Social Network Analysis – Network Dynamics

Network Growth Models

Acknowledgement: Tanmoy Chakraborty

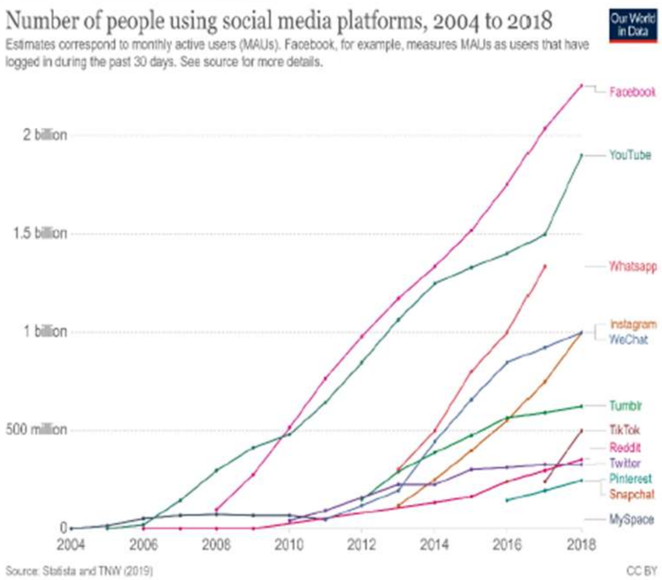
Overview

- Properties of Real-world Networks
- Erdős and Rényi Model
- Watts-Strogatz Model
- Barabasi-Albert Network Model

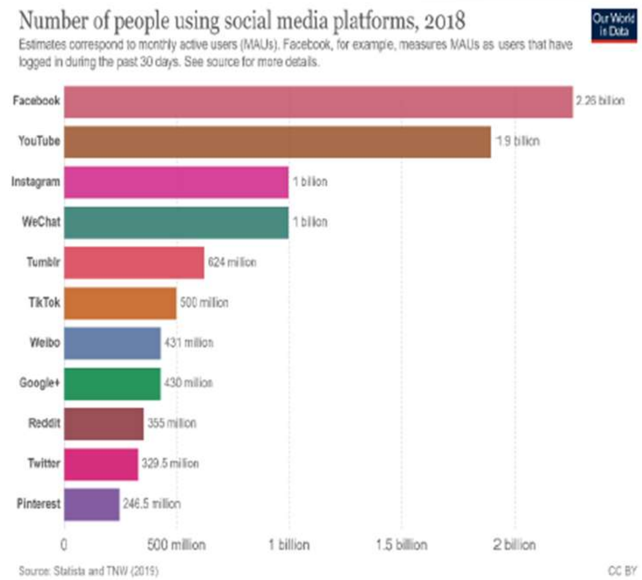
Overview

- Properties of Real-world Networks
- Erdős and Rényi Model
- Watts-Strogatz Model
- Barabasi-Albert Network Model

Rise of Online Social Networks



- Online social networks growing rapidly year-by-year
- Their sizes are huge
- In 2014, Facebook possessed 1.39 billion active users and 400 billion friendship links
- March 2015, Twitter has 288 million active users and 60 billion followers



Such huge volume of these online networks restrict the active research with real social networks

https://ourworldindata.org/rise-of-social-media

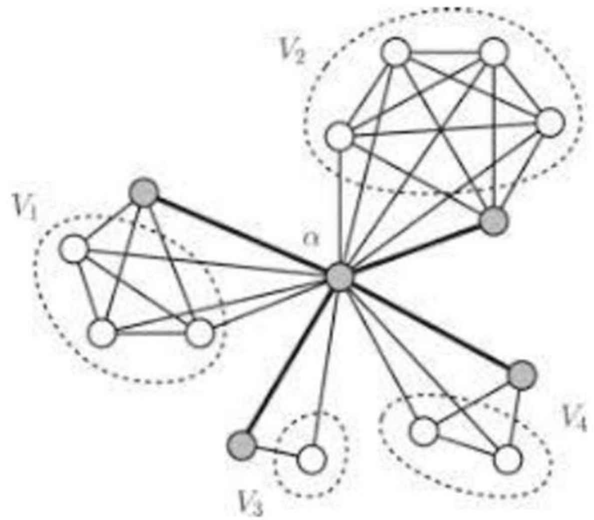
Synthetic Networks

- Generated using theoretical network models
- Often possesses strong underlying mathematical foundation
- Often can simulate important real world network characteristics
- Help getting insights of the real-life networks
- Allow experimentation through simulation when real networks are unavailable
- Can establish network insights on concrete theoretical foundations

Properties of Real-world Social Networks

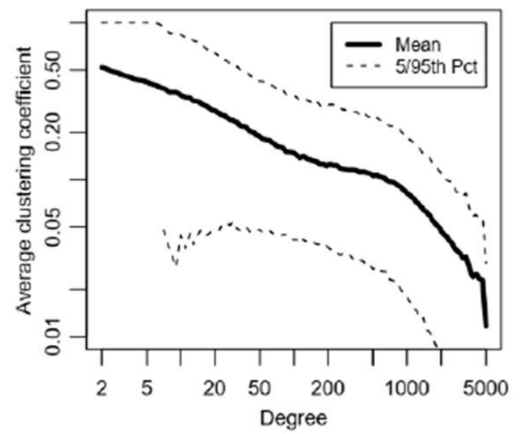
I. High average local clustering coefficient
II. Small world property
III. Scale free property

High Average Local Clustering Coefficient



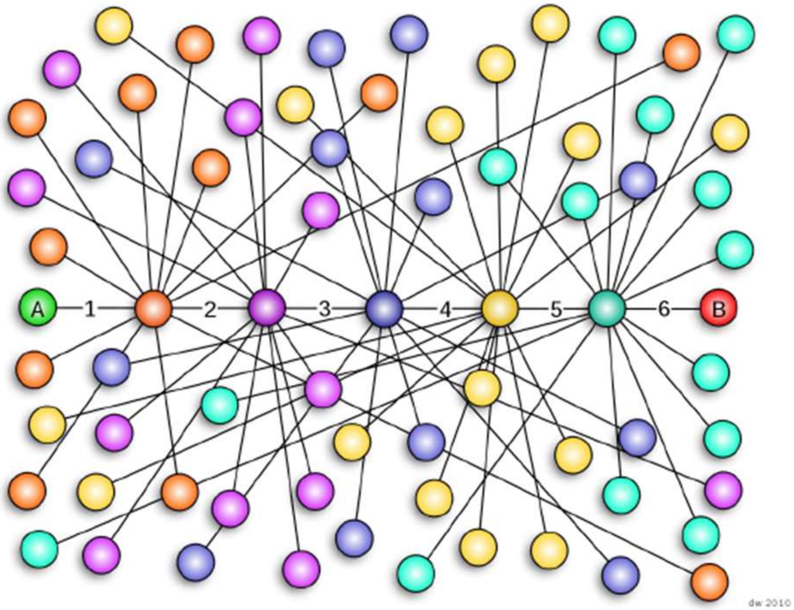
https://www.sciencedirect.com/science/article/pii/S0166218X16304589

- Neighbors of a node tend to be highly connected with each other at individual level
- In Facebook social graph, for users with 100 friends have average local clustering coefficient of 0.14
- For a graph of 150 million nodes, the number is unexpectedly high



Clustering coefficient of Facebook Social Graph
Ugander et al. (2011)

Small-world Property



The "six degrees of separation" model

https://en.wikipedia.org/wiki/Small-world_experiment

- An outcome of Stanley Milgram's small-world experiment (1967) to measure the probability of two random persons being known to each other
- Can also be viewed in light of the average path length between two randomly chosen nodes in a network
- A network G is said to follow the small-world property if the average path length of the network is logarithmically proportional to the network size.

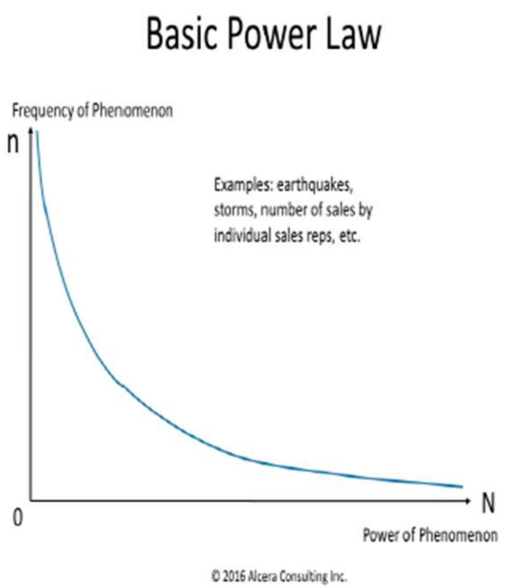
$$Average\ Path\ Length \propto \log(Network\ Size)$$

- Average chain of contacts in Microsoft Messenger was 6.6 people [Leskovec and Horvitz 2007]
- Average distance between two random Facebook users in 2016 was 4.57 with 3.57 intermediaries [Experiment by Facebook in 2016]

Scale-free Property

- A scale-free network is a network where the number of connections to each node follows a power law distribution:
- **Power law distribution**: Most nodes have few connections, while a few nodes, called hubs, have many connections.
- For an undirected network, we can just write the degree distribution as $P_{deg}(k) \propto k^{-\gamma}$, where γ is some exponent.
- This form of $P_{deg}(k)$ decays slowly as the degree k increases, increasing the likelihood of finding a node with a very large degree.
- Networks with power-law distributions a called scale-free because power laws have the same functional form at all scales.
- Most real-world networks are scale free

Power Laws



https://exploitingchange.com/2016/09/14

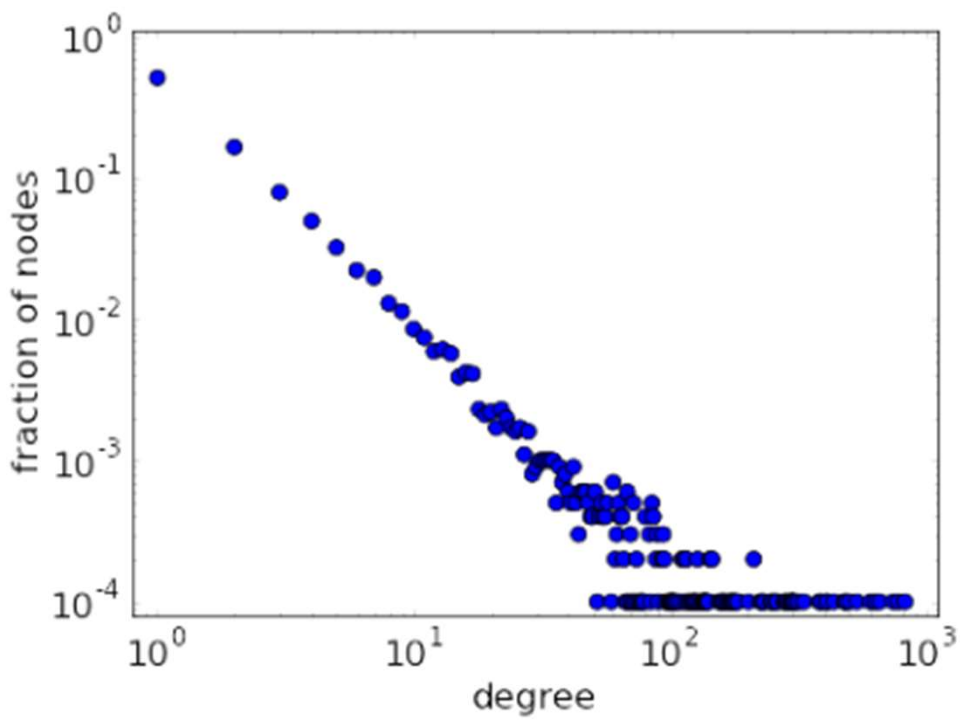
- ❑ Power Law: a relative change in the value of one variable leads to the proportional change in the value of other variable
- ❑ Independent of the initial values of both the variable
- ❑ Mathematically,

$$y \propto x^{-b}, b \in \mathbb{R}$$

- ❑ Functions that follows power-law are scale-invariant
- ❑ Pareto Principle (or the 80/20 rule) in Economics: 80% of the outcomes are results of 20% of the causes
- ❑ Power law principle is also coined as Pareto Distribution

- Scale invariance is a property of objects or laws that remain the same when the scale of length, energy, or other variables are multiplied by a common factor.
- In other words, the behavior or structure of a system does not change regardless of the scale at which it is observed.

Scale-free Property



- The degree distribution of a scale-free network of N=10,000 nodes and power-law exponent γ=2
- One can recognize that a degree distribution has a power-law form by plotting it on a log-log scale. As shown in the above scatter plot, the points will tend to fall along a line.

Real-world Networks

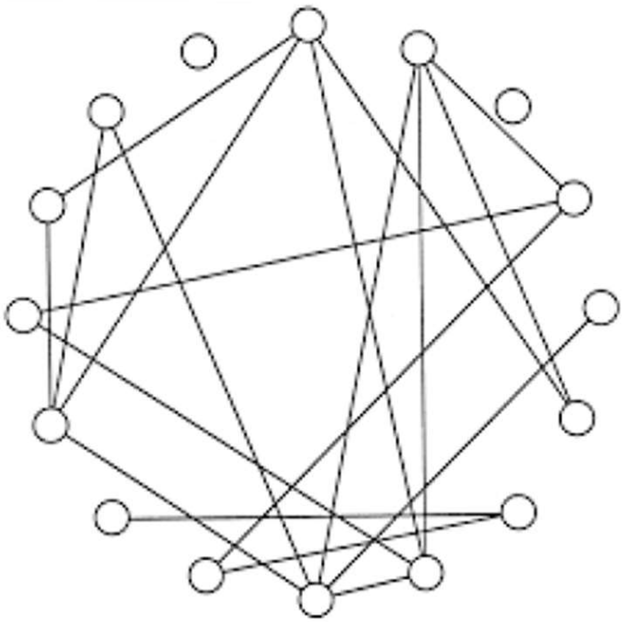
Network	N	L	‹k›	‹d›	d_{max}	ln*N*/ln‹*k*›
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

- k: Average degree d: Average path length

Overview

- Properties of Real-world Networks
- Erdős and Rényi Model
- Watts-Strogatz Model
- Barabasi-Albert Network Model

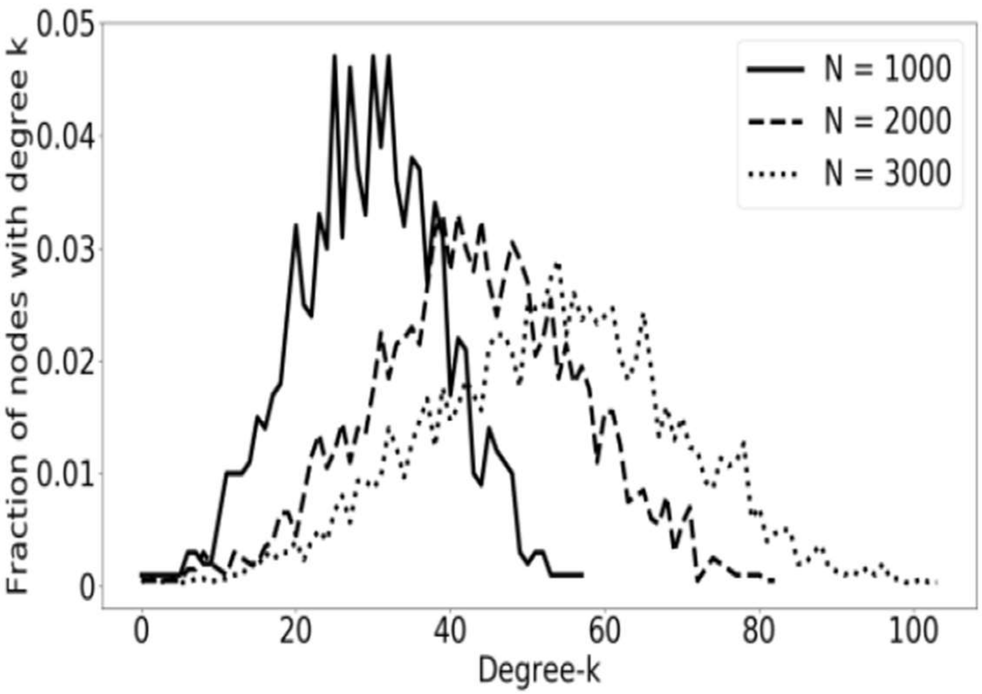
Random Network Model



An instance of $G(16, \frac{1}{7})$ network

- ❑ Also popularly known as Erdős and Rényi model (or ER model)
- ❑ A number of variants of the model
- ❑ Popular variants:
  - ❑ $G(N, K)$ model [Erdős and Rényi 1959]: From the set of all networks of N nodes and K edges, a network is chosen uniformly at random
  - ❑ $G(N, p)$ model [Gilbert 1959]: Network has N nodes, and any random pair of nodes has a probability p of being adjacent independently with any other pair of nodes in the network
- ❑ Both the variants behave identically in the limiting case
- ❑ $G(N, p)$ model considered as the standard random network model

Erdős-Rényi Network: Degree Distribution



Comparison of degree distribution

- For a node in $G(N,p)$ to have degree k, the corresponding node must be adjacent to k other nodes of the network
- We can choose these nodes in $\binom{N-1}{k}$ ways
- Probability $P(k)$ of a node to have degree k is given by

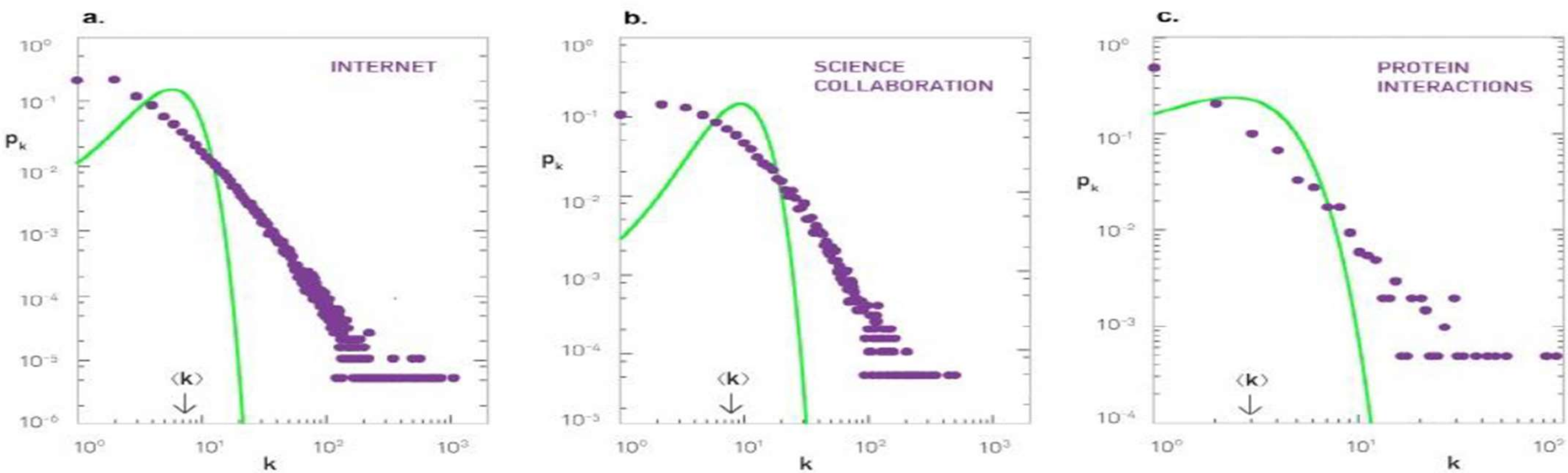
$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- Letting $N \to \infty$,

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

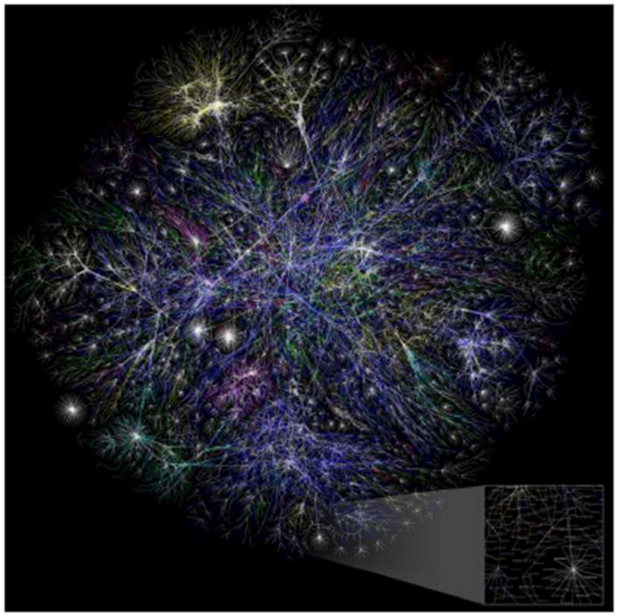
Degree distribution is Binomial

Erdős-Rényi Networks vs. Real-life Networks: Degree Distribution



Real-life networks are often scale-free; however, Erdős-Rényi Networks are not
http://networksciencebook.com/chapter/3#not-poisson

Erdős-Rényi Networks vs. Real-life Networks: Presence of Outliers



Partial map of the Internet based on the January 15, 2005. Hubs are highlighted
https://en.wikipedia.org/wiki/Hub_(network_science)

- In an Erdős-Rényi Network, probability of having a node with a high degree is extremely low
- probability of a node with 2000 neighbours is 10^{-27}!
- In real-world networks, such nodes exist (Hubs)
- Celebrities in social networks

Erdős-Rényi Network: Emergence of Giant Component

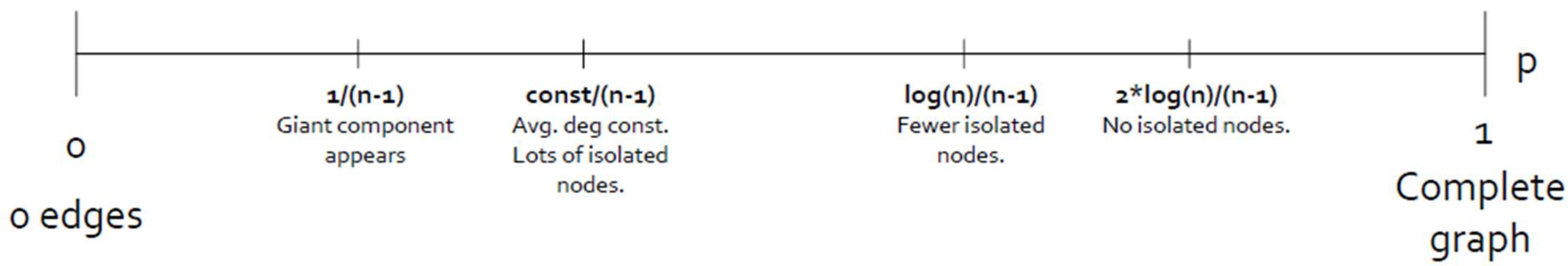
Theorem: A giant component emerges in a random network when the average degree of the network is greater than or equal to unity, i.e., $\langle k \rangle \geq 1$.

- For emergence of a giant component, only one link per node on-an-average is sufficient!! The above condition necessary, too
- The emergence is not a smooth, gradual process; it follows a second-order phase transition
- Regimes of evolution:
  - Subcritical Regime ($0 < \langle k \rangle < 1$) $\Rightarrow$ A number of small isolated clusters in the network, as the number of links is much less than the number of nodes
  - Critical point ($\langle k \rangle = 1$) $\Rightarrow$ A distinguishable giant component emerges
  - Supercritical Regime ($\langle k \rangle > 1$) $\Rightarrow$ A growing giant component, and less and less smaller isolated clusters and nodes
  - Connected Regime ($\langle k \rangle > \ln N$) $\Rightarrow$ The giant component absorbs all nodes and components, the network becomes connected

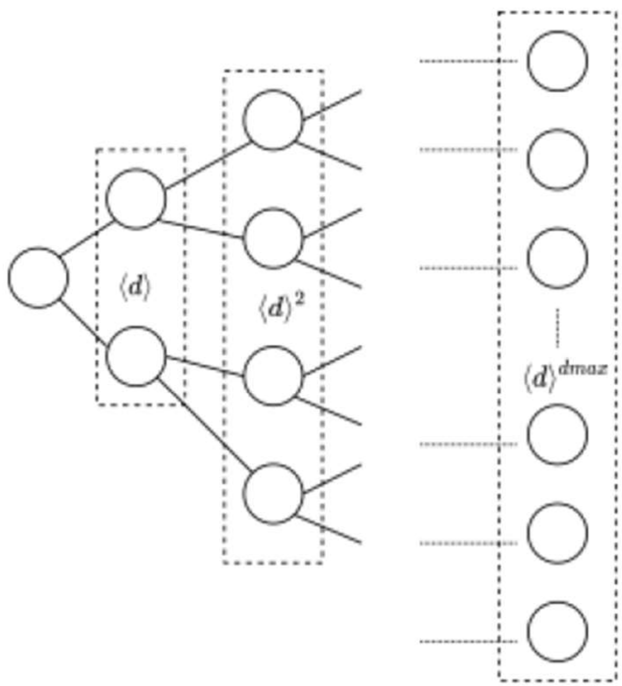
A second-order phase transition is a continuous transition from one phase to another without the two phases coexisting in equilibrium

Graph Structure – Dependence on p

- **Graph structure as *p* changes:**



Erdős-Rényi Network: Average Path Length



A depiction of random network in tree format

- When l_{max} represent the maximum path length of $G(N,p)$,

$$1 + \langle d \rangle + \langle d \rangle^2 + \langle d \rangle^3 + \cdots\cdots\cdots + \langle d \rangle^{l_{max}} = N$$

- When $\langle d \rangle \gg 1$, the above yields,

$$l_{max} \approx \frac{logN}{log\langle d \rangle}$$

- Further approximation yields,

$$\langle l \rangle \propto logN$$

Theorem: Erdős-Rényi Networks follow small-world property.

Erdős-Rényi Network: Clustering Coefficient

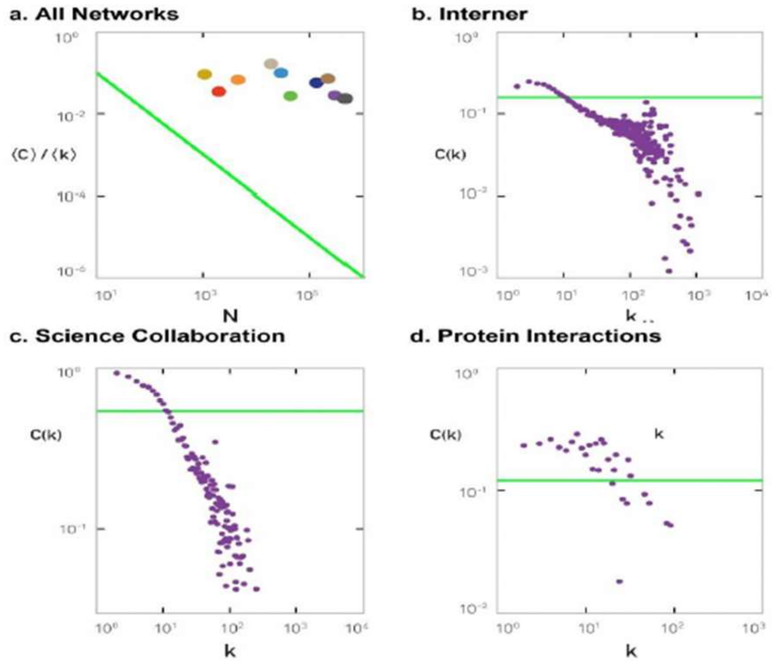
- In $G(N,p)$
  - Number of possible edges between neighbours of a node: $\binom{\langle d \rangle}{2}$
  - Expected number of edges between these nodes: $p \times \binom{\langle d \rangle}{2}$
- The above yields, the local clustering coefficient for a node $v_i \in G$

$$C_i = p \approx \frac{\langle d \rangle}{N}$$

Theorem: The local clustering coefficient for any node in an Erdős-Rényi Network is inversely proportional to the size of the network

Note: The local clustering coefficient for any node in an Erdős-Rényi Network does not depend on the degree of the node

Erdős-Rényi Networks vs. Real-life Networks: Clustering Coefficient



Local Clustering coefficients in real-life networks
http://networksciencebook.com/chapter/3#clustering-3-9

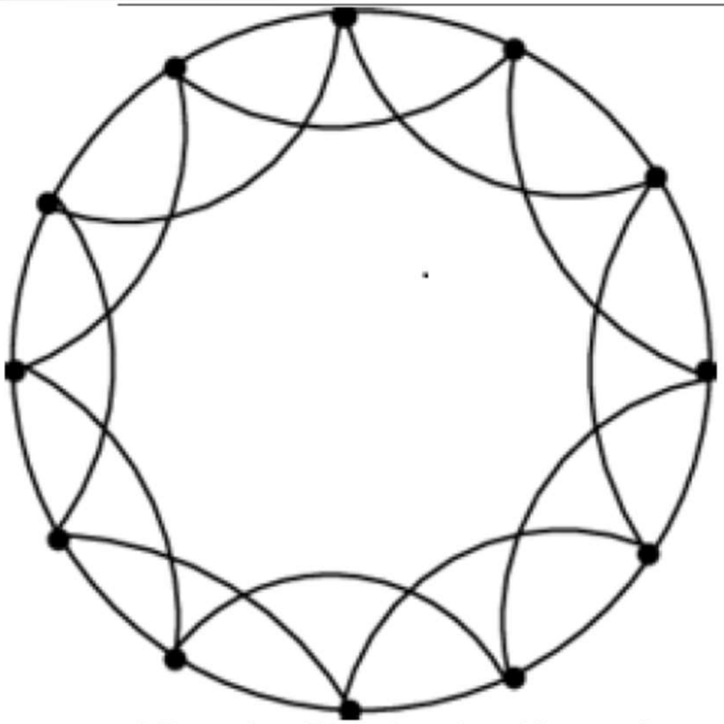
- local clustering coefficients in Erdős-Rényi Networks decreases with the increase in network size
- Nodes having high local clustering coefficient exist in enormously large real-life networks
- Echo chambers in large social networks like Facebook

An echo chamber is "an environment where a person only encounters information or opinions that reflect and reinforce their own."

Overview

- Properties of Real-world Networks
- Erdős and Rényi Model
- Watts-Strogatz Model
- Barabasi-Albert Network Model

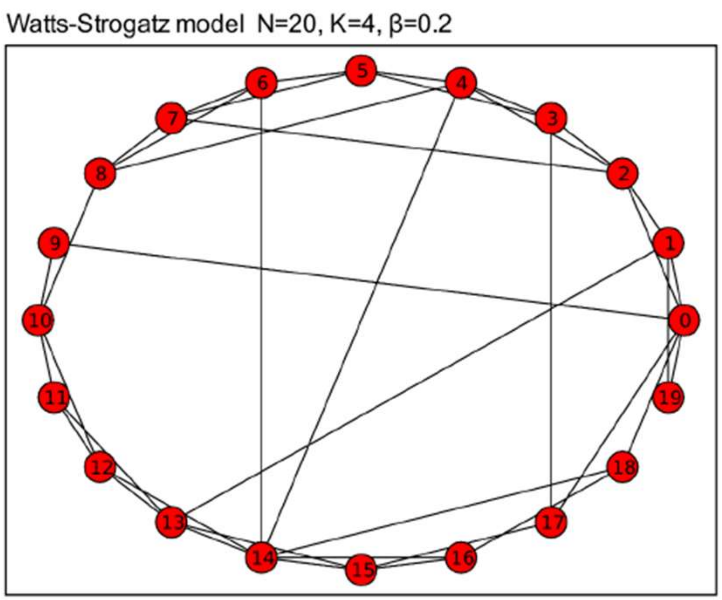
Ring Lattice Network



4-Regular Ring Lattice Network

- ❑ A ring lattice network consists of N nodes labeled $0, 1, 2, \cdots, N-1$ arranged in circular order
- ❑ Every node in the network is connected to exactly k other nodes, immediate $\frac{k}{2}$ rightmost nodes and $\frac{k}{2}$ leftmost nodes relative to the position of the node in the network

Watts-Strogatz Network Model: Network Formation



https://www.hindawi.com/journals/mpe/2014/693743/fig2/

1) Start with a k-regular lattice network of size N
2) List the nodes of the lattice as $1, 2, \cdots, N$
3) Choose i^{th} node from the list, $i = 1, 2, \cdots, N$
4) Select edges that that link i^{th} node to some j^{th} node $(j > i)$
5) With a fixed rewiring probability β, rewire the other end of these edges
6) Avoid formation self-loops and link-duplication
7) Repeat steps 3 through 6 until all the nodes are scanned

Watts-Strogatz Network Model: Properties

- Locally clustered network due to the underlying ring lattice structure

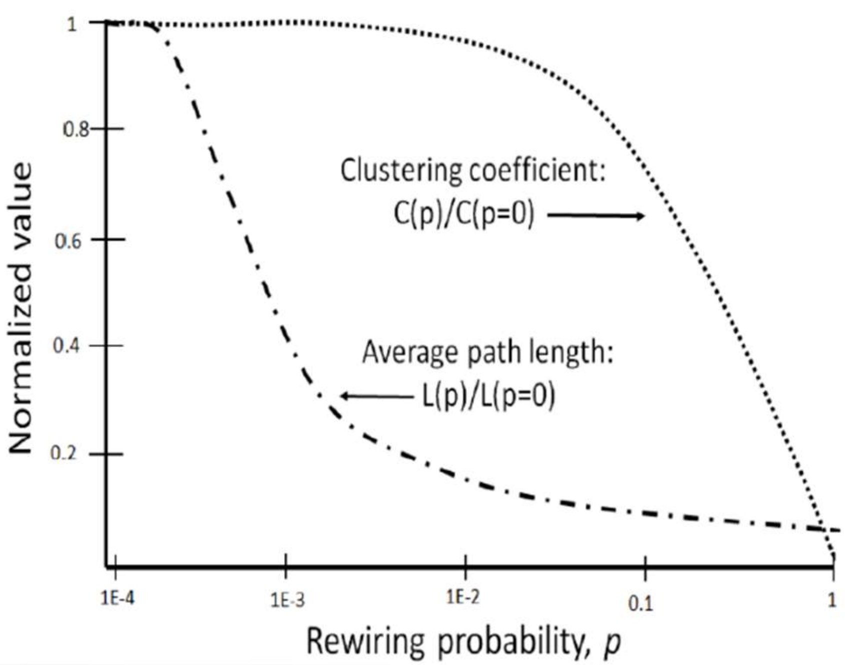
- Random rewiring of links reduces the average path length

- $\frac{\beta N k}{2}$ number of non-lattice edges introduced due to random random rewiring

Watts-Strogatz Network Model: Average Path Length

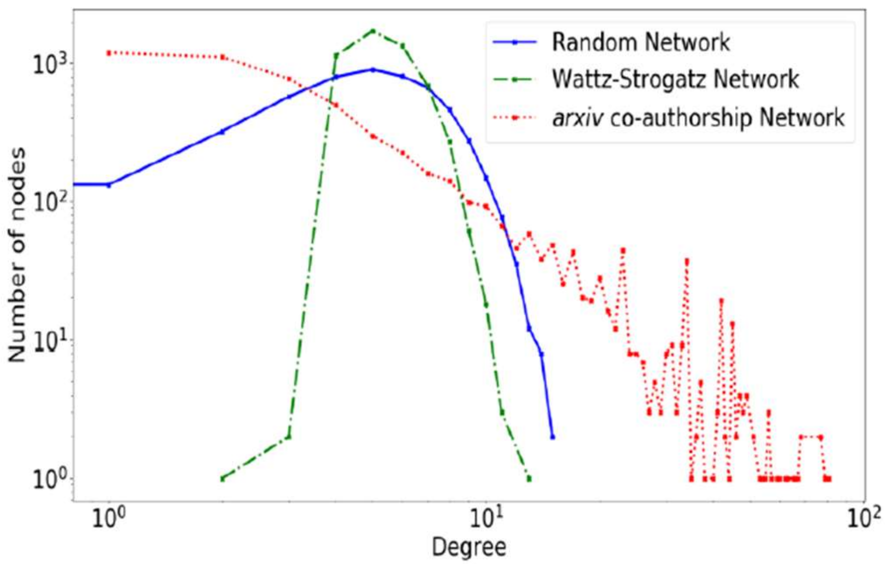
- Average path length in WS network $\approx \frac{N}{2k}$ when $\beta \to 0$
- The above scales linearly with the size of the network
- If $\beta \to 1$, average path length boils down to $\approx \frac{\ln N}{\ln k}$
- If $0 < \beta < 1$, with increase in the value of β, the average path length reduces sharply

Watts-Strogatz Network Model: Clustering Coefficient



- Average path length in WS network $\approx \frac{N}{2k}$ when $\beta \to 0$
- The above scales linearly with the size of the network
- If $\beta \to 1$, average path length boils down to $\approx \frac{\ln N}{\ln k}$
- If $0 < \beta < 1$, with increase in the value of β, the average path length reduces sharply

Watts-Strogatz Networks vs. Real-world Networks



Comparisons of Degree Distribution

- Watts-Strogatz networks have only few outliers, whereas, real-life networks have significantly high number of outliers
- Watts-Strogatz networks follow small-world property if $\beta \rightarrow 1$, however, real-world networks always follow small-world
- Degree distribution in Watts-Strogatz does not follow power law, whereas, real networks often does that
- Both Watts-Strogatz networks and real-world networks have high clustering coefficient

Overview

- Properties of Real-world Networks
- Erdős and Rényi Model
- Watts-Strogatz Model
- Barabasi-Albert Network Model

Barabasi-Albert Network Model

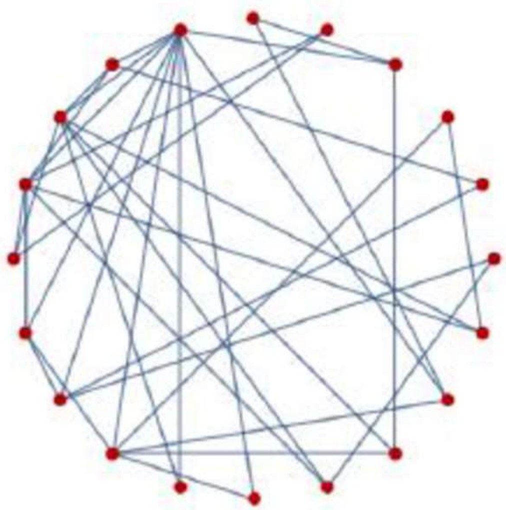
- Real-life social networks often evolve with time
- Originates with a small seed network
- The network grows as new nodes and edges gets attached to the network with time
- Barabasi-Albert model follows the same principle of network evolution
- Also known as Preferential Attachment model or rich gets richer model
- Generates scale-free networks

Barabasi-Albert Model: Network Formation

- Initially, the network has a limited number of nodes, m_0
- At every time-step, a new node with m edges enters the network
- New edges gets attached with the existing nodes based on the principle of preferential attachment as follows:
  - the probability that the new edge attaches to an existing node v_i with degree d_i is:

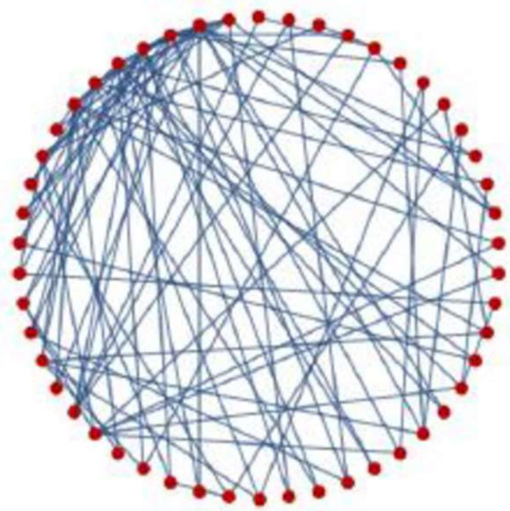
$$P(v_i) = \frac{d_i}{\sum_j d_j}$$

Barabasi-Albert Model: Network Growth



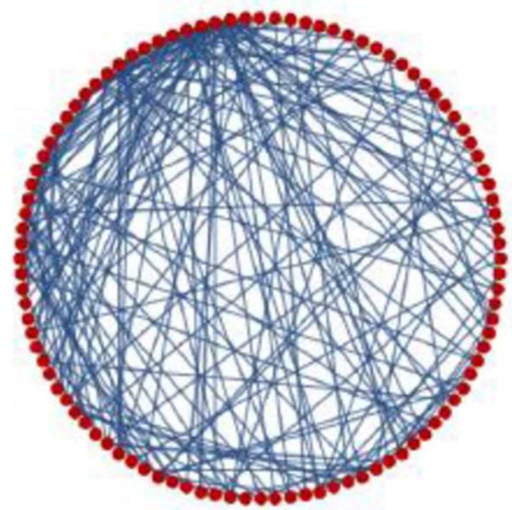
$|G| = 20$

Average Path Length = 2.16



$|G| = 50$

Average Path Length = 2.69



$|G| = 100$

Average Path Length = 3.02

https://www.researchgate.net/publication/259742981_Information_Theory_Kolmogorov_Complexity_and_Algorithmic_Probability_in_Network_Biology

Barabasi-Albert Model: Degree Dynamics

- Captures the continuous changes in degree distribution with addition of nodes and edges into the network with time
- Let the node v_i joined the network at time t_i and has degree d_i at any instance of time
- Then the degree of the above node at current instance is given by the equation

$$d_i(t) = d_i(t_i)\left(\frac{t}{t_i}\right)^{\beta} \cdots\cdots\cdots (*)$$

Where β is the dynamical exponent, and is set as $\frac{1}{2}$ in general

- Expected number of preferential attachment during joining of a node is usually denoted m
- Then the revised equation assumes the form: $d_i(t) = m\left(\frac{t}{t_i}\right)^{1/2} \cdots\cdots\cdots (**)$

Barabasi-Albert Model: Scale-free Property

- Note the equation $(*)$ as follows:

$$d_i(t) = d_i(t_i)\left(\frac{t}{t_i}\right)^{\beta}$$

- This implies

$$d_i(t) \propto t^{\beta}$$

- The above establishes that the growth model follows power law

Barabasi-Albert Model: Scale-free Property

- It implies further from the equation $(*)$

$$d_i(t) \propto \left(\frac{1}{t_i}\right)^{\beta}$$

- earlier the node joins the network, higher would be its degree (First mover advantage)

- It implies even further from equation $(*)$

$$\frac{\partial d_i(t)}{\partial t} \propto \sqrt{\frac{1}{(t_i \cdot t)}}$$

- Thus, the rate of acquiring of new edges by a node slows down with time

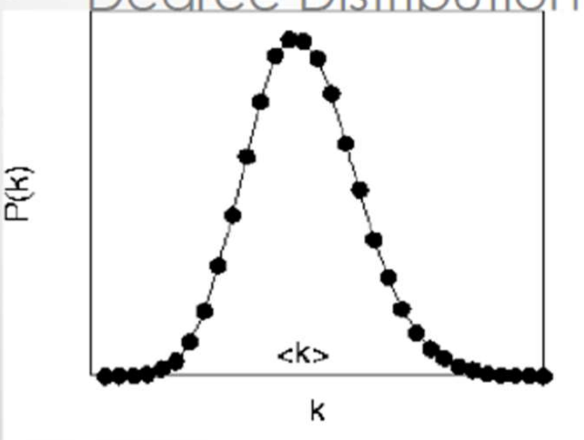
Barabasi-Albert Model: Limitations

- Model assumes that only one node is added at a time; difficult to apply if more than one node arrive simultaneously
- Assumes a linear growth model; may not be realistic in many applications
- Predicts a fixed exponent in the power-law degree distribution; whereas, across real-world networks, exponent varies between 1 and 3
- Model does not capture the temporal decay of preference of a node
- Model does not consider the competitive characteristics of a real-world node that flourish in short notice

Model Comparison

Erdös-Rényi

- Random graphs
- Has small diameter
- Low clustering coefficient
- Degree Distribution



Watts–Strogatz

- Regular graphs
- Random links
- Explains "small world" along with high clustering coefficient

Barabási–Albert

- Preferential Attachment
- Rich-get-richer (Copying model)
- Explains Power Law degree distributions

Reading

- Social Network Analysis. Tanmoy Chakraborty.
  - Chapter 3