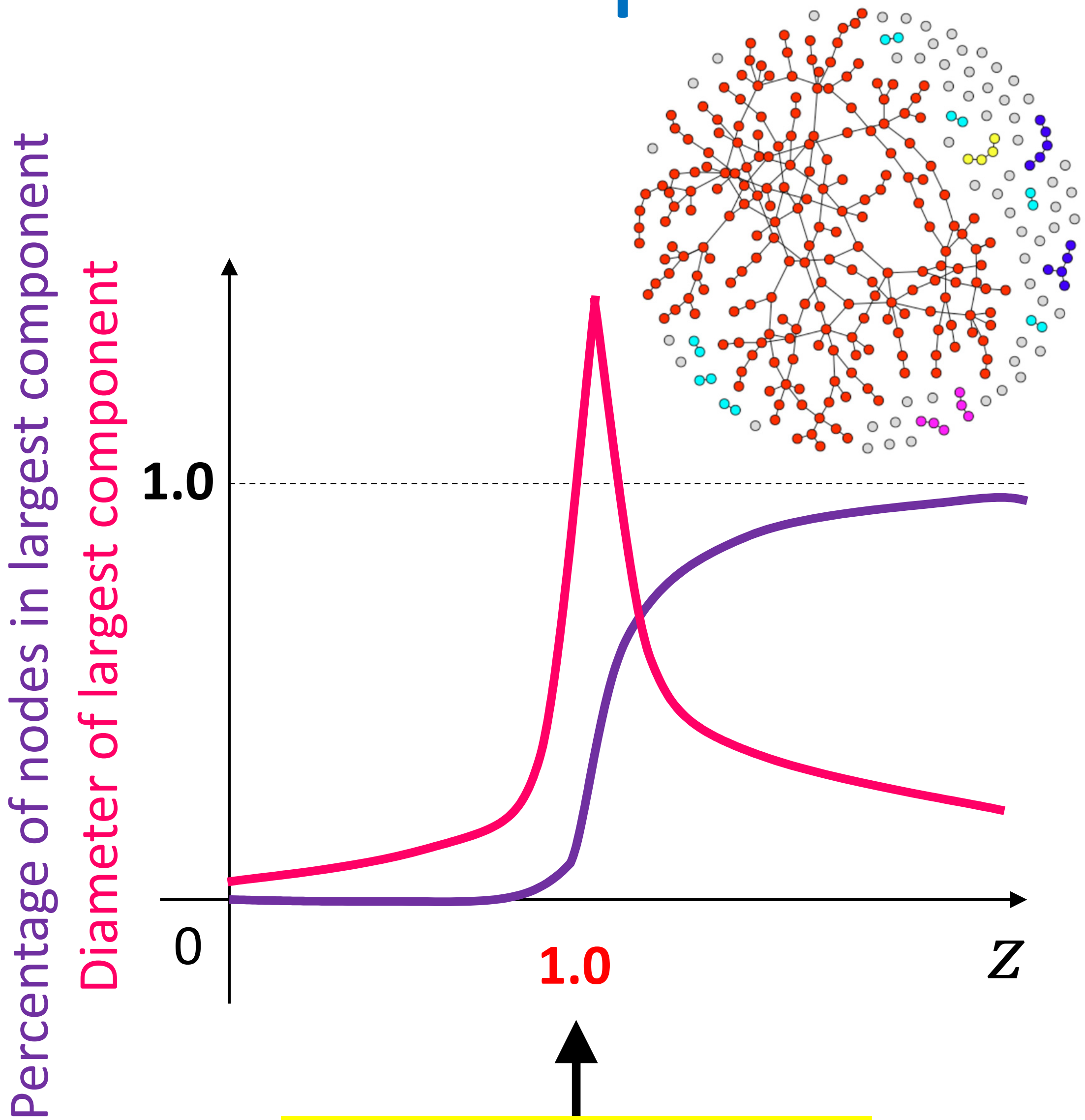
- **small**, isolated clusters
- **small** diameters
- **short** path lengths

At $z = 1$:

- a ***giant component*** appears
- diameter **peaks**
- path lengths are **long**

For $z > 1$:

- almost **all** nodes **connected**
- diameter **shrinks**
- path lengths **shorten**

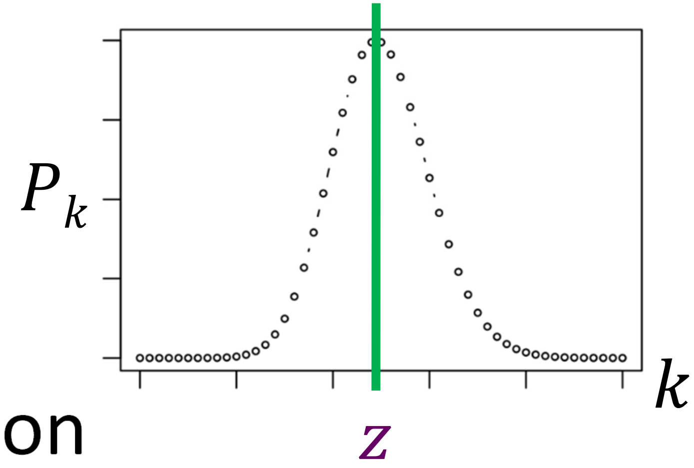



phase transition

# $G_{np}$: Poisson Degree Distribution

- Consider $G_{np}$ for a fixed $\boldsymbol{p}$ and $\boldsymbol{n}$ ($n$ is large)
- The absence or presence of an edge is independent for all edges
  - *Prob*(node $i$ connects to all other $n-1$ nodes) = $p^{n-1}$
  - *Prob*(node $i$ is isolated) = $(1-p)^{n-1}$
  - *Prob*(node $i$ has degree $\boldsymbol{k}$) follows a Binomial distribution

$$p_k = C_k^{n-1} p^k (1-p)^{n-1-k}$$




When $n$ is large ($n \to \infty$),
$p_k$ becomes a Poisson degree distribution

# $G_{np}$: Clustering Coefficient = $p$

- Recall: $C_i = \frac{2e_i}{k_i(k_i-1)}$

$e_i$ is the number of edges between node $i$'s neighbors

- Edges in $G_{np}$ appear I.I.D. with probability $p$

- So: $e_i = p\,\frac{k_i(k_i-1)}{2}$

Number of distinct pairs of neighbors of node $i$ of degree $k_i$

Each pair is connected with prob. $p$

- Then: $C = \frac{p k_i(k_i-1)}{k_i(k_i-1)} = p \quad \left(= \frac{z}{n-1} \approx \frac{z}{n}\right)$

Clustering coefficient of a random graph is small

# Clustering Coefficient for Different Networks

$$C = p = \frac{z}{n-1}$$

| network | $n$ | $z$ | clustering coefficient $C$ measured | clustering coefficient $C$ random graph |
|---|---|---|---|---|
| Internet (autonomous systems)[a] | 6 374 | 3.8 | 0.24 | 0.00060 |
| World-Wide Web (sites)[b] | 153 127 | 35.2 | 0.11 | 0.00023 |
| power grid[c] | 4 941 | 2.7 | 0.080 | 0.00054 |
| biology collaborations[d] | 1 520 251 | 15.5 | 0.081 | 0.000010 |
| mathematics collaborations[e] | 253 339 | 3.9 | 0.15 | 0.000015 |
| film actor collaborations[f] | 449 913 | 113.4 | 0.20 | 0.00025 |
| company directors[f] | 7 673 | 14.4 | 0.59 | 0.0019 |
| word co-occurrence[g] | 460 902 | 70.1 | 0.44 | 0.00015 |
| neural network[c] | 282 | 14.0 | 0.28 | 0.049 |
| metabolic network[h] | 315 | 28.3 | 0.59 | 0.090 |
| food web[i] | 134 | 8.7 | 0.22 | 0.065 |

# Real-world Networks vs. $G_{np}$

1) Compute the average degree $z$ in the real-world graph
2) Compute $p$ using $p = z/(n-1)$
3) Generate the random graph using $p$

| | Original Network | | | | Simulated Random Graph | |
|---|---|---|---|---|---|---|
| **Network** | *Size* | *Average Degree* | *Average Path Length* | $C$ | *Average Path Length* | $C$ |
| Film Actors | 225,226 | 61 | 3.65 | 0.79 | 2.99 | 0.00027 |
| Medline Coauthorship | 1,520,251 | 18.1 | 4.6 | 0.56 | 4.91 | $1.8 \times 10^{-4}$ |
| E.Coli | 282 | 7.35 | 2.9 | 0.32 | 3.04 | 0.026 |
| C.Elegans | 282 | 14 | 2.65 | 0.28 | 2.25 | 0.05 |

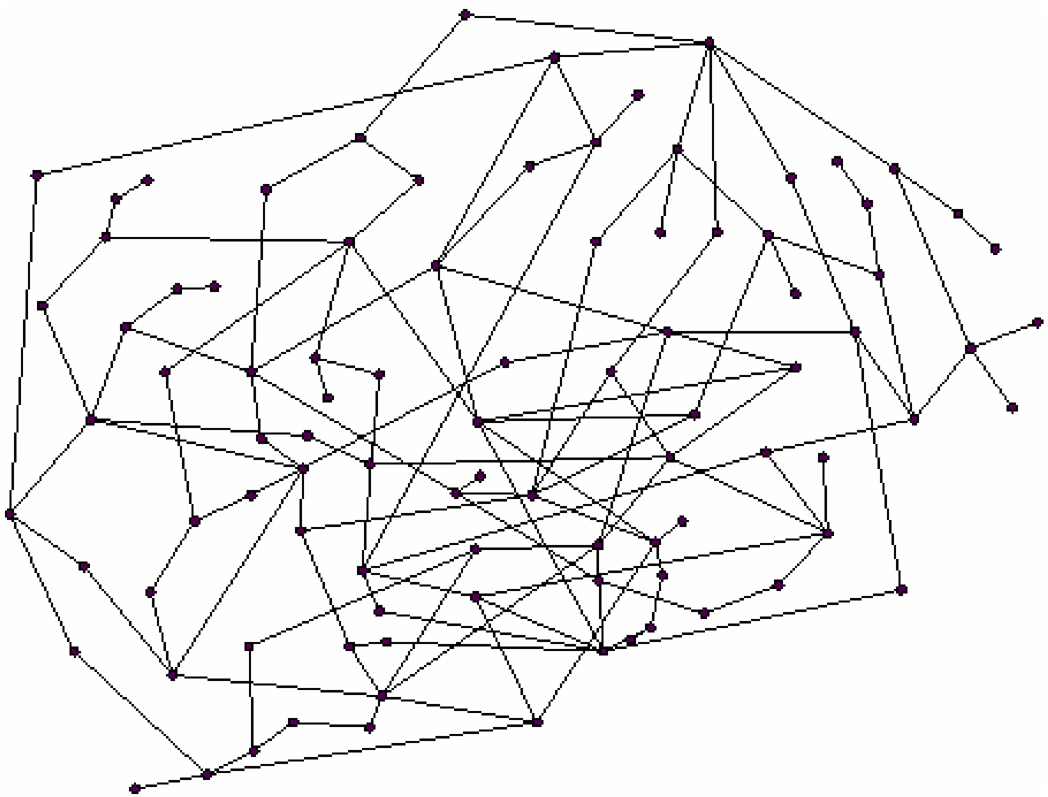
# Real-world Networks vs. $G_{np}$

- Are real networks like random graphs? **Real**
  - Average Path Length: Low $l \approx \frac{\log n}{\log z}$ — Low ☺
  - Clustering Coefficient: Low $C = p$ — High ☹
  - Degree Distribution: Binomial/Poisson — Power-law ☹
- So, are real networks random?
  - The answer is obvious: **NO!**

# Barabasi-Albert Model
## Scale-free Network

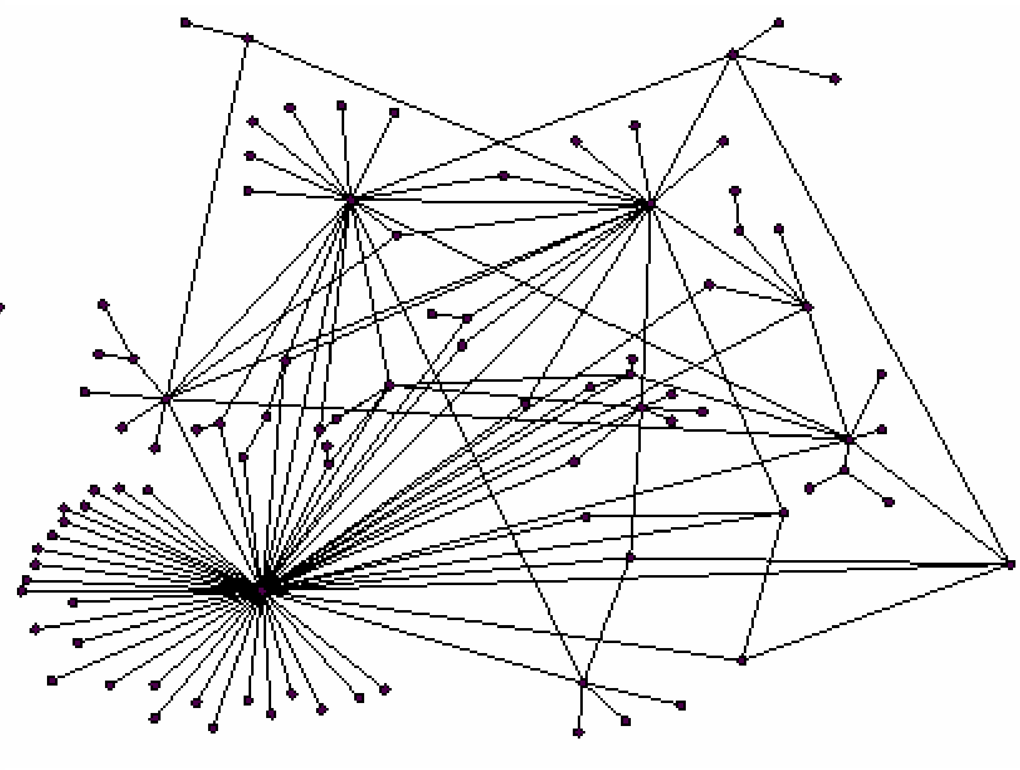
- (**26318** cites) A. L. Barabasi and R. Albert. “Emergence of scaling in random networks.” Science, 1999.
- (**7259** cites) R. Albert, H. Jeong, A. L. Barabási. “Error and attack tolerance of complex networks.” Nature, 2000.
- (**6815** cites) S.H. Strogatz. “Exploring complex networks.” Nature, 2001.
- (**2143** cites) A. L. Barabasi. “Scale-Free Networks.” Scientific American, 2003.

# Random Graph vs. Scale-free Network



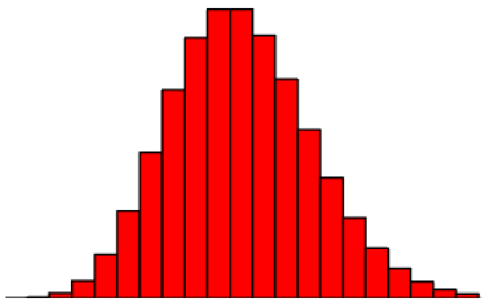
Poisson network

(Erdos-Renyi random graph)

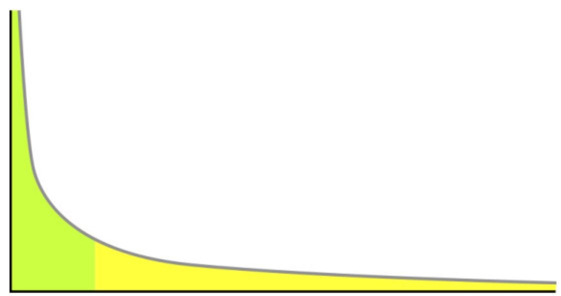


Scale-free network

(power-law network)



Degree distribution is Poisson



Degree distribution is Power-law

# Should the number of nodes be fixed?

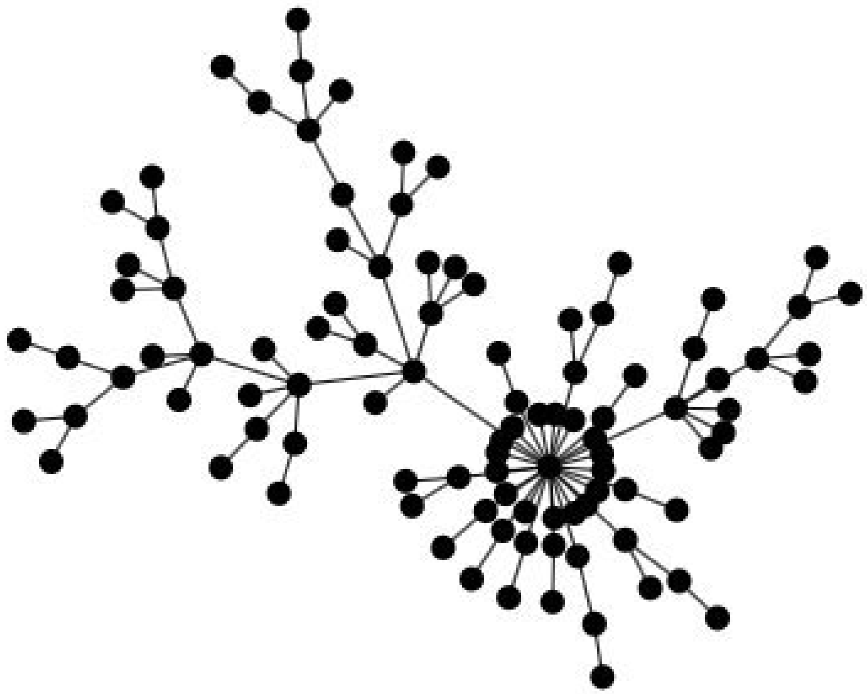
- ER model, random graphs, assumes the fixed number of nodes, and the probability that two nodes are connected is independent of the nodes' degree
- However, most real-world networks grow by continuously adding new nodes
  - E.g., WWW, citations
- Can we model such network growing?

# Rich Gets Richer

- Sociologist *Herbert Simon*:
  - Power-laws arise from "Rich get richer" (cumulative advantage)
- New nodes are more likely to link to nodes that already have high degree
- Examples
  - New citations to a paper are proportional to the number it already has

**Herding:** If a lot of people cite a paper, then it must be good, and therefore I should cite it too

  - Matthew effect (馬太效應)



# Matthew Effect

## 《新約聖經·馬太福音》第13章第12節

“ 凡有的，還要加給他，叫他有餘；凡沒有的，連他所有的也要奪去。 ”

## 老子《道德經》第77章

「天之道，損有餘而補不足。人之道則不然，損不足以奉有餘。孰能有餘以奉天下，唯有道者。」

“Eminent scientists will often get more credit than a comparatively unknown researcher, even if their work is similar; it also means that credit will usually be given to researchers who are already famous.” (Wikipedia)

「相對於那些不知名的研究者，聲名顯赫的科學家通常得到更多的聲望，即使他們的成就是相似的，同樣地，在同一個項目上，聲譽通常給予那些已經出名的研究者，例如，一個獎項幾乎總是授予最資深的研究者，即使所有工作都是一個研究生完成的。」(維基百科)

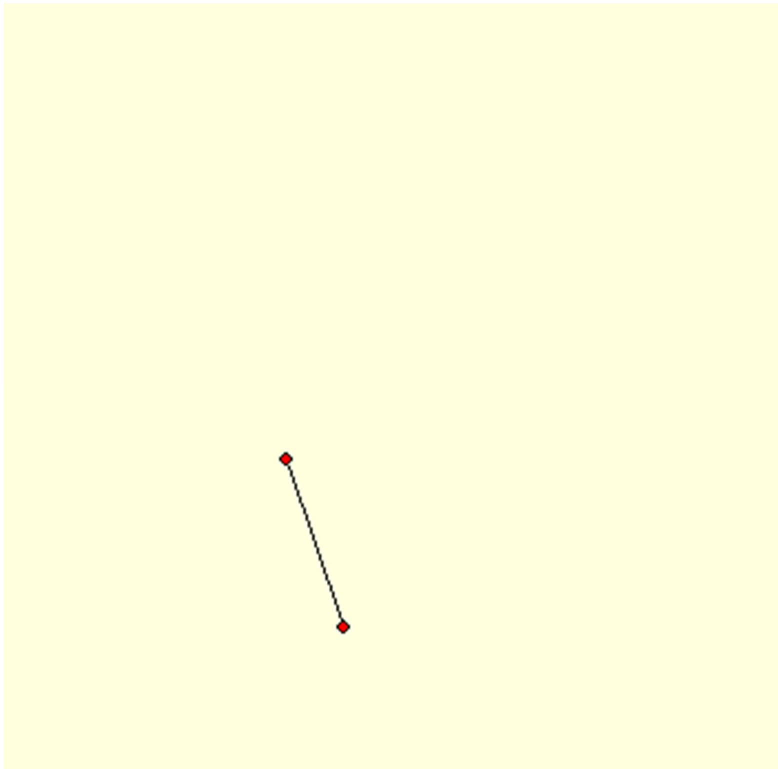
# Barabasi-Albert (BA) Model

- Idea: Preferential Attachment
  - When a new user joins the network, the probability of connecting to existing nodes is proportional to existing nodes' degrees
  - New nodes tend to connect themselves to a given node that has higher degree

https://upload.wikimedia.org/wikipedia/commons/thumb/4/48/Barabasi_Albert_model.gif/300px-Barabasi_Albert_model.gif

# Barabasi-Albert (BA) Model

1) Start with a small number ($\boldsymbol{m_0}$) of fully connected nodes
2) Each time, add **one new** node with $\boldsymbol{m}$ ($\leq \boldsymbol{m_0}$) edges
   - Connect the new node $v$ to $\boldsymbol{m}$ different nodes already present in the system
   - Based on the degree $k_i$ of node $v_i$ i.e., Connect $v$ to a random node $v_i$ with probability:
     $$P(v_i) = \frac{k_i}{\sum_j k_j}$$
3) After $\boldsymbol{t}$ time steps
   - *There are* $\boldsymbol{N} = \boldsymbol{t} + \boldsymbol{m_0}$ nodes,
     $\boldsymbol{mt} + C_2^{m_0}$ edges
     total degree = $2mt + 2C_2^{m_0}$



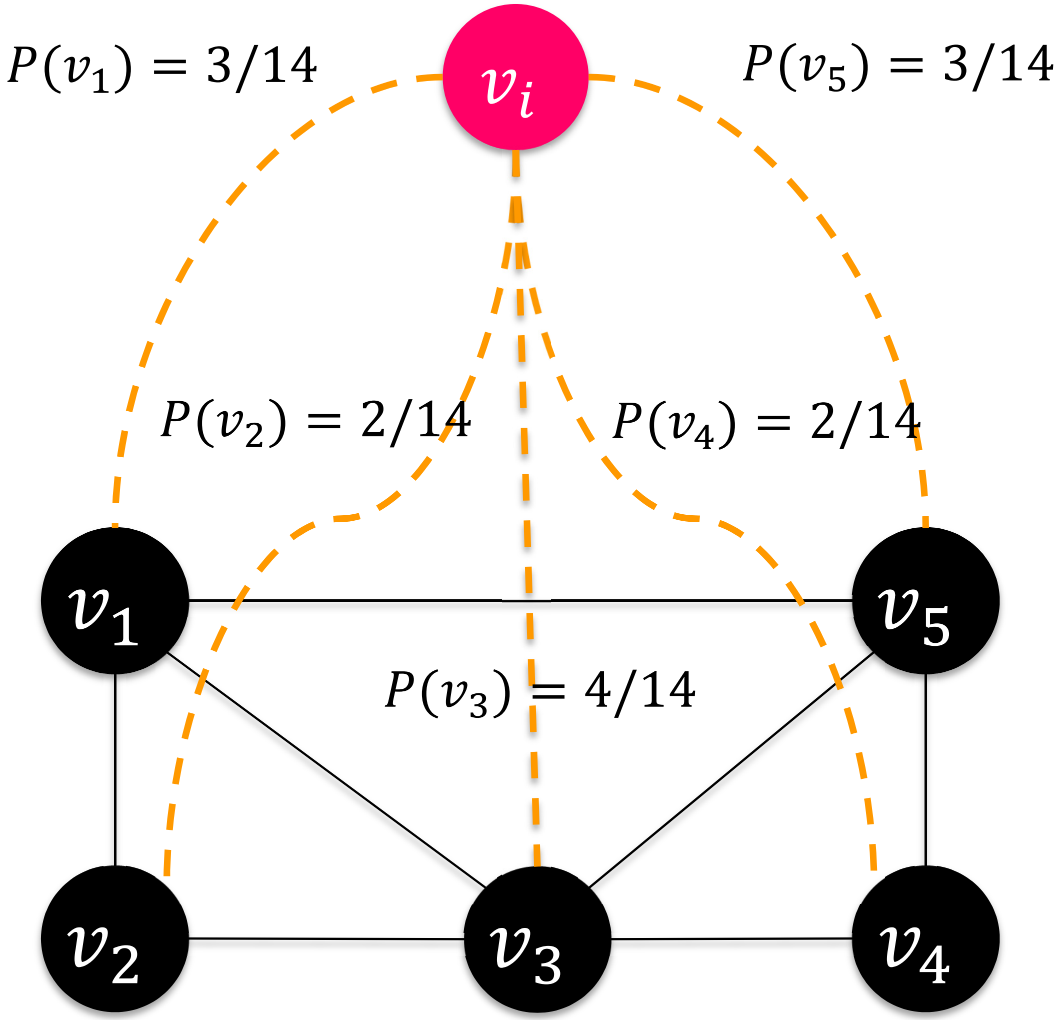
http://www.can.fudan.edu.cn/lix/ComplexNet.gif

# Preferential Attachment: Example

- New node $v_i$ arrives

$$P(v_i) = \frac{k_i}{\sum_j k_j}$$

- $P(v_1) = 3/14$
- $P(v_2) = 2/14$
- $P(v_3) = 4/14$
- $P(v_4) = 2/14$
- $P(v_5) = 3/14$

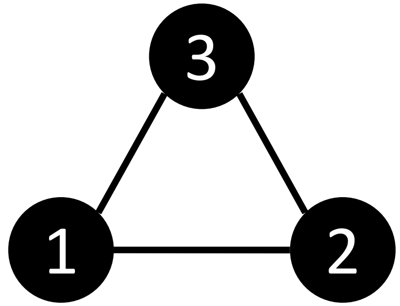


# Implementation Details

- Use an array to keep track of the end nodes of every edge
- Select one from this array at random
- Prob. of selecting is proportional to #(appearing times), corresponding to its degree

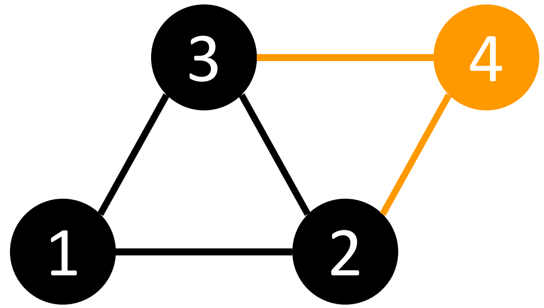
# Generating Barabasi-Albert Networks

- Start with a set of $m_0 = 3$ fully connected nodes



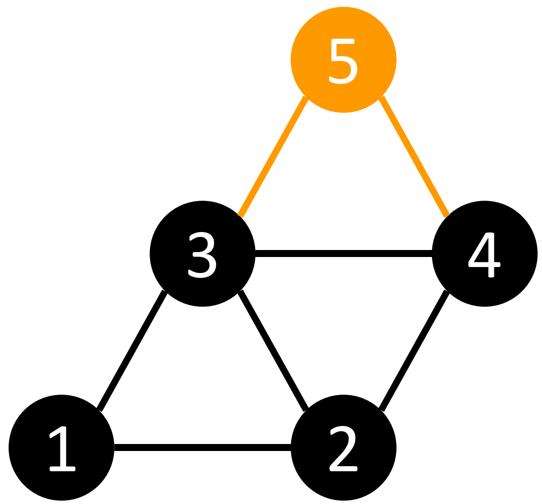

1 1 2 2 3 3

- Add a new node **4**, it has $m = 2$ edges
  - Prob(selecting any node)=1/3
  - Suppose they are node2 and node3




1 1 2 2 **2** 3 3 **3 4 4**

- Add a new node **5**, it has $m = 2$ edges
  - Prob(selecting node**1**)=2/10=1/5
  - Prob(selecting node**2**)=3/10
  - Prob(selecting node**3**)=3/10
  - Prob(selecting node**4**)=2/10=1/5




1 1 2 2 2 3 3 3 **3** 4 4 **4 5 5**

- Add a new node **6** ...
  - 2/14, 3/14, 4/14, 3/14, 2/14

# BA Model

**Algorithm** Preferential Attachment

**Require:** Graph $G(V_0, E_0)$, where $|V_0| = m_0$ and $d_v \geq 1 \; \forall \; v \in V_0$, number of expected connections $m \leq m_0$, time to run the algorithm $t$

**return** A scale-free network<br>
`//Initial graph with` $m_0$ `nodes with degrees at least 1`<br>
$G(V, E) = G(V_0, E_0)$;<br>
**for** 1 to $t$ **do**<br>
    $V = V \cup \{v_i\}$; `// add new node` $v_i$<br>
    **while** $d_i \neq m$ **do**<br>
        Connect $v_i$ to a random node $v_j \in V, i \neq j$ ( i.e., $E = E \cup \{e(v_i, v_j)\}$ ) with probability $P(v_j) = \frac{d_j}{\sum_k d_k}$.<br>
    **end while**<br>
**end for**<br>
Return $G(V, E)$

# Real-world vs. Scale-free Networks

1) Compute the expected degree $m$ from real networks
2) Generate the scale-free network based on $m$ and the number of nodes in real networks

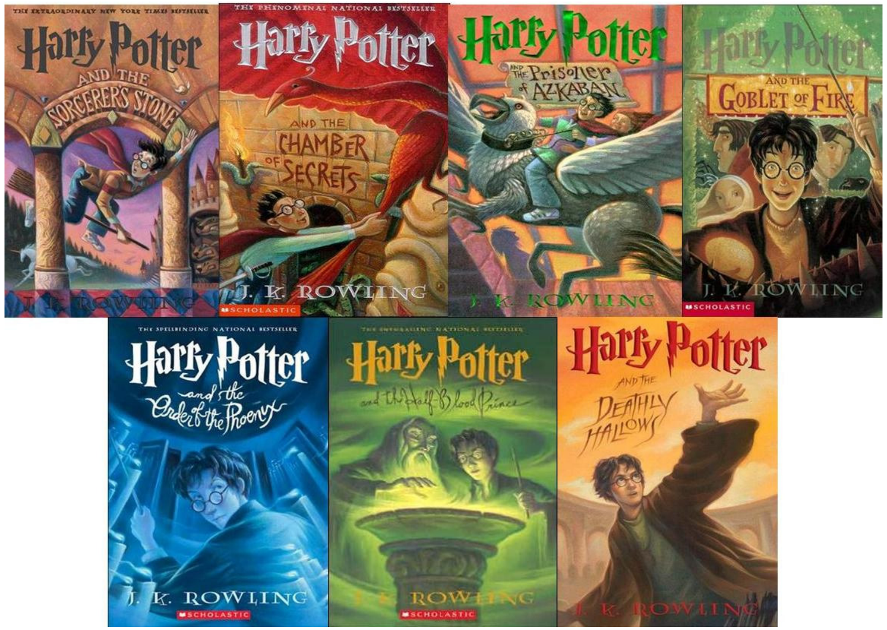
| | **Original Network** | | | | **Simulated Graph** | |
|---|---|---|---|---|---|---|
| **Network** | *Size* | *Average Degree* | *Average Path Length* | $C$ | *Average Path Length* | $C$ |
| Film Actors | 225,226 | 61 | 3.65 | 0.79 | 4.90 | $\approx 0.005$ |
| Medline Coauthorship | 1,520,251 | 18.1 | 4.6 | 0.56 | 5.36 | $\approx 0.0002$ |
| E.Coli | 282 | 7.35 | 2.9 | 0.32 | 2.37 | 0.03 |
| C.Elegans | 282 | 14 | 2.65 | 0.28 | 1.99 | 0.05 |

# Real-world vs. Scale-free Networks

- Are real networks like BA graphs? **Real**
  - Average Path Length: Low $l \approx \frac{\ln N}{\ln \ln N}$ Low ☺
  - Clustering Coefficient: Low $C = \frac{m_0 - 1}{8} \frac{(\ln t)^2}{t}$ High ☹
  - Degree Distribution: Power-law Power-law ☺
- Other characteristics of BA graphs
  - The generated graph is connected (giant component)
  - The older gets richer

# Unpredictability of the Rich-Get-Richer Effects

- The initial stages of one's rise to popularity are fragile
- Once a user is well established, the rich-get-richer dynamics of popularity is likely to push the user even higher
- **BUT** getting the rich-get-richer process started in the first place is full of potential accidents and near-misses




If we could roll time back to 1997, and then run history forward again, would the *Harry Potter* books again sell hundreds of millions of copies?

**See more:** Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. "Experimental study of inequality and unpredictability in an artificial cultural market." *Science* 311.5762 (2006): 854-856.

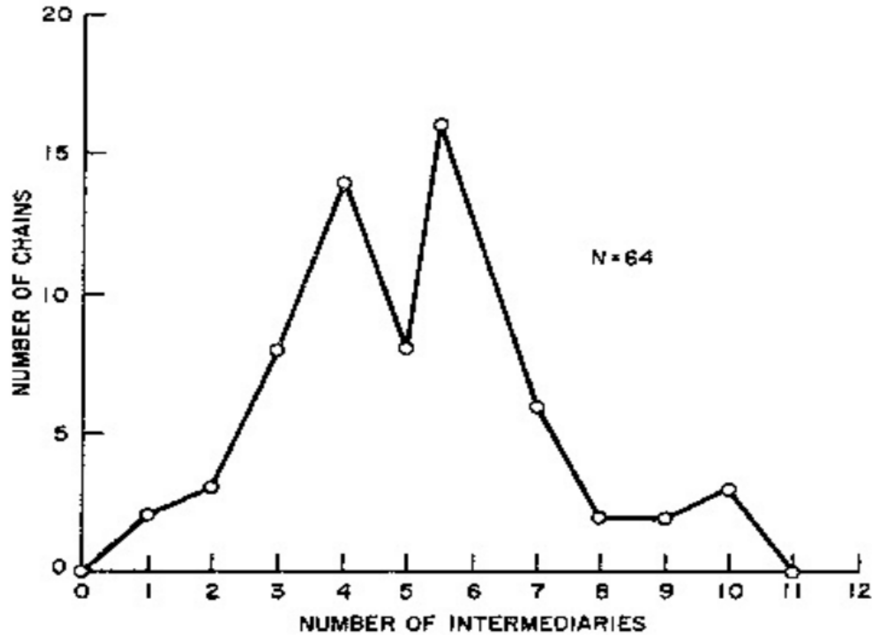
# Watts-Strogatz Model

## Small-world Network

- (**31894** cites) D. J. Watts, S. H. Strogatz. “Collective dynamics of 'small-world' networks.” Nature, 1998.
- (**5370** cites) D. J. Watts. “Small worlds: the dynamics of networks between order and randomness.” Princeton University Press, 1999.
- (**1406** cites) D. J. Watts. “Networks, Dynamics, and the Small-World Phenomenon.” American Journal of Sociology, 1999.
- (**871** cites) P. S. Dodds, R. Muhamad, D. J. Watts. “An Experimental Study of Search in Global Social Networks.” Science, 2003.

# Small-world Phenomenon

- "Six Degree of Separation"
- Milgram's Experiment [1969]
  - 64/296 (25%) reach
  - Average path length = **6**




# Observation in Real-world Networks

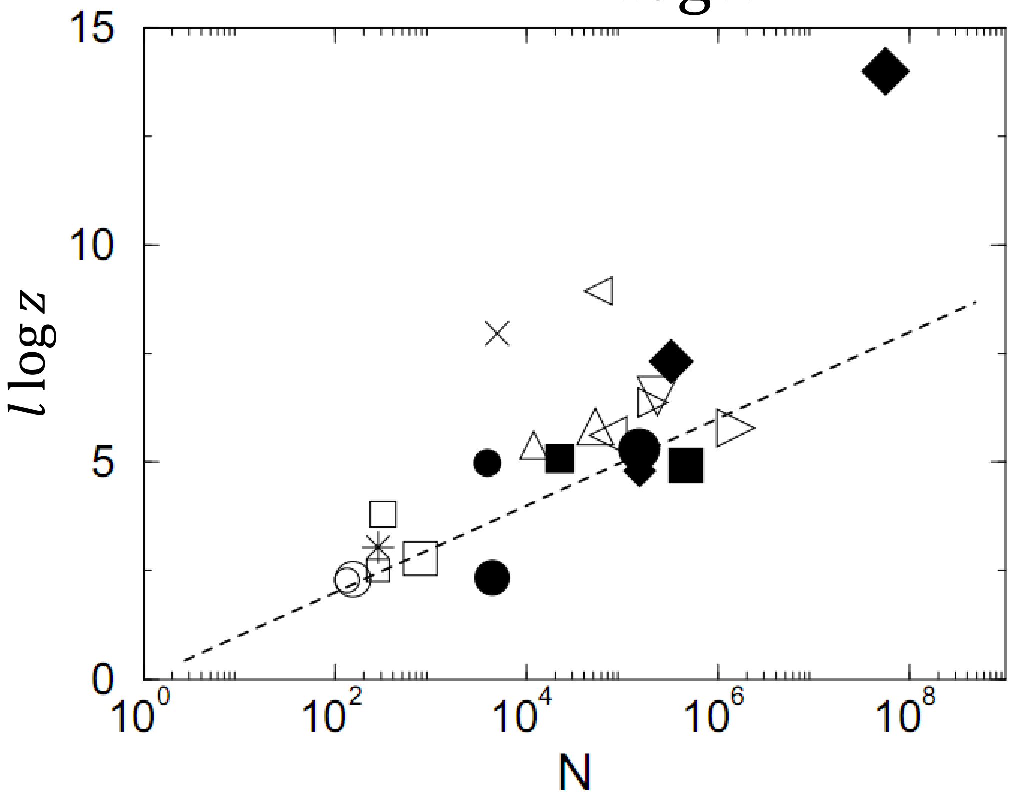
$l \approx \frac{\log n}{\log z}$ $\quad$ $C = p \approx \frac{z}{n}$

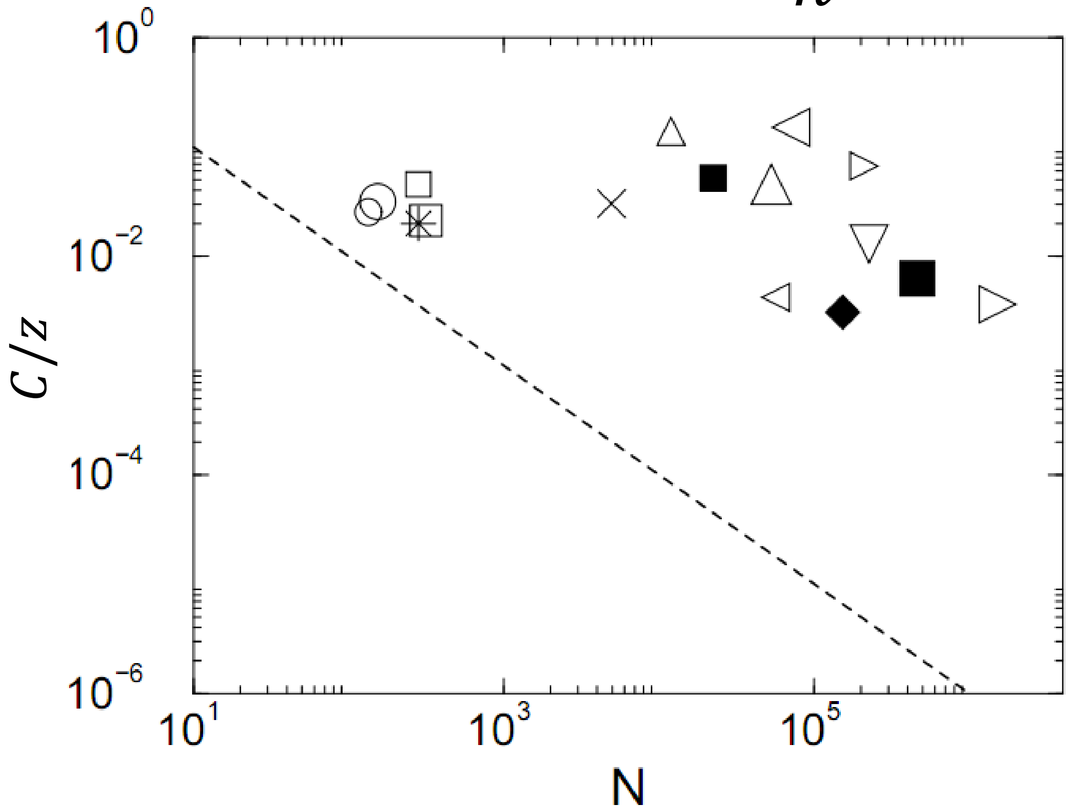
| Network | Size | $\langle k \rangle$ | $\ell$ | $\sim$ $\ell_{rand}$ | $C$ | $>$ $C_{rand}$ |
|---|---|---|---|---|---|---|
| WWW, site level, undir. | 153, 127 | 35.21 | 3.1 | 3.35 | 0.1078 | 0.00023 |
| Internet, domain level | 3015 - 6209 | 3.52 - 4.11 | 3.7 - 3.76 | 6.36 - 6.18 | 0.18 - 0.3 | 0.001 |
| Movie actors | 225, 226 | 61 | 3.65 | 2.99 | 0.79 | 0.00027 |
| LANL coauthorship | 52, 909 | 9.7 | 5.9 | 4.79 | 0.43 | $1.8 \times 10^{-4}$ |
| MEDLINE coauthorship | 1, 520, 251 | 18.1 | 4.6 | 4.91 | 0.066 | $1.1 \times 10^{-5}$ |
| SPIRES coauthorship | 56, 627 | 173 | 4.0 | 2.12 | 0.726 | 0.003 |
| NCSTRL coauthorship | 11, 994 | 3.59 | 9.7 | 7.34 | 0.496 | $3 \times 10^{-4}$ |
| Math coauthorship | 70, 975 | 3.9 | 9.5 | 8.2 | 0.59 | $5.4 \times 10^{-5}$ |
| Neurosci. coauthorship | 209, 293 | 11.5 | 6 | 5.01 | 0.76 | $5.5 \times 10^{-5}$ |
| *E. coli*, substrate graph | 282 | 7.35 | 2.9 | 3.04 | 0.32 | 0.026 |
| *E. coli*, reaction graph | 315 | 28.3 | 2.62 | 1.98 | 0.59 | 0.09 |
| Ythan estuary food web | 134 | 8.7 | 2.43 | 2.26 | 0.22 | 0.06 |
| Silwood park food web | 154 | 4.75 | 3.40 | 3.23 | 0.15 | 0.03 |
| Words, cooccurence | 460.902 | 70.13 | 2.67 | 3.03 | 0.437 | 0.0001 |
| Words, synonyms | 22, 311 | 13.48 | 4.5 | 3.84 | 0.7 | 0.0006 |
| Power grid | 4, 941 | 2.67 | 18.7 | 12.4 | 0.08 | 0.005 |
| *C. Elegans* | 282 | 14 | 2.65 | 2.25 | 0.28 | 0.05 |

# Observation in Real-world Networks

$$l_{rand} \approx \frac{\log n}{\log z}$$

$$C_{rand} = p \approx \frac{z}{n}$$




- Average Path Length

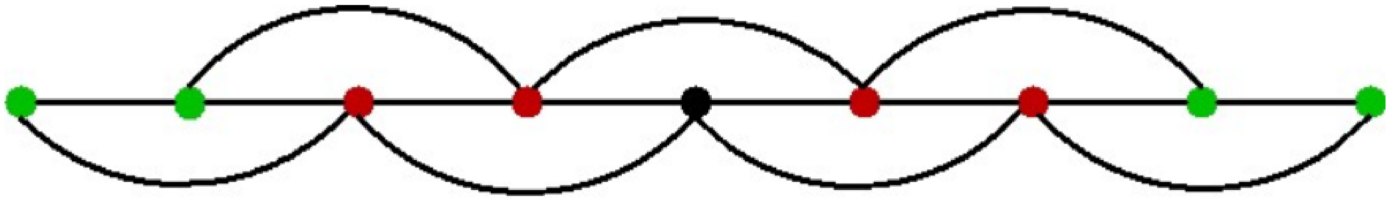
$l_{\mathbf{rand}} \sim l_{\mathbf{real\text{-}world}}$

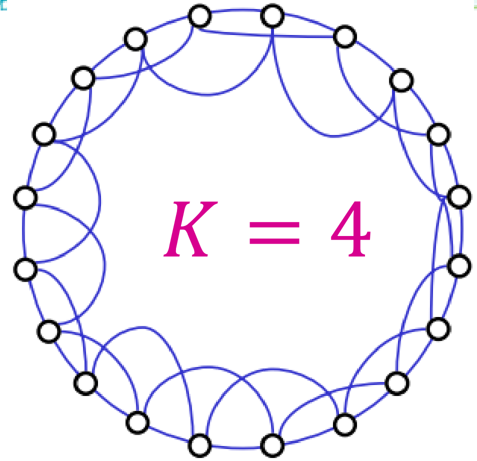
- Clustering Coefficient

$C_{\text{real-world}}$ appears to be independent of the network size

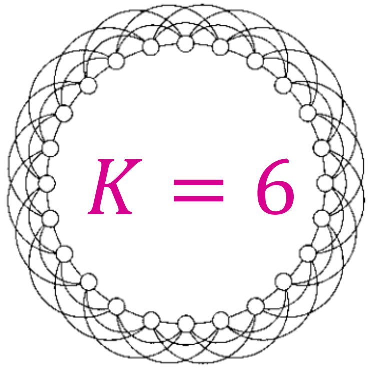
$\Rightarrow$ The Characteristic of **Ring Lattice**!

# Ring Lattice






- Ring lattice with $N$ nodes
  - Each has $\boldsymbol{K}$ neighbors
  - $|E| = KN/2$




- Diameter $1 + \sum_{i=1}^{l_{max}} 4 \approx N \;\; \rightarrow l_{max} \approx \frac{N}{4} = \frac{N}{K}$
- Average Path Length $l \approx \frac{N}{2K}$ Increase much faster than random and real-world networks

APL is **High!**

- Clustering Coefficient

$$C = \frac{C_2^K - (1 + 2 + \cdots + \frac{K}{2})}{C_2^K} = \frac{3(K-2)}{4(K-1)} \approx \frac{3}{4}$$

Independent of network size

CC is **High!**

# Watts-Strogatz Model

Summary of observations

- One extreme: **Random graph**
  - **Low** avg path length, **low** clustering coefficient
- Other extreme: "**Regular**" network (**Lattice**)
  - **High** avg path length, **high** clustering coefficient
- Real-world case: **Small-world Network**
  - **Low** avg path length, **high** clustering coefficient

# Watts-Strogatz Model

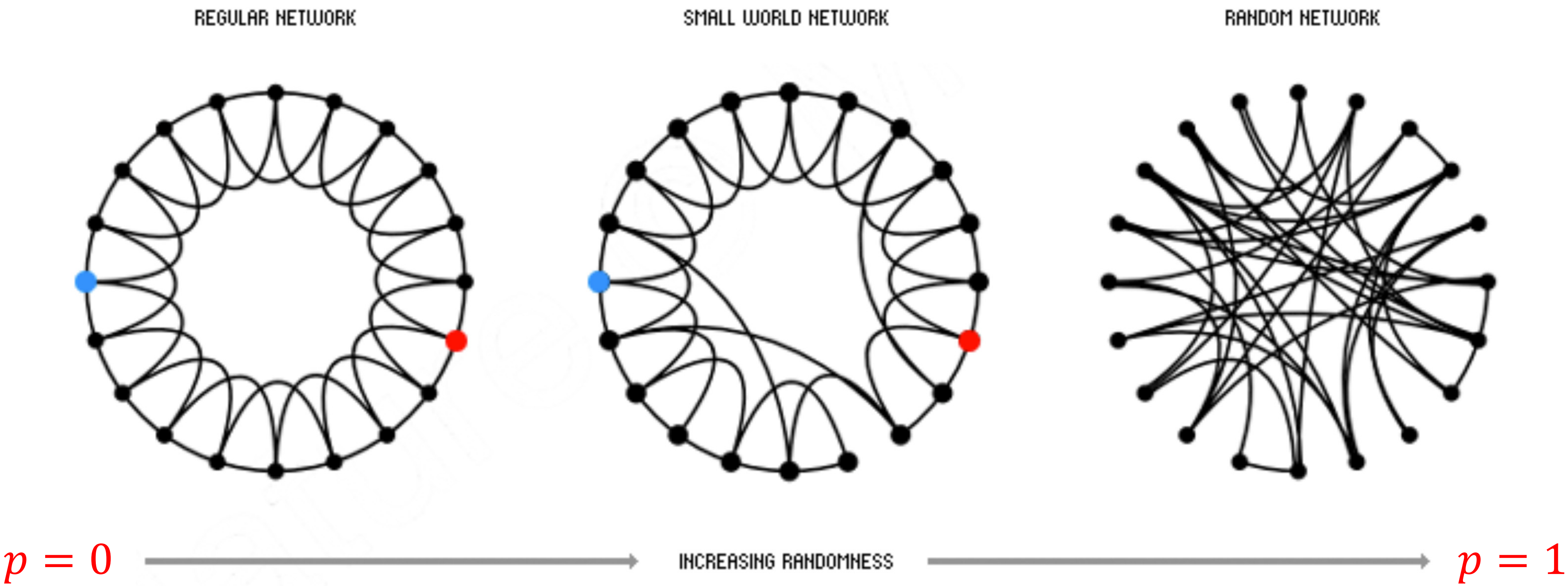
- Interpolate between lattice and random graph!

1. **Start with order** → Keep Locality
   - Start with a ring lattice with $N$ nodes ($K$ neighbors)
   - Friends of friends are friends
   - Ensure sparse and connected network

2. **Random Rewiring** → Create Short-cuts
   - Randomly rewire each edge with probability $p$
   - Shortcuts to join remote parts of the lattice
   - Disallow self-loop and duplicate edges
   - Introduce $pNK/2$ long-range edges

     p*(number of edges)=p*(total_degree / 2)=p*NK/2

# Watts-Strogatz Model




Rewiring allows us to “interpolate” between a regular lattice and a random graph

| | | | |
|---|---|---|---|
| High APL $l \approx \frac{N}{2K}$ | **Low APL** | Low APL $l \approx \frac{\log n}{\log z}$ |
| High CC $C \approx \frac{3}{4}$ | **High CC** | Low CC $C = \frac{z}{n}$ |

# WS Model

**Algorithm** Small-World Generation Algorithm

**Require:** Number of nodes $|V|$, mean degree $c$, parameter $\beta$

**return** A small-world graph $G(V, E)$

$G$ = A regular ring lattice with $|V|$ nodes and degree $c$

**for** node $v_i$ (starting from $v_1$), and all edges $e(v_i, v_j)$, $i < j$ **do**

  $v_k$ = Select a node from $V$ uniformly at random.

  **if** rewiring $e(v_i, v_j)$ to $e(v_i, v_k)$ does not create loops in the graph or multiple edges between $v_i$ and $v_k$ **then**

    rewire $e(v_i, v_j)$ with probability $\beta$: $E = E - \{e(v_i, v_j)\}$, $E = E \cup \{e(v_i, v_k)\}$;

  **end if**

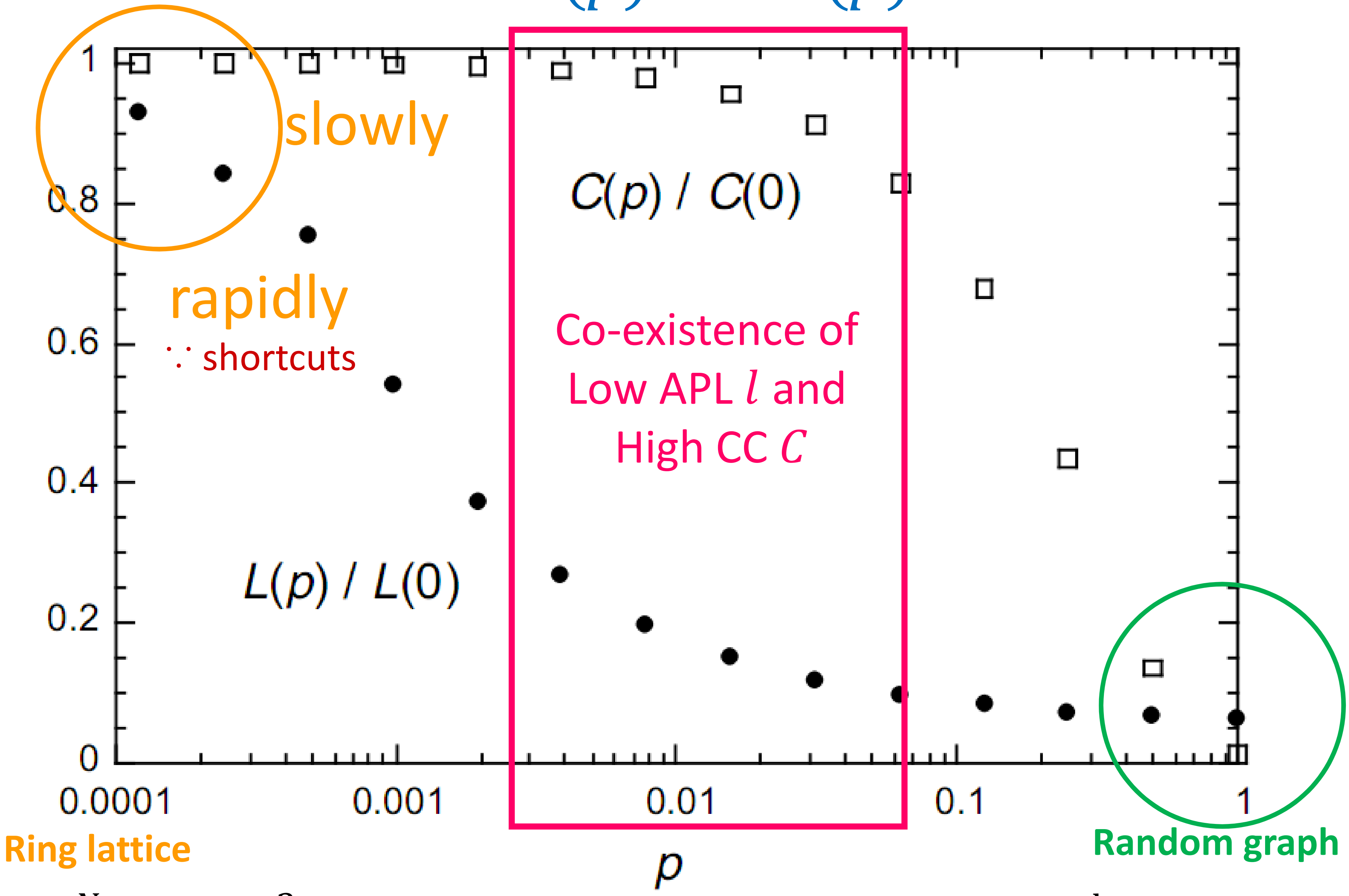
**end for**

Return $G(V, E)$

# Intuition of WS Model

1. Most people are friends with their immediate neighbors (neighbored edges)
   - Neighbors on the same street, colleagues, etc.
2. Everyone has one or two friends who are a long way away (long-range edges by rewiring)
   - People in other countries, old acquaintances

# Trade-off between $l(p)$ and $C(p)$ in WS Model



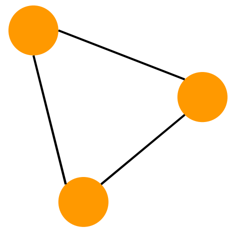

$l(0) \approx \frac{N}{2K}$ $\quad C(0) \approx \frac{3}{4}$

(i.e., % of edges rewired)

$l(1) \approx \frac{\log n}{\log z}$ $\quad C(1) = \frac{z}{n}$

# WS Model: Clustering Coefficient

- The probability that a connected triple stays connected after rewiring consists of



  1) The probability that none of the 3 edges were rewired is $(1-p)^3$
  2) The probability that other edges were rewired back to form a connected triple
     - Very small and can be ignored

- Clustering coefficient: $C(p) \approx (1-p)^3 C(0)$

# Small-world vs. Real-world Networks

- Given a real-world network in which average degree is $z$ and clustering coefficient $C$ is given,
  - Set $C(p) = C$ and determine $p$ using equation

$$C(p) \approx (1-p)^3 C(0)$$

- Given $p$, $z(=K)$, and $n$ (size of network), we can simulate the WS model

$$l(0) \approx \frac{N}{2K} \quad C(0) \approx \frac{3}{4}$$

| | Original Network | | | | Simulated Graph | |
|---|---|---|---|---|---|---|
| **Network** | *Size* | *Average Degree* | *Average Path Length* | $C$ | *Average Path Length* | $C$ |
| Film Actors | 225,226 | 61 | 3.65 | 0.79 | 4.2 | 0.73 |
| Medline Coauthorship | 1,520,251 | 18.1 | 4.6 | 0.56 | 5.1 | 0.52 |
| E.Coli | 282 | 7.35 | 2.9 | 0.32 | 4.46 | 0.31 |
| C.Elegans | 282 | 14 | 2.65 | 0.28 | 3.49 | 0.37 |

# Degree Distribution of Small-world Network

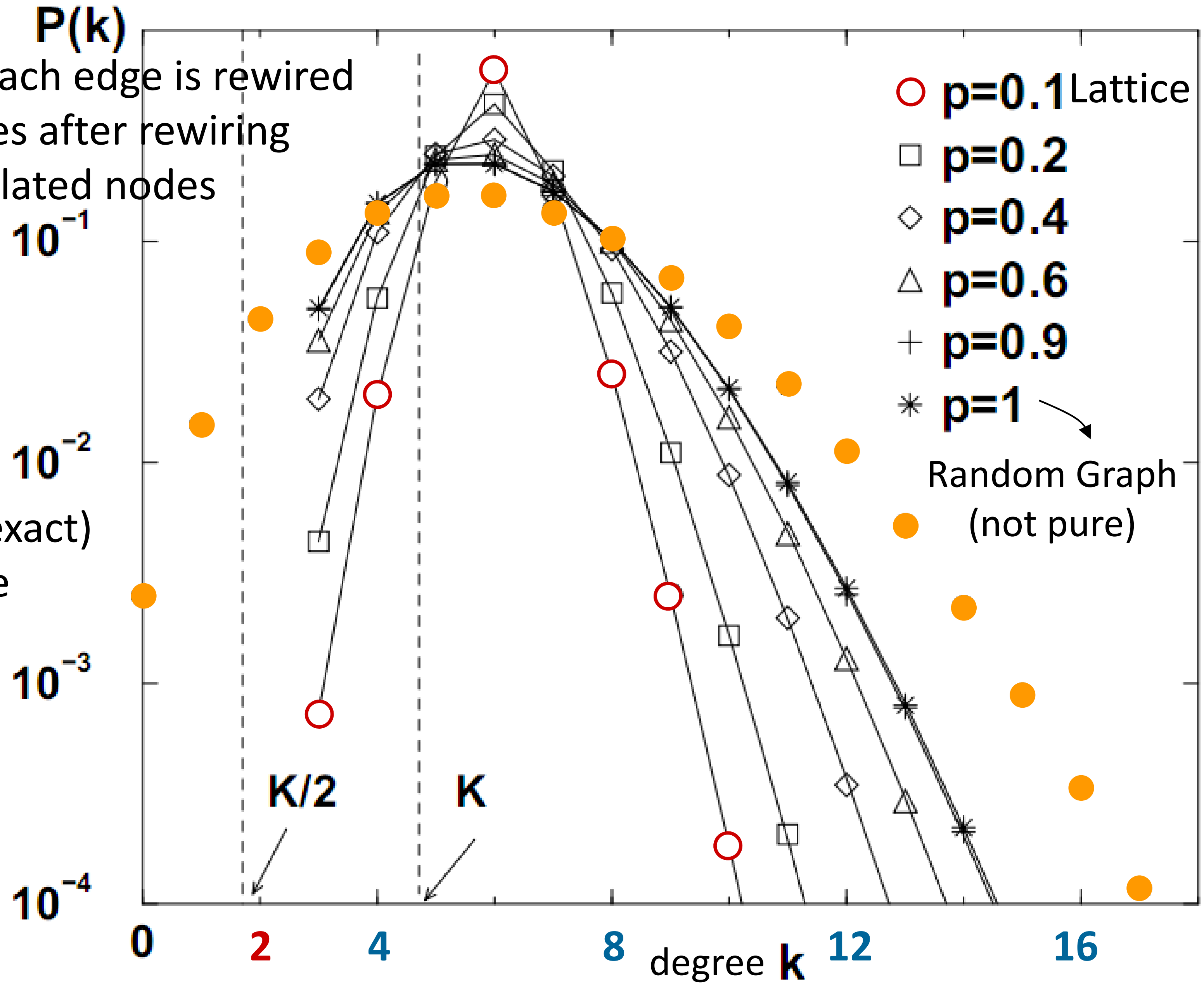
∵ Only one end of each edge is rewired

→ At least $K/2$ edges after rewiring

→ For $K > 2$, no isolated nodes

$|E| = KN/2$

●: Random graph (exact) with average degree



# Real-world vs. Small-world Networks

- Are real networks like WS graphs?
  - Average Path Length: Low
  - Clustering Coefficient: High
  - Degree Distribution: not Power-law

| Real | |
|---|---|
| Low | ☺ |
| High | ☺ |
| Power-law | ☹ |

# Network Generative Models: Summary

| | Random Graph ER Model | Scale-free Network BA Model | Small-world Network WS Model | Ring Lattice |
|---|---|---|---|---|
| Average Path Length | Short $\frac{\log N}{\log z}$ | Short $\frac{\ln N}{\ln \ln N}$ | Short | Long $\frac{N}{2K}$ |
| Clustering Coefficient | Low $\frac{z}{n}$ | Low $\frac{m_0 - 1}{8} \frac{(\ln t)^2}{t}$ | High $(1-p)^3 C(0)$ | High $\frac{3}{4}$ |
| Degree Distribution | Binomial/ Poisson | Power-Law | ≈Binomial/ Poisson | Fixed Value |