# Random Networks.

16 November 2016

**Lecturer:** Dimitra Maoutsa

**Course:** Network Dynamics & Complex Systems
Theoretical and Computational Tools
Georg-August-Universität Göttingen

# Outline

**Network Properties.**

**Random Network Models.**

**Erdős-Rényi $G(n, p)$.**

**Small World Networks.**

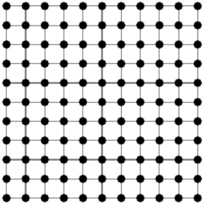
- Watts-Strogatz Model.
- Newman-Watts Model.

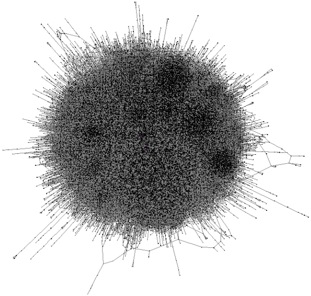
**Scale-free or Growing Random Networks.**

- Barabasi-Albert Model.

## Random Networks

Real world networks do not really have a regular structure.





If we'd try to visualise them, they'd look like hairballs.

Therefore, we need to define quantities to describe their structure.

# Network Properties.

Some key aspects of real complex networks:

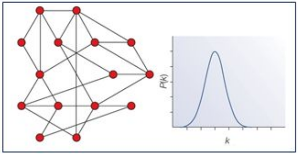
- Degree Distribution
- Clustering Coefficient
- Geodesic distances
- Assortativity
- Motifs

## Degree Distribution $P_k$.

The probability that a randomly selected node has degree $k$.
$k$ = node degree = # of connections
Gives a rough profile of how the connectivity is distributed within the network.

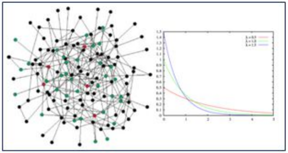
**Poisson:**

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



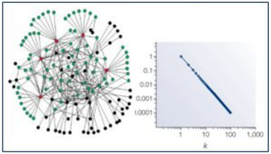

**Exponential:**

$$P(k) \propto e^{-k/d}$$



**Power-law:**

$$P(k) \propto k^{-c}, k \neq 0, c > 1$$



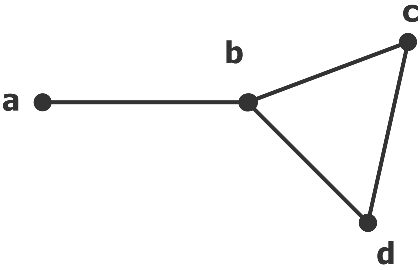
## Transitivity or Clustering Coefficient $CC$.

- **Local Clustering Coefficient ( Watts-Strogatz ) $CC_l$ [1].**
  Number of edges between neighbours of a node, divided by total number of possible edges between those neighbours

$$CC_l = < \frac{\text{\# of pairs of neighbours of i that are connected}}{\text{\# of pairs of neighbours of i}} >_i \quad (1)$$

$$CC_l = < \frac{\sum_{j_1, j_2 \in N_i} a_{j_1 j_2}}{k_i(k_i - 1)/2} >_i, \quad (2)$$

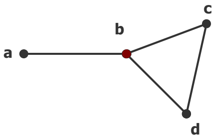
where $N_i$: neighbours of $i$, and $k_i$: degree of $i$.



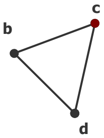

$$CC_l = \frac{1}{n}\frac{\sum_{j_1,j_2 \in N_i} a_{j_1 j_2}}{k_i(k_i-1)/2} = \frac{7}{12}$$

$CC_{l,a} = 0$

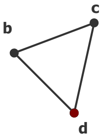

$CC_{l,b} = \frac{1}{3}$




$CC_{l,c} = 1$



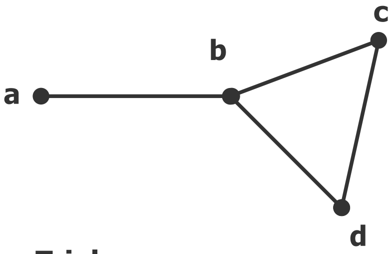

$CC_{l,d} = 1$



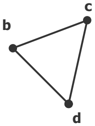
- **Global Clustering Coefficient ( Newman )** $CC_g$ **.**

$$CC_g = \frac{3 \times \text{\#of triangles}}{\text{\#of triples}} \tag{3}$$

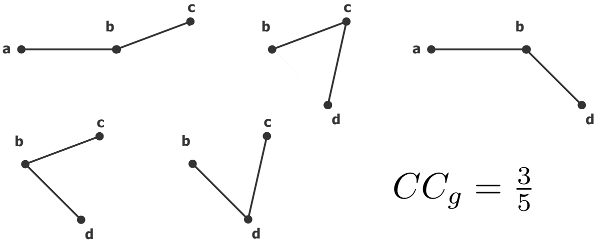
  - o Nodes $i, j, l$ form a ***triangle*** if each pair of nodes is connected.
  - o Nodes $i, j, l$ form a ***triple*** around $i$ if $i$ is connected to $j$ and $l$.




**Triangles:**




**Triples:**




$$CC_g = \frac{3}{5}$$

Alternative formulations of the same clustering coefficient:

$$CC = \frac{6 \times \#\ \text{of triangles}}{\#\ \text{of paths of length } 2} \tag{4}$$

$$CC = \frac{\#\ \text{closed paths of length } 2}{\#\ \text{paths of length } 2} \tag{5}$$

For undirected, unweighted networks:

$$\#\text{of triples} = \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{l=1}^{n}[A^2]_{il} - Tr(A^2)\right) \tag{6}$$

$$\#\text{of triangles} = \frac{1}{6}Tr(A^3) \tag{7}$$

## Geodesic Distances.

- **Shortest path length $\delta$.**
  Fewest possible number of steps from node $i$ to node $j$.

  Let $p(v_i, v_j) = \{v_i, v_{i+1}, ..., v_{j-1}, v_j\}$ be a path from node $v_i$ to node $v_j$.
  The shortest path between these nodes is defined as:

$$\delta(v_i, v_j) = \begin{cases} min\{|p(v_i, v_j)|\}, & \text{if there is a path } p \text{ from } v_i \text{ to } v_j \\ \infty, & \text{otherwise} \end{cases},$$

  where $|p(v_i, v_j)|$ denotes the length of the path $p(v_i, v_j)$.
  ( $\rightarrow$ Dijkstra's algorithm - single source-single end shortest path )

- **Characteristic path length $L$.** Average shortest path in whole network:

$$L = \langle\langle\delta(v_i, v_j)\rangle_i\rangle_j = \frac{1}{n(n-1)} \sum_{i \neq j} \delta(v_i, v_j), \tag{8}$$

where $n$ the size of the network.

- **Diameter $D$.** The longest shortest path in the network:

$$D = max_{v_i, v_j}\{\delta(v_i, v_j)\} \tag{9}$$

( $\rightarrow$ Floyd-Warshall algorithm - all-to-all shortest paths )

## Assortativity or Degree correlation.

In social networks nodes with similar degree connect preferably, while in technological and biological networks the reverse tends to prevail.

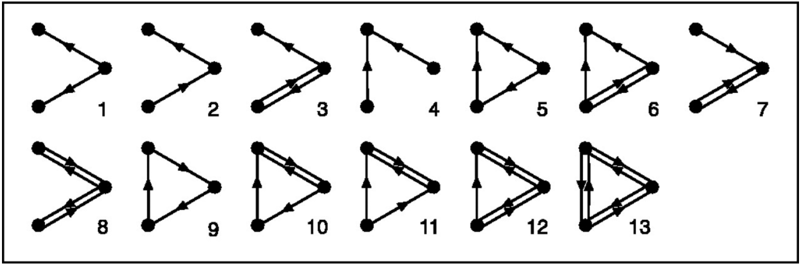
Pearsonś correlation coefficient of node degrees across links:

$$r = \frac{Cov(k_i, k_j)}{\sigma_{k_i}\sigma_{k_j}} \tag{10}$$

| Network | $n$ | $r$ |
| --- | --- | --- |
| Physics coauthorship (a) | 52 909 | 0.363 |
| Biology coauthorship (a) | 1 520 251 | 0.127 |
| Mathematics coauthorship (b) | 253 339 | 0.120 |
| Film actor collaborations (c) | 449 913 | 0.208 |
| Company directors (d) | 7 673 | 0.276 |
| Internet (e) | 10 697 | −0.189 |
| World-Wide Web (f) | 269 504 | −0.065 |
| Protein interactions (g) | 2 115 | −0.156 |
| Neural network (h) | 307 | −0.163 |
| Marine food web (i) | 134 | −0.247 |
| Freshwater food web (j) | 92 | −0.276 |
| Random graph (u) | | 0 |
| Callaway *et al.* (v) | | $\delta/(1+2\delta)$ |
| Barabási and Albert (w) | | 0 |

## Motifs.

Recurring patterns of interactions (*subgraphs*) that are significantly overrepresented with respect to a background model.




[Figure from [2]]

**Finding $s$-node motifs in a network:**

- Scan all $s$-node subgraphs of the given network and count # of appearances of each motif (considering isomorphic architectures).
- Generate an ensemble of random networks with some of the same properties as the original network.
- Count # of appearances of each $s$-node motif in every randomly generated network.
- Compare motif occurrence with background statistics from random ensemble (identify statistical significance).

# Random Network Models.

# Erdős-Rényi $G(n, p)$.

A network made of $n$ nodes.
Each node pair is connected randomly and independently with probability $p$.

$G(n, p)$ is the ensemble of networks with $n$ vertices in which each simple graph $G$ appears with probability:

$$P(G) = p^m (1-p)^{\binom{n}{2} - m} \tag{11}$$

## Exercise.

- Create and plot a few ER random networks.
- Measure their properties for varying $p$:
  - Degree distribution
  - Characteristic path length
  - Clustering coefficient
  - Degree correlation

## Characteristics of Erdős-Rényi random model.

- **Expected number of edges $< m >$.**

Probability of drawing a graph with $m$ edges from the $G(n,p)$ ensemble:

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}. \tag{12}$$

Mean of $P(m)$:

$$< m >= \sum_{m=0}^{\binom{n}{2}} mP(m) = \binom{n}{2} p \tag{13}$$

- **Mean Degree $c$.**

  Mean degree in a graph with exactly $m$ edges:

$$< k >= \frac{2m}{n} \tag{14}$$

  Mean degree in $G(n,p)$:

$$c =< k >= \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n}\binom{n}{2} p = (n-1)p \tag{15}$$

- **Degree distribution $P_k$.**
  Prob. of a node to be connected to $k$ other nodes:

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{16}$$

In the large $n$ limit, $p = c/(n-1) \to 0$ and thus:

$$ln[(1-p)^{n-1-k}] = (n-1-k) ln(1 - \frac{c}{n-1}) \approx -(n-1-k)\frac{c}{n-1} \approx -c \tag{17}$$

$$P_k = \binom{n-1}{k} p^k e^{-c} \approx \frac{(n-1)^k}{k!} \left(\frac{c}{n-1}\right)^k e^{-c} = e^{-c}\frac{c^k}{k!} \tag{18}$$

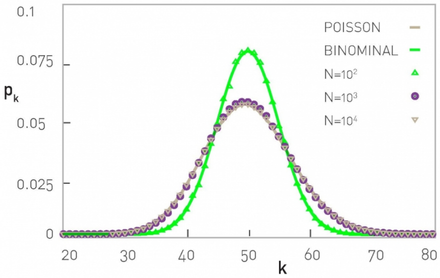
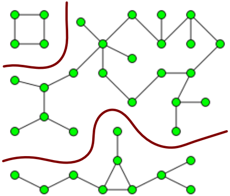



Figure 1: Degree distribution $P_k$ is binomially distributed. For $N \ll c = < k >$ the degree distribution may be well approximated by a Poisson distribution with rate parameter $c$.

## Exercise.

Generate several $G(1000, p)$ networks for varying $p \in [0, 0.005]$ and plot the size of the largest component identified in each network against the average degree $c$.

**Definition.** A component is a subset of nodes in a network, so that there is a path between any two nodes that belong to the component, but one cannot add any more nodes to it that would have the same property.



Graph with 3 components

**Homework.**

Look up the definition the Laplacian of a graph and how it can reveal information regarding the connectedness of a network.

## Giant Component.

**Definition.**

A network component whose size grows in proportion to $n$ is called a **giant component**.

- For $p = 0$, there are no edges in the network, thus we have $n$ components of size $1$.
- For $p = 1$, the network is fully connected, therefore the network has one component of size $n$.
- But what happens in-between?

- Assuming that we are in the large $n$ limit, let us calculate the size of the giant component.

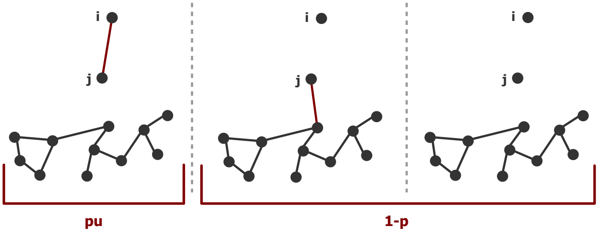
- Define $u$: average fraction of network nodes that **DON'T** belong to the giant component.

  - For $p = 0 \rightarrow u = 1$

- Alternatively, we may interpret $u$ as the probability that a randomly selected node is not part of the giant component.

A node $i$ does not belong to the giant component if it is not connected to it via any other node.
Therefore, any other vertex $j$ in the network is:

- either not connected to $i \rightarrow (1-p)$
- or connected to $i$, but not part of the giant component $\rightarrow pu$



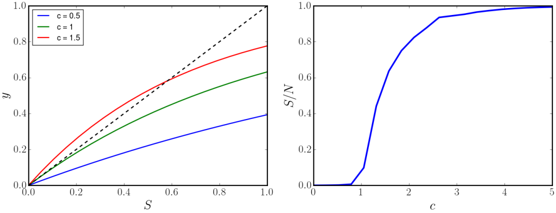
Thus, the probability that a node is not connected via any other $n-1$ nodes to the giant component is:$(1-p+pu)^{n-1}$. Therefore,

$$u = (1-p+pu)^{n-1} = \left(1 + \frac{c}{n-1}(1-u)\right)^{n-1} \tag{19}$$

$$ln(u) = (n-1)ln\left(1 - \frac{c}{n-1}(1-u)\right) \approx -(n-1)\frac{c}{n-1}(1-u) = -c(1-u) \tag{20}$$

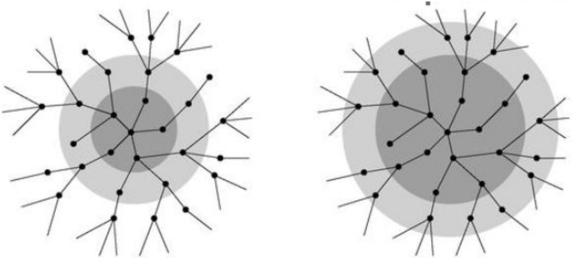
$$u = e^{-c(1-u)} \tag{21}$$

Fraction of vertices in the giant component: $S = 1 - u = 1 - e^{-cS}$




For small $c$, there is only one solution at $S = 0 \rightarrow$ no giant component.
For large enough $c$, there are two solutions: one at $S = 0$ and one at $S > 0$.
The transition between the two regimes corresponds to the point where the gradients of the curve for $c = 1$ and $y = S$ agree at $S = 0$, i.e. when $\frac{d}{dS}(1 - e^{-cS}) = 1 \Longrightarrow c - e^{-cS} = 1$.

Thus the transition occurs when $c = 1$.



We proved that below $c = 1$ no giant component can arise, but not that there has to be a giant component for $c > 1$.
Select $s$ connected nodes. Divide into *core* (nodes linked only to nodes within the selected set - dark grey) and *periphery* (the rest - light grey).
Periphery grows by factor $p(n-s) \approx c$ in the large $n$ limit, if we enlarge core and iterate [3].

- **Diameter.**

$$D \approx \frac{ln(n)}{ln(c)} \tag{22}$$

- **Clustering Coefficient.** Probability that two network neighbours of a vertex are also neighbours of each other.

$$CC = p = \frac{c}{n-1} \tag{23}$$

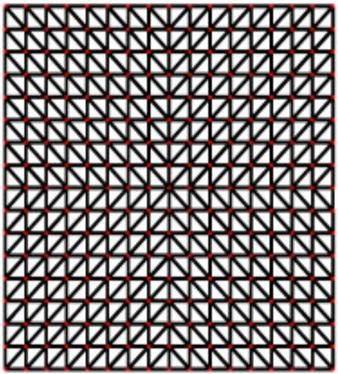
## Random Networks

low path length — high clustering coefficient

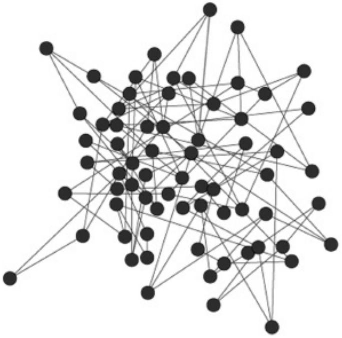
| Network | Size | Separation | Clustering |
|---|---|---|---|
| Company directors | 7673 | 4.60 | 0.588 |
| Movie actors | 449,913 | 3.48 | 0.199 |
| Physics authors | 52,909 | 5.92 | 0.452 |
| Biomedical authors | 1,520,253 | 4.42 | 0.088 |

Figure 2: Real world networks exhibit relatively **small characteristic path lengths** and **large clustering coefficients**.

# Small World Networks.



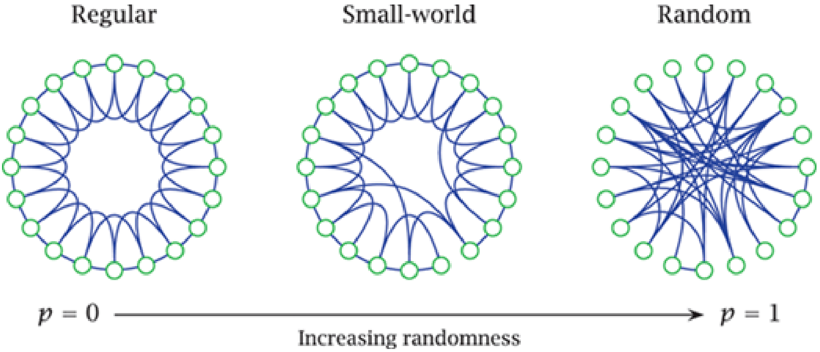
**Large clustering coefficient**
**Large characteristic path length**



**Small clustering coefficient**
**Small characteristic path length**

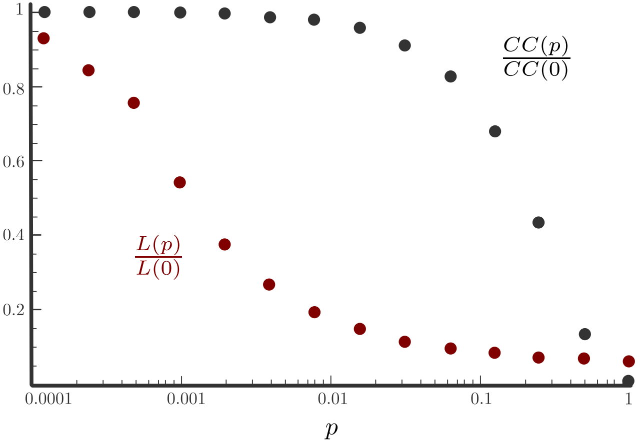
## Watts-Strogatz Model.

Moving from a regular, locally connected graph to a random, globally connected graph [1].

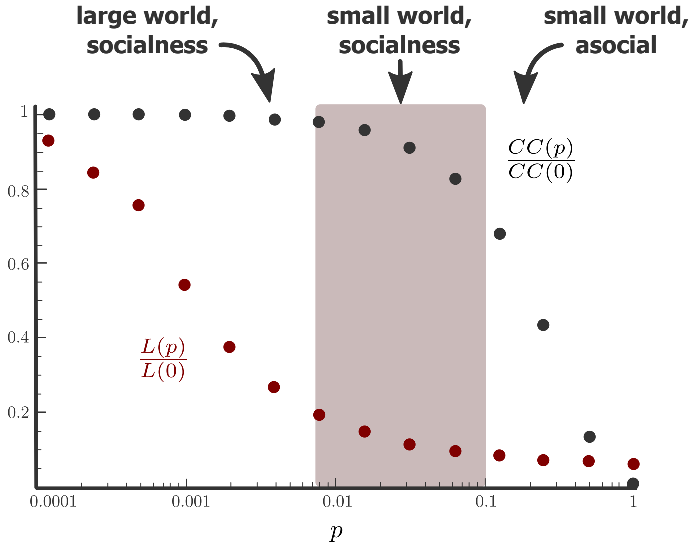


**Exercise.**

- Create a ring-shaped network made of $n = 1000$ nodes.
- Connect each node to $k = 10$ nearest neighbors.
- Randomly rewire edges one-by-one with probability $p$.
- Monitor what happens to the characteristic path length and the average clustering coefficient with increasing rewiring probability (for the same number of nodes). In particular, observe how the quantities $CC(p)/CC(0)$ and $L(p)/L(0)$ change with $p$ averaging over several realisations of each network setting.

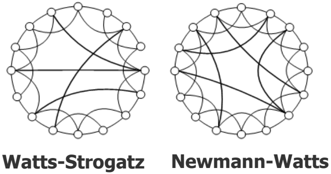


1
0.8
0.6
0.4
0.2
0.0001
0.001
0.01
0.1
1
$p$
$\frac{CC(p)}{CC(0)}$
$\frac{L(p)}{L(0)}$


large world, socialness
small world, socialness
small world, asocial
1
0.8
0.6
0.4
0.2
$\frac{CC(p)}{CC(0)}$
$\frac{L(p)}{L(0)}$
0.0001
0.001
0.01
0.1
1
$p$

Indeed, the existence of a few far leaping links significantly **decreases the length of shortest paths** for most pairs of nodes, while the **clustering coefficient remains relatively high**.

## Newman-Watts Model.




- A variant of the Watts-Strogatz random network, where instead of rewiring edges, for every existing edge in the regular lattice we add randomly an additional edge with probability $p$ [4].
- Thus on average we add $\frac{1}{2}ncp$ shortcuts.

- **Degree distribution.**
  - The additional shortcuts induce an increment of $ncp$ in the total degree of the network and therefore $cp$ shortcuts on average end up at any particular node.
  - The specific number of shortcuts attached to any vertex is Poisson distributed with mean $cp$:

$$P_s = e^{-cp}\frac{(cp)^s}{s!}. \tag{24}$$

  - The total degree of a vertex will be $k = s + c \rightarrow s = k - c$.

$$P_k = e^{-cp}\frac{(cp)^{k-c}}{(k-c)!}, \text{ for } k \leq c. \tag{25}$$

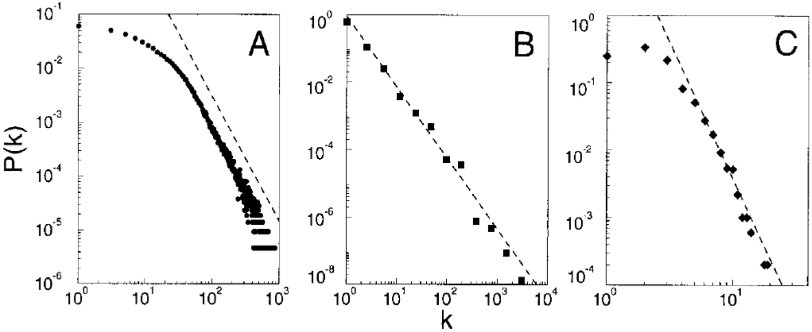
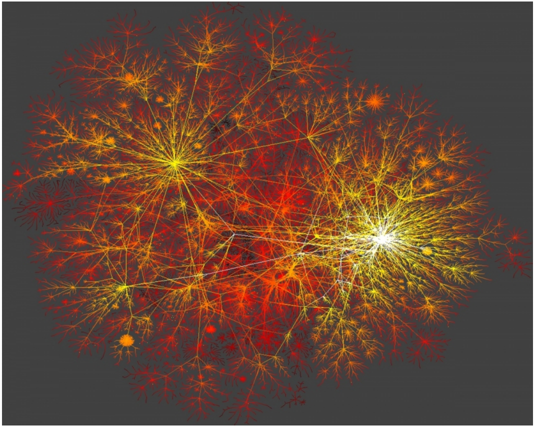



**Fig. 1.** The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

## Scale-free or Growing Random Networks.



- Previous network models assumed **static** topologies, in which edges among a fixed number of nodes are formed randomly following some principle.
- However, most real-world networks do not emerge instantaneously, but rather form **dynamically**. New nodes are born over time and project links to already existing ones.
  - *Example.* The creation of network of webpages (WWW). When a new webpage is created, it most likely includes links to existing web pages.
  - In a similar manner, professional or technological networks are formed.
- Moreover, in the previous models the nodes upon which an edge was attached were selected randomly.

- However, in realistic networks it is more likely that new nodes **preferentially attach** to highly connected (popular) ones.
  - *Example.* A new webpage will probably link to some highly connected page like Google.
  - Indeed, in studies over many different Web snapshots taken at different timepoints, the degree distribution of the network closely resembled a **power law distribution**, i.e. the fraction of webpages with $k$ links $\propto k^{-\gamma}$.
- Evolution over time introduces a natural heterogenicity to nodes according to their áge in the network.

## Barabasi-Albert Model.

**Growth and Preferential Attachment [5]**

- Start with $m_0$ disconnected nodes.
- **Growth:** A new node appears at each timestep $t = 1, 2, ...$
- **Preferential Attachment:** Each node projects $m(\leq m_0)$ links to already present nodes with probability of connecting to $i$-th node $\propto k_i$.

Essentially we have a *rich-gets-richer* scenario; some nodes will gradually turn into hubs.

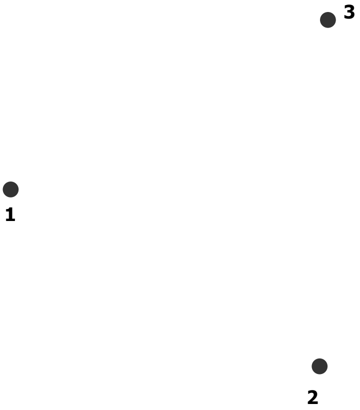
**Definition.**

Attachment probability to a node $i$ with degree $k_i$ at timestep $t$: $P_{attach}(i,t)$.

$$P_{attach}(i,t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}, \tag{26}$$

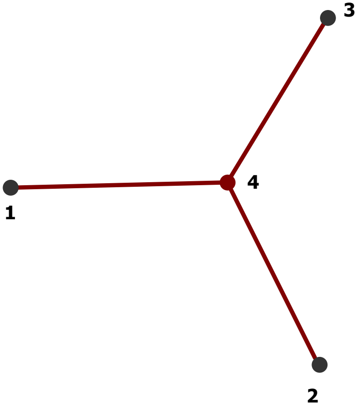
where $N(t)$ is the total number of nodes in the network at timestep $t$.

$m = m_0 = 3$



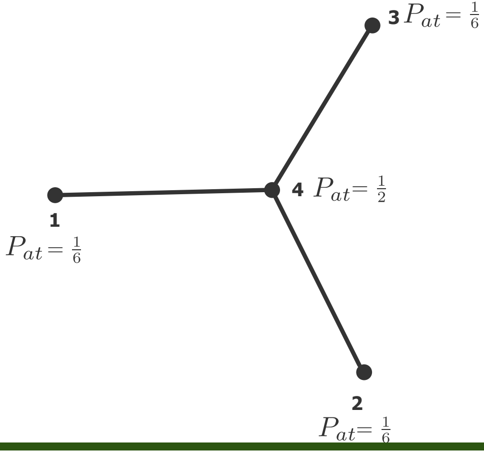
$m = m_0 = 3$

$t = 1$



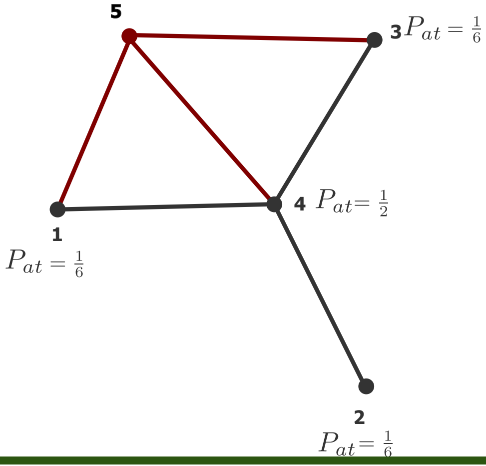
$m = m_0 = 3$

$t = 1$



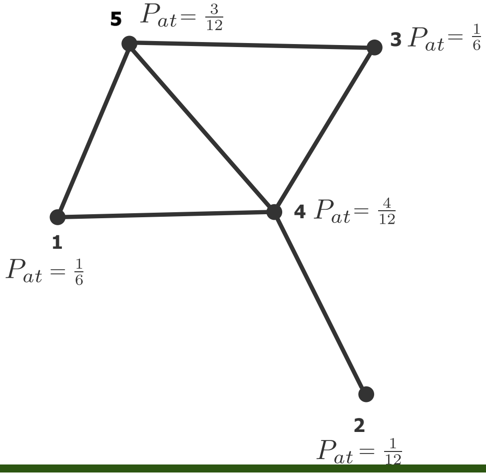
$m = m_0 = 3$

$t = 2$



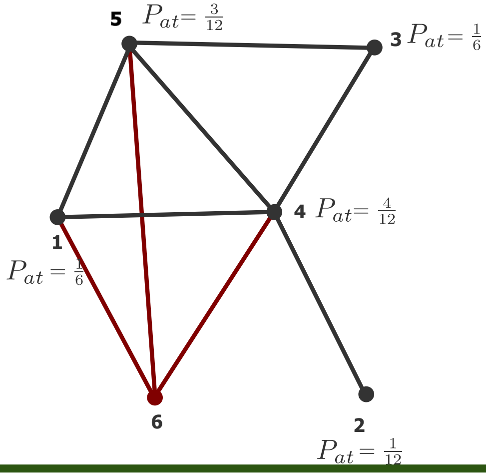
$m = m_0 = 3$

$t = 2$


5 $P_{at} = \frac{3}{12}$
3 $P_{at} = \frac{1}{6}$
4 $P_{at} = \frac{4}{12}$
1
$P_{at} = \frac{1}{6}$
2
$P_{at} = \frac{1}{12}$

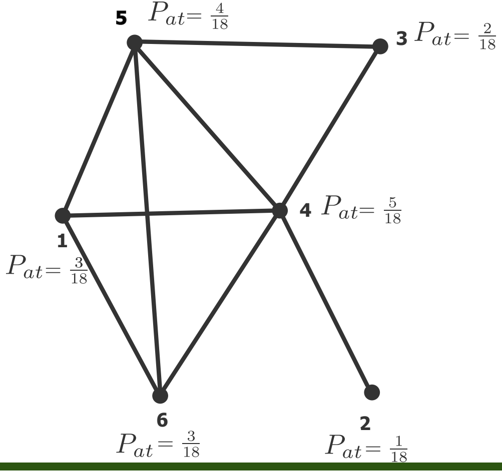
$m = m_0 = 3$

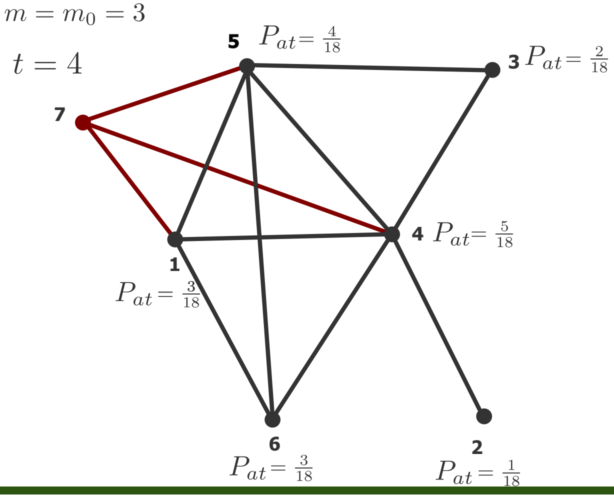
$t = 3$


5 $P_{at} = \frac{3}{12}$
3 $P_{at} = \frac{1}{6}$
4 $P_{at} = \frac{4}{12}$
1
$P_{at} = \frac{1}{6}$
6
2
$P_{at} = \frac{1}{12}$

$m = m_0 = 3$

$t = 3$


5 $P_{at}= \frac{4}{18}$
3 $P_{at}= \frac{2}{18}$
4 $P_{at}= \frac{5}{18}$
1 $P_{at}= \frac{3}{18}$
6 $P_{at}= \frac{3}{18}$
2 $P_{at}= \frac{1}{18}$


$m = m_0 = 3$
$t = 4$
5 $P_{at} = \frac{4}{18}$
3 $P_{at} = \frac{2}{18}$
7
4 $P_{at} = \frac{5}{18}$
1 $P_{at} = \frac{3}{18}$
6 $P_{at} = \frac{3}{18}$
2 $P_{at} = \frac{1}{18}$

## Degree Dynamics.

Approximate the degree with a continuous real variable.
Rate of increase of the degree of node $i$ is:

$$\frac{\partial k_i}{\partial t} = m\frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m\frac{k_i(t)}{2mt} = \frac{k_i(t)}{2t} \tag{27}$$
$$\Longrightarrow k_i(t) = c_i t^{1/2}$$

Denoting with $t_i$ the timetstep of creation of node $i(\leq m_0)$, we have:
$k_i(t_i) = m = c_i t_i^{1/2} \Longrightarrow c_i = \frac{m}{t_i^{1/2}}$ and thus

$$k_i(t) = m(\frac{t}{t_i})^{1/2}. \tag{28}$$

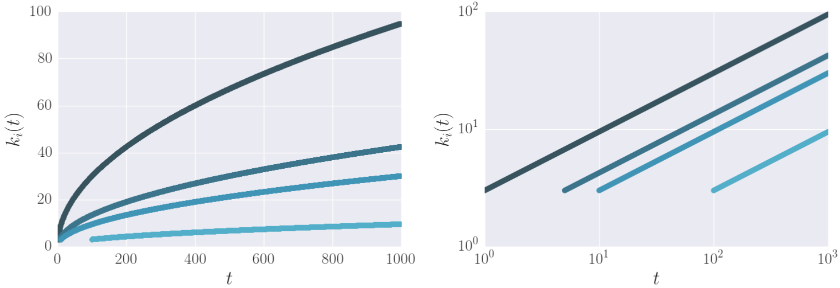



Figure 3: Temporal evolution of degree $k_i$ of four nodes created at timesteps $t = \{1, 5, 10, 100\}$ and $m = 3$. All degrees follow a sublinear power-law with dynamical exponent $1/2$.

## Degree distribution.

We will first calculate the number of nodes with degree smaller than $k$, i.e $k_i(t) < k$. From equation (28), we may write

$$t_i > t_* = \frac{m^2 t}{k^2}, \tag{29}$$

where $t_*$ the appearance timestep of a node with degree $k$ at timestep $t$. All latter nodes will have degree smaller than $k$ at $t$, according to the previously derived degree dynamics.

Since at each timestep one node is added, the number of nodes with degree smaller than $k$ is:

$$t - \frac{m^2 t}{k^2} \tag{30}$$

Altogether there are $N = m_0 + t$ nodes, which in the large $t$ limit becomes $N \approx t$. Therefore, the probability that a randomly selected node has degree $k$ or smaller follows:

$$P(k) = 1 - \frac{m^2}{k^2} \tag{31}$$

Finally, we obtain the degree distribution $P_k$ by taking the derivative of the cumulative degree distribution $P(k)$:

$$P_k = \frac{\partial P(k)}{\partial k} = \frac{2m^2}{k^{2+1}} \propto 2m^2 k^{-3}, \text{ for } m_0 \ll t. \tag{32}$$

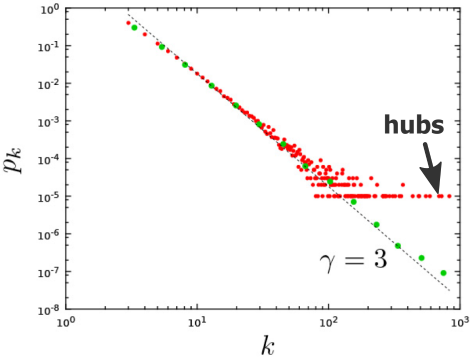
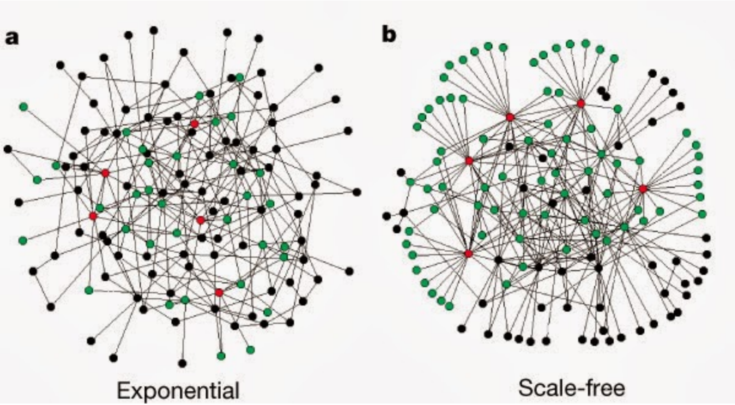



Figure 4: Degree distribution $P_k$ of a scale-free network. [Red:$P_k$ vs $k$, green: $P_k$ vs $k$ after binning in equidistant $k$s in the log scale, grey - actual power law with exponent $\gamma = 3$].

**Remarks.**

- The obtained degree distribution is independent of both the time $t$ and the size $N$ of the network.
- The model generates a scale-free network with exponent $\gamma = 3$. The exponent is independent of the parameters $m_0$ and $m$.
- The **clustering coefficient** scales with $CC_l \approx \frac{(lnN)^2}{N}$ [6].

**Robustness.**




[Figure from [7]]

## Homework.

Error tolerance of Scale-free and Erdős-Rényi random networks.
To study how the network interconnectedness changes upon network failures, generate random Erdos-Renyi and Scale-free graphs of $N = 1000$ nodes and $m = 2000$ links. For each network simulate two different attack strategies:

- Progressively remove randomly selected network nodes. (Failure)
- Perform informed removals by deleting nodes in order of decreasing degree. (Attack)

Observe the diameter and the average path length as the fraction $f$ of removed network changes for both network models and both attack schemes. Which network is more resilient to each attack scheme?

**Exercise.**

Measure degree correlation (assortativity) for the following networks:

- Erdős-Rényi random networks
- Watts-Strogatz small-world networks
- Barabasi-Albert scale-free networks Repeat measurements multiple times and plot histograms of assortativity.

## Homework.

1. Consider a random graph $G(n, p)$ with mean degree $c$.

   a. Show that in the limit of large $n$ the expected number of triangles in the network is $\frac{1}{6}c^3$. This means that the number of triangles is constant, neither growing nor vanishing in the limit of large $n$.

   b. Show that the expected number of connected triples in the network is $\frac{1}{2}nc^2$.

   c. Calculate the global clustering coefficient and confirm that in the limit of large $n$ it agrees with $CC = \frac{c}{n-1}$.

# Overview.

- **Erdős-Rényi $G(n,p)$:** Fixed number of nodes $n$, each pair connected with probability $p$.
  - Characteristic path length scales with $log(n)$.
  - Clustering coefficient vanishes in the large $n$ limit (for fixed $c$).
- **Small-world $SW(n,k,p)$:** Regular $n$ node lattice with node degree $k$ and probability $p$ for rewiring each edge.
  - Characteristic path length scales with $log(n)$.
  - Large Clustering coefficients.

- **Scale-Free $SF(m_0, m, t)$:** Starting initially from $m_0$ isolated nodes, network grows by progressive addition of one node of degree $m$ at each timestep $t$. Each new node connects preferentially to $m$ existing nodes with prob. of connection proportional to the degree of each node.
    - Power law degree distribution

# References

[1] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[2] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[3] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[4] Mark EJ Newman and Duncan J Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.

[5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[6] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.

[7] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

# Questions?