Modelling and Analysis of Complex Networks

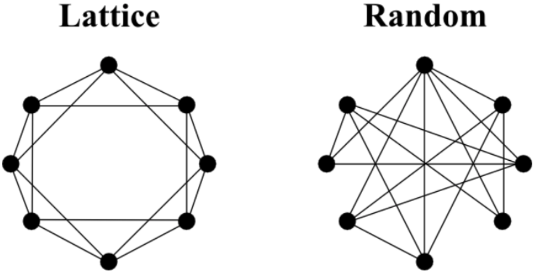
— Semester 3, Master of Data Science —

Jun Pang
University of Luxembourg

The Small-Word Model

The Small-Word Model

Can we have high clustering while also having short paths?



- Regular graphs: high clustering coefficient, high diameter
- Random graphs: low clustering coefficient, low diameter

The Small-Word Experiment [Milgram'67]

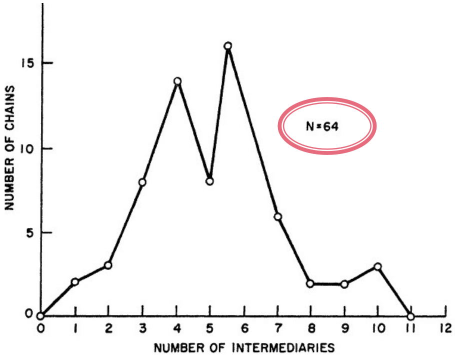
What is the typical shortest path length between any two people?

- Picked 300 people in Omaha, Nebraska and Wichita, Kansas
- Ask them to get a letter to a stock-broker in Boston by passing it through friends
- How many steps did it take?

The Small-Word Experiment [Milgram'67]

64 chains completed i.e., 64 letters reached the target

- It took 6.2 steps on average, thus "6 degrees of separation"



- People who owned stock had shorter paths (5.4 vs. 6.7)
- People from the Boston area have even shorter paths (4.4)

Should We Be Surprised?

Assume each person is connected to 100 other persons

- Step 1: reach 100 persons
- Step 2: reach 100*100 = 10,000 persons
- Step 3: reach 100*100*100 = 1,000,000 persons
- Step 4: reach 100*100*100*100 = 100M persons
- Step 5: reach 100*100*100*100*100 = 10B persons

What is wrong here? We ignore clustering!

- Not all edges point to new persons
  (92% of Facebook friendships happen through a friend-of-a-friend)

Should We Be Surprised?

Assume each person is connected to 100 other persons

- Step 1: reach 100 persons
- Step 2: reach 100*100 = 10,000 persons
- Step 3: reach 100*100*100 = 1,000,000 persons
- Step 4: reach 100*100*100*100 = 100M persons
- Step 5: reach 100*100*100*100*100 = 10B persons

What is wrong here? We ignore clustering!

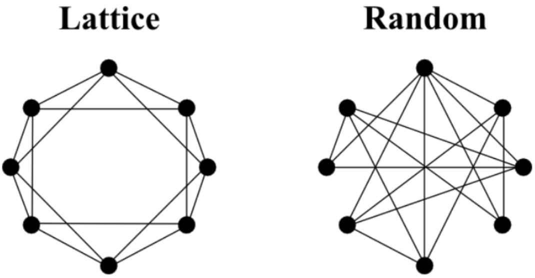
- Not all edges point to new persons
  (92% of Facebook friendships happen through a friend-of-a-friend)

Clustering Implies Edge Locality

- MSN network has 7 orders of magnitude larger clustering than the corresponding $G(N, p)$.
- More examples:

Network	h_{actual}	h_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power Grid	18.70	12.40	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

The Small-Word Model



- Regular graphs: high clustering coefficient, high diameter
- Random graphs: low clustering coefficient, low diameter
- How can we have both?

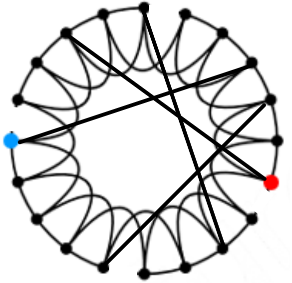
The Watts-Strogatz Model

A model to capture large clustering coefficient and short distances observed in real networks

- It interpolated between an ordered lattice and a random graph
- Fixed parameters: N (network size), $\overline{k}$ initial coordination number (node degree)
- Variable parameters: p (rewiring probability)

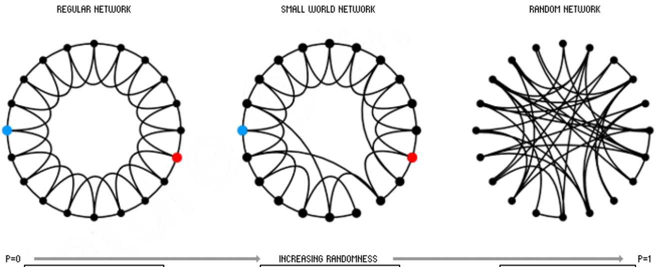
The Watts-Strogatz Model

- Start with a ring lattice with N nodes, in which every node is connected to its first $\overline{k}$ neighbours ($\frac{\overline{k}}{2}$ on either side)
- Randomly rewire each edge of the lattice with probability p (self-connections and duplicate edges are excluded)



The Watts-Strogatz Model

By varying p the network can be transformed from a completely ordered ($p = 0$) to a completely random ($p = 1$) structure.



High clustering High diameter	High clustering Low diameter	Low clustering Low diameter
$h = \frac{N}{2\bar{k}}$ $\quad C = \frac{1}{2}$		$h = \frac{\log N}{\log \alpha}$ $\quad C = \frac{\bar{k}}{N}$

The Watts-Strogatz Model

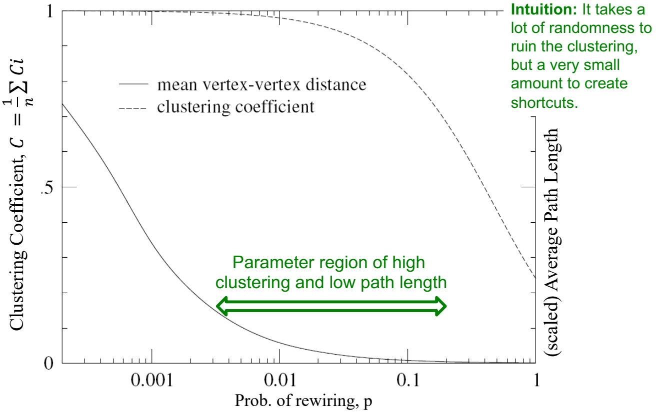
Clustering coefficient:

$$\overline{C} = \frac{\frac{1}{4}N\overline{k}(\frac{1}{2}\overline{k} - 1) \times 3}{\frac{1}{2}N\overline{k}(\overline{k} - 1) + N\overline{k}^2 p + \frac{1}{2}N\overline{k}^2 p^2} = \frac{3(\overline{k} - 2)}{4(\overline{k} - 1) + 8\overline{k}p + 4\overline{k}p^2}$$

- Independent of N
- It recovers the ring value if $p \to 0$

The Watts-Strogatz Model

Average path length: No closed solution (from numerical simulations)



The Watts-Strogatz Model

Degree distribution:

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

if $k \geq \overline{k}$, and $P(k) = 0$ if $k < \overline{k}$.

- $p = 0$: each node has the same degree $\overline{k}$
- $p > 0$: approximates a Poisson distribution (random graphs)

The Watts-Strogatz Model

Degree distribution:

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

if $k \geq \overline{k}$, and $P(k) = 0$ if $k < \overline{k}$.

- $p = 0$: each node has the same degree $\overline{k}$
- $p > 0$: approximates a Poisson distribution (random graphs)

The Watts-Strogatz Model

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small
Configuration Model	Custom, can be broad	short	small
Watts & Strogatz (in SW regime)	Poissonian	short	large

The Watts-Strogatz Model

- Gives rise to small-world networks with high clustering
- Provides insight between clustering and the small world
- Captures the structure of many realistic networks
- Does not lead to the correct degree distribution

The Watts-Strogatz Model

- Gives rise to small-world networks with high clustering
- Provides insight between clustering and the small world
- Captures the structure of many realistic networks
- Does not lead to the correct degree distribution

Scale-free Networks

Scale-free Networks

A network is called scale-free when its degree distribution follows (to some extent) a Power-law distribution.

$$P(k) \sim Ck^{-\gamma} = C\frac{1}{k^{\gamma}}$$

Power-law distribution (PDF), γ called the exponent of the distribution

Scale-fee refers specifically to the degree distribution having a Power-law decay in its tail, namely for 'large' k.

Scale-free Networks

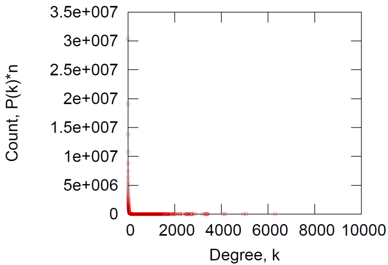
A network is called scale-free when its degree distribution follows (to some extent) a Power-law distribution.

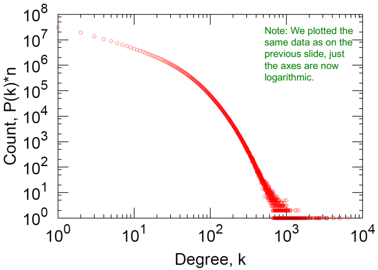
$$P(k) \sim Ck^{-\gamma} = C\frac{1}{k^{\gamma}}$$

Power-law distribution (PDF), γ called the exponent of the distribution

Scale-fee refers specifically to the degree distribution having a Power-law decay in its tail, namely for 'large' k.

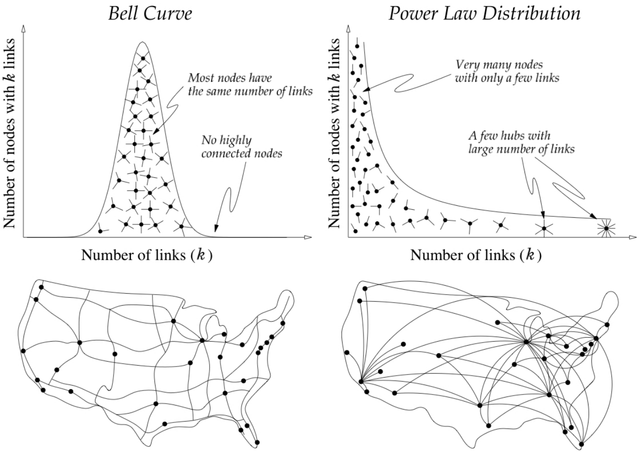
MSN: Degree Distribution





Linear ($P(k) \sim k^{-\gamma}$) vs. log-log degree distribution ($\log P(k) \sim -\gamma \log(k)$)

Scale-free Distribution



AL. Barabási, Linked (2002)

Degree fluctuations have no characteristic scale (scale-invariant).

Scale-free Distribution (Discrete)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

Log-log coordinates:

$$\log P(k) = -\gamma \log k + \log C$$

Normalisation:

$$\sum_{k=1}^{\infty} P(k) = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(r) = 1; C = \frac{1}{\zeta(\gamma)}$$

Riemann zeta function, $\gamma > 1$:

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Scale-free Distribution (Discrete)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

Log-log coordinates:

$$\log P(k) = -\gamma \log k + \log C$$

Normalisation:

$$\sum_{k=1}^{\infty} P(k) = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(r) = 1; C = \frac{1}{\zeta(\gamma)}$$

Riemann zeta function, $\gamma > 1$:

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Scale-free Distribution (Discrete)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

Log-log coordinates:

$$\log P(k) = -\gamma \log k + \log C$$

Normalisation:

$$\sum_{k=1}^{\infty} P(k) = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(r) = 1; C = \frac{1}{\zeta(\gamma)}$$

Riemann zeta function, $\gamma > 1$:

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Scale-free Distribution (Discrete)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

Log-log coordinates:

$$\log P(k) = -\gamma \log k + \log C$$

Normalisation:

$$\sum_{k=1}^{\infty} P(k) = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(r) = 1; C = \frac{1}{\zeta(\gamma)}$$

Riemann zeta function, $\gamma > 1$:

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Scale-free Distribution (Continuous Approximation)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

To have a proper degree distribution ($k \in \mathbb{R}$), we need

$$\int P(k) = 1 = \int Ck^{-\gamma} = C \int k^{-\gamma}$$

In most cases, there is a lower bound k_{min} from which the law holds. From this, we define the normalisation constant:

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{min}^{\gamma - 1}$$

Scale-free Distribution (Continuous Approximation)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

To have a proper degree distribution ($k \in \mathbb{R}$), we need

$$\int P(k) = 1 = \int Ck^{-\gamma} = C \int k^{-\gamma}$$

In most cases, there is a lower bound k_{min} from which the law holds. From this, we define the normalisation constant:

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{min}^{\gamma - 1}$$

Scale-free Distribution (Continuous Approximation)

Initial definition ($k \in \mathbb{N}$):

$$P(k) \sim Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

To have a proper degree distribution ($k \in \mathbb{R}$), we need

$$\int P(k) = 1 = \int Ck^{-\gamma} = C \int k^{-\gamma}$$

In most cases, there is a lower bound k_{min} from which the law holds. From this, we define the normalisation constant:

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{min}^{\gamma - 1}$$

Scale-free Distribution (Continuous Approximation)

Put together:

$$P(k) \sim Ck^{-\gamma}$$

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{min}^{\gamma-1}$$

Hence, power law normalised probability density function (PDF)

$$P(k) = (\gamma - 1) k_{min}^{\gamma-1} k^{-\gamma}$$

$$\boxed{P(k) = \frac{\gamma - 1}{k_{min}} \left(\frac{k}{k_{min}} \right)^{-\gamma}}$$

Scale-free Distribution (Continuous Approximation)

Put together:

$$P(k) \sim Ck^{-\gamma}$$

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{min}^{\gamma-1}$$

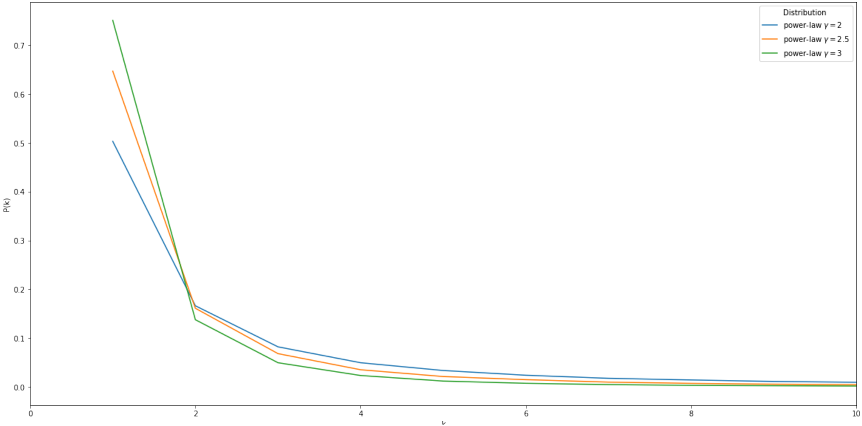
Hence, power law normalised probability density function (PDF)

$$P(k) = (\gamma - 1)k_{min}^{\gamma-1} k^{-\gamma}$$

$$\boxed{P(k) = \frac{\gamma - 1}{k_{min}} \left(\frac{k}{k_{min}}\right)^{-\gamma}}$$

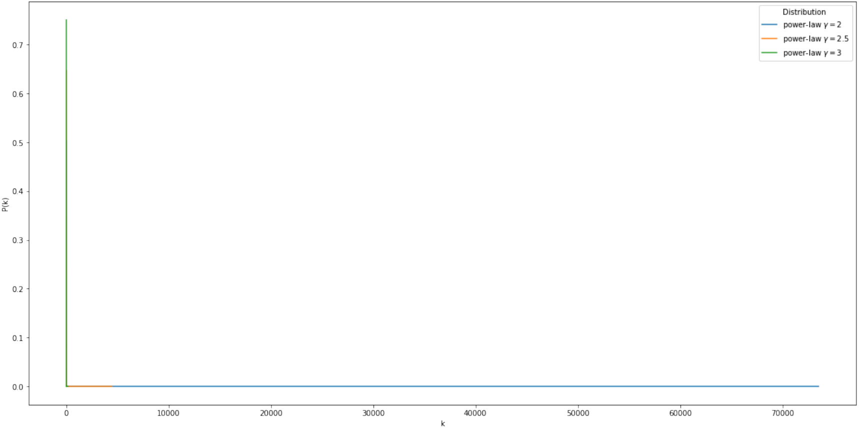
Scale-free Networks

Power-law plotted with a linear scale, for $k \leq 10$ (100,000 samples)



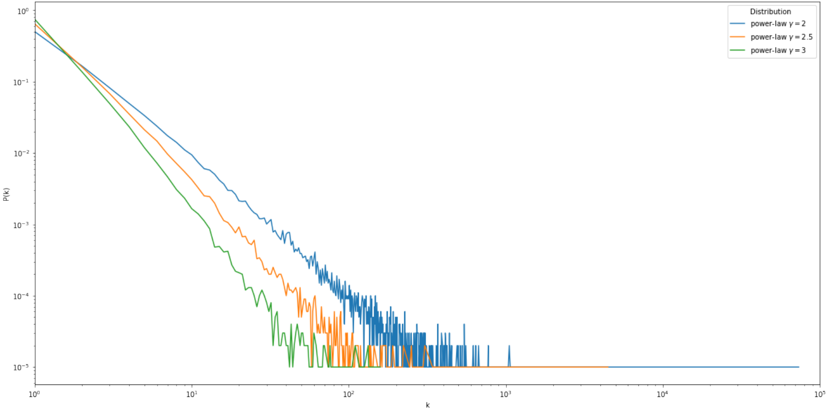
Scale-free Networks

Power-law plotted with a linear scale, for $k < 100,000$



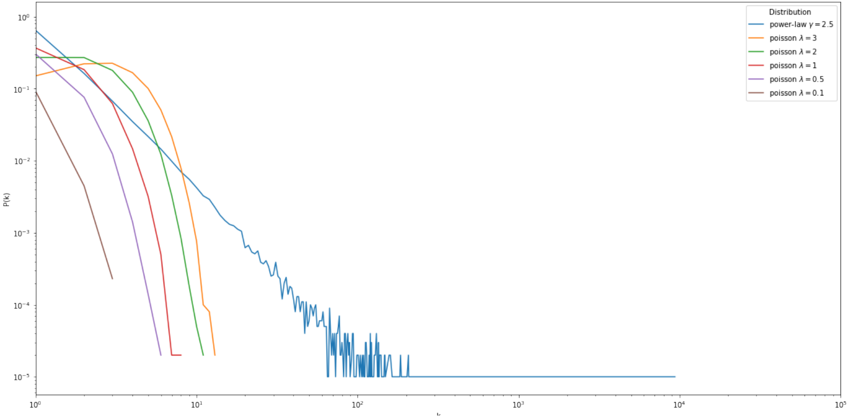
Scale-free Networks

Power-law plotted with a log-log scale, for $k < 100,000$

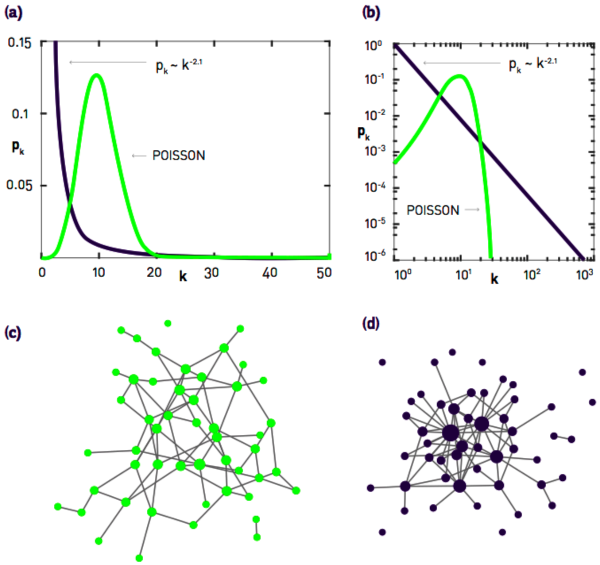


Scale-free Networks

Compare a Poisson distribution and a Power-law



Scale-free Networks



Power law vs. Poisson distribution

Scale-free Networks

How does the network size affect the size of its hubs?

- Probability of network having a node with degree $k \geq k_{max}$:

$$P(k \geq k_{max}) = \int_{k_{max}}^{\infty} P(k)dk$$

- Expected number of nodes with $k \geq k_{max}$:

$$N \cdot P(k \geq k_{max}) = 1$$

- Expected largest node degree in power-law networks:

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$$

- Expected largest node degree in random networks:

$$k_{max} = k_{min} + \frac{\ln N}{\gamma}$$

Scale-free Networks

How does the network size affect the size of its hubs?

- Probability of network having a node with degree $k \geq k_{max}$:

$$P(k \geq k_{max}) = \int_{k_{max}}^{\infty} P(k)dk$$

- Expected number of nodes with $k \geq k_{max}$:

$$N \cdot P(k \geq k_{max}) = 1$$

- Expected largest node degree in power-law networks:

$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}$$

- Expected largest node degree in random networks:

$$k_{max} = k_{min} + \frac{\ln N}{\gamma}$$

Scale-free Networks

How does the network size affect the size of its hubs?

- Probability of network having a node with degree $k \geq k_{max}$:

$$P(k \geq k_{max}) = \int_{k_{max}}^{\infty} P(k) dk$$

- Expected number of nodes with $k \geq k_{max}$:

$$N \cdot P(k \geq k_{max}) = 1$$

- Expected largest node degree in power-law networks:

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$$

- Expected largest node degree in random networks:

$$k_{max} = k_{min} + \frac{\ln N}{\gamma}$$

Scale-free Networks

How does the network size affect the size of its hubs?

- Probability of network having a node with degree $k \geq k_{max}$:

$$P(k \geq k_{max}) = \int_{k_{max}}^{\infty} P(k)dk$$

- Expected number of nodes with $k \geq k_{max}$:

$$N \cdot P(k \geq k_{max}) = 1$$

- Expected largest node degree in power-law networks:

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$$

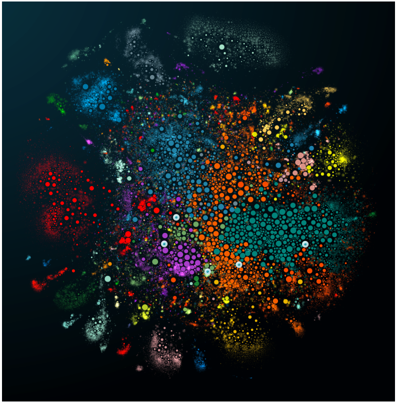
- Expected largest node degree in random networks:

$$k_{max} = k_{min} + \frac{\ln N}{\gamma}$$

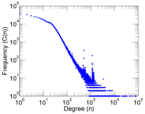
Scale-free Networks

Online social networks

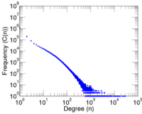
- Nodes: individuals
- Links: online interactions



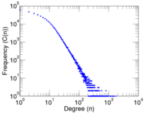
Social network of Steam
http://85.25.226.110/mapper



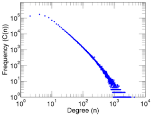
Catster/Dogster Familylinks/Friendships



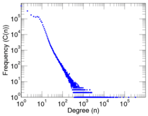
Chinese Wikipedia internal links



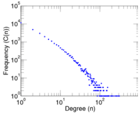
CiteSeer



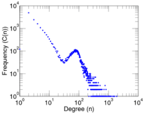
DBLP



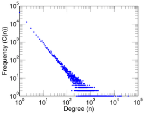
DBpedia



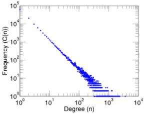
Digg



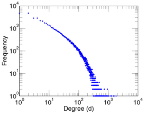
Edinburgh Associative Thesaurus



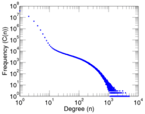
Enron



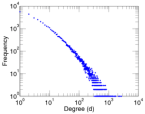
Epinions trust



Facebook friendships



Facebook social graph



Facebook wall posts

Scale-free Networks

Computing γ of an observed network:

- Naïve method: find the slope of the line of the log-log plot (overfitting based on a few values in the long tail)
- More advanced method: Maximum Likelihood Estimation

Fitting to the Power-Law Distribution, Goldstein et al.

https://arxiv.org/vc/cond-mat/papers/0402/0402322v1.pdf

Scale-free Networks

Computing γ of an observed network:

- Naïve method: find the slope of the line of the log-log plot (overfitting based on a few values in the long tail)
- More advanced method: Maximum Likelihood Estimation

Fitting to the Power-Law Distribution, Goldstein et al.

https://arxiv.org/vc/cond-mat/papers/0402/0402322v1.pdf

Scale-free Networks

Network	Size	$\langle k \rangle$	κ	γ_{out} (Exponent)	γ_{in} (Exponent)	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

Albert, R. et.al. Rev. Mod. Phy. (2002)

Exponents γ of real-world networks are usually between 2 and 3.

Scale-free Networks

Network	Size	$\langle k \rangle$	κ	Exponent γ_{out}	Exponent γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

Albert, R. et.al. Rev. Mod. Phy. (2002)

Exponents γ of real-world networks are usually between 2 and 3.

Scale-free Networks

Why $2 \leq \gamma \leq 3$?

- $\gamma < 2$: the distribution is so skewed that we expect to find nodes with a degree larger than the size of the network (not possible in finite networks).
- To detect a scale-free network its degree distribution needs to span through several (at least 2-3) orders of magnitude
- $\gamma > 3$: large degrees become so rare that the size of the sample (i.e., size of observed network) must be enormous to indeed observe such a node, for example

  $\gamma = 4, k_{min} = 1, k_{max} = 1,000, \; N = (\frac{k_{max}}{k_{min}})^{\gamma-1} = 10^{12}$

Scale-free Networks

Why $2 \leq \gamma \leq 3$?

- $\gamma < 2$: the distribution is so skewed that we expect to find nodes with a degree larger than the size of the network (not possible in finite networks).
- To detect a scale-free network its degree distribution needs to span through several (at least 2-3) orders of magnitude
- $\gamma > 3$: large degrees become so rare that the size of the sample (i.e., size of observed network) must be enormous to indeed observe such a node, for example

  $\gamma = 4, k_{min} = 1, k_{max} = 1,000, \; N = (\frac{k_{max}}{k_{min}})^{\gamma-1} = 10^{12}$

Scale-free Networks

Why $2 \leq \gamma \leq 3$?

- $\gamma < 2$: the distribution is so skewed that we expect to find nodes with a degree larger than the size of the network (not possible in finite networks).
- To detect a scale-free network its degree distribution needs to span through several (at least 2-3) orders of magnitude
- $\gamma > 3$: large degrees become so rare that the size of the sample (i.e., size of observed network) must be enormous to indeed observe such a node, for example

  $\gamma = 4, k_{min} = 1, k_{max} = 1,000,\ N = (\frac{k_{max}}{k_{min}})^{\gamma-1} = 10^{12}$

Scale-free Networks

Degree of a randomly chosen node:

$$k = \langle k \rangle \pm \sigma_k, \qquad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Poisson degree distribution (random networks) has a scale $\langle k \rangle$:

$$k = \langle k \rangle \pm \sqrt{\langle k \rangle}$$

Power-law network with $2 < \gamma < 3$ is scale-free:

$$k = \langle k \rangle \pm \infty$$

Scale-free Networks

Degree of a randomly chosen node:

$$k = \langle k \rangle \pm \sigma_k, \qquad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Poisson degree distribution (random networks) has a scale $\langle k \rangle$:

$$k = \langle k \rangle \pm \sqrt{\langle k \rangle}$$

Power-law network with $2 < \gamma < 3$ is scale-free:

$$k = \langle k \rangle \pm \infty$$

Scale-free Networks

Degree of a randomly chosen node:

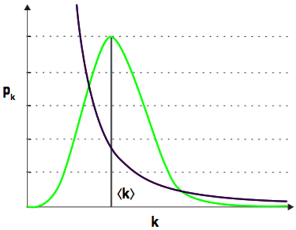
$$k = \langle k \rangle \pm \sigma_k, \qquad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Poisson degree distribution (random networks) has a scale $\langle k \rangle$:

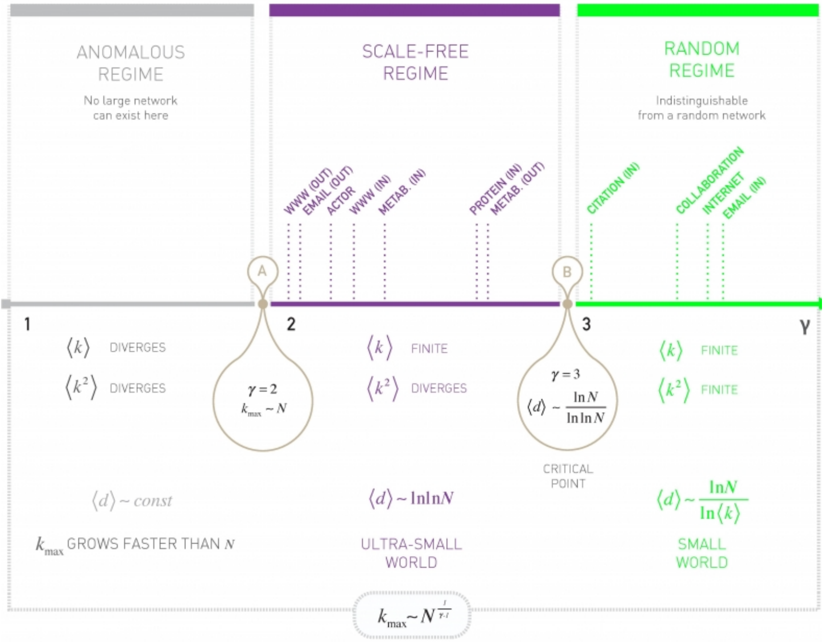
$$k = \langle k \rangle \pm \sqrt{\langle k \rangle}$$

Power-law network with $2 < \gamma < 3$ is scale-free:

$$k = \langle k \rangle \pm \infty$$



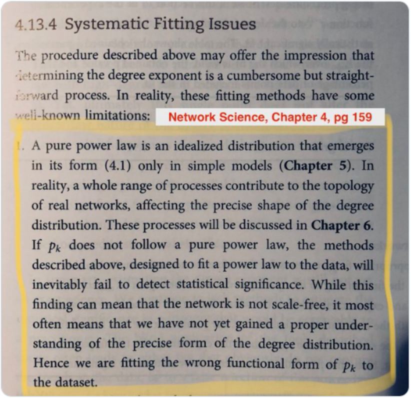
Scale-free Networks



Are Real Networks Really Scale-free?

Albert-László Barabási
@barabasi

@aaronclauset Every 5 years someone is shocked to re-discover that a pure power law does not fit many networks. True: Real networks have predictable deviations. Hence forcing a pure power law on these is like...fitting a sphere to the cow. Sooner or later the hoof will stick out.


4.13.4 Systematic Fitting Issues

The procedure described above may offer the impression that determining the degree exponent is a cumbersome but straightforward process. In reality, these fitting methods have some well-known limitations:

Network Science, Chapter 4, pg 159

1. A pure power law is an idealized distribution that emerges in its form (4.1) only in simple models (**Chapter 5**). In reality, a whole range of processes contribute to the topology of real networks, affecting the precise shape of the degree distribution. These processes will be discussed in **Chapter 6**. If p_k does not follow a pure power law, the methods described above, designed to fit a power law to the data, will inevitably fail to detect statistical significance. While this finding can mean that the network is not scale-free, it most often means that we have not yet gained a proper understanding of the precise form of the degree distribution. Hence we are fitting the wrong functional form of p_k to the dataset.

Albert-László Barabási @barabasi · Jan 15, 2018
Replying to @barabasi
Chapter 6 in Network Science networksciencebook.com/chapter/6 discusses what you should be fitting to the degree distribution of *real* scale-free networks. You are right: Pure power laws are predictably rare. Scale-free networks are not.
1 21 45

Aaron Clauset @aaronclauset · Jan 15, 2018
Replying to @barabasi
Yes, science is hard and real data often messy. But it is worrying how criticisms of harsh statistical evaluations can be interpreted as a belief that "disagreement with data" (as Feynman would put it) should not be held against a favored theory or model.
3 5 18

Albert-László Barabási @barabasi · Jan 15, 2018
We are on the same page. The question is, what you test and what you conclude. There are multiple processes that contribute to the degree distribution that modify the power law. Hence testing for power laws only you are ignoring them all, leading to misleading takeway message.
2 4 10

Aaron Clauset @aaronclauset · Jan 15, 2018
Perhaps. I feel good about the accuracy of our conclusions: we used rigorous statistical methods, tested 5 distributions, considered 5 levels of evidence, across nearly 1000 network datasets. The goal was to be thorough and to treat the SF hypothesis as falsifiable.
1 3 14

Albert-László Barabási @barabasi · Jan 15, 2018
The effort is amazing. The conclusions are less so. The feather falls slower than the rock, yet gravitation is not wrong. We add friction. You need to fit for each system the Pk that is right for it. That is hard, I know. Otherwise you ignore 20 year of work by hundreds.
2 4 6

Aaron Clauset @aaronclauset · Jan 15, 2018
It seems easy to get confused here: an empirical power-law degree distribution is evidence for SF structure, but no deviation from the power law can be evidence against SF structure? It is reasonable to believe a fundamental phenomena would require less customized detective work.

Are Real Networks Really Scale-free?

- Rigorous statistical tests show that observed degree distributions are not compatible with a Power-law distribution.
- Compared with different distributions, in particular log-normal, most degree distributions are likely to be generated by something else than Power-laws.
- Networks are real objects, not mathematical abstraction, therefore they are sensitive to noise (real-life limits, ...)
- Power-law is a good, simple model of degree distributions of a class of networks

Jacomy, M. Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. Big Data & Society, 2020.

Are Real Networks Really Scale-free?

- Rigorous statistical tests show that observed degree distributions are not compatible with a Power-law distribution.
- Compared with different distributions, in particular log-normal, most degree distributions are likely to be generated by something else than Power-laws.

- Networks are real objects, not mathematical abstraction, therefore they are sensitive to noise (real-life limits, ...)
- Power-law is a good, simple model of degree distributions of a class of networks

Jacomy, M. Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. Big Data & Society, 2020.

Are Real Networks Really Scale-free?

- Rigorous statistical tests show that observed degree distributions are not compatible with a Power-law distribution.
- Compared with different distributions, in particular log-normal, most degree distributions are likely to be generated by something else than Power-laws.

- Networks are real objects, not mathematical abstraction, therefore they are sensitive to noise (real-life limits, ...)
- Power-law is a good, simple model of degree distributions of a class of networks

Jacomy, M. Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. Big Data & Society, 2020.

Scale-free Networks

The Big Deal:

- We move beyond describing of networks to finding mechanisms for why certain networks are the way they are.

The Barabási-Albert Model

Emergence of Hubs

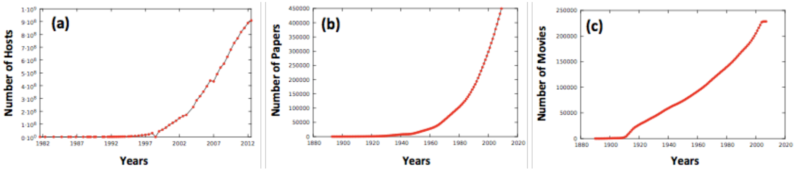
What did we miss with the earlier network models?

- Networks are not static but growing in time, as new nodes are entering the system
- Preferential attachment: nodes are not connected randomly, but tend to link to more attractive nodes.

Emergence of Hubs

What did we miss with the earlier network models?

- Networks are not static but growing in time, as new nodes are entering the system
- Preferential attachment: nodes are not connected randomly, but tend to link to more attractive nodes.



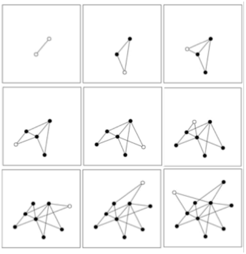
Barabási, Network Science Book (2016).

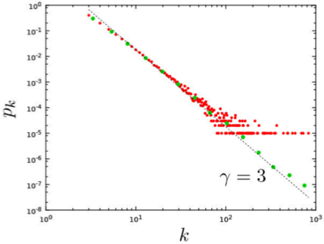
The Barabási-Albert Model

- Start with m_0 connected nodes
- At each time step we add a new node with m ($\leq m_0$) links that connect it to m nodes already in the network
- The probability $\pi(k)$ that one of the links of the new node connects to node i depends on its degree k_i as $\pi(k_i) = \frac{k_i}{\sum_j k_j}$

The Barabási-Albert Model

- Start with m_0 connected nodes
- At each time step we add a new node with m ($\leq m_0$) links that connect it to m nodes already in the network
- The probability $\pi(k)$ that one of the links of the new node connects to node i depends on its degree k_i as $\pi(k_i) = \frac{k_i}{\sum_j k_j}$





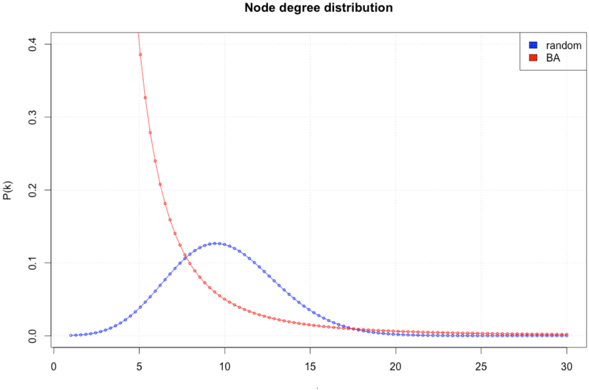
The emerging network will be scale-free with $\gamma = 3$, independent on m_0 and m.

Barabási, Network Science Book (2016).

The Barabási-Albert Model

Degree distribution:

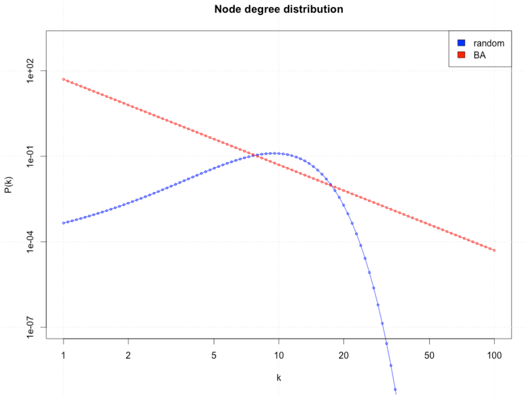
$$P(k) = \frac{2m^2}{k^3}$$



The Barabási-Albert Model

Degree distribution:

$$P(k) = \frac{2m^2}{k^3}$$



The Barabási-Albert Model

Path length:

$$
\langle l \rangle \sim \begin{cases}
const. & \gamma = 2 \\
\dfrac{\ln\ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3 \\
\dfrac{\ln N}{\ln\ln N} & \gamma = 3 \\
\ln N & \gamma > 3
\end{cases}
$$

Ultra Small World

- $\gamma = 2$: Size of the biggest hub is of order O(N). Most nodes can be connected within two layers of it, thus the average path length will be independent of the system size.
- $2 < \gamma < 3$: The average path length increases slower than logarithmically. In a random network all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network the vast majority of the path go through the few high degree hubs, reducing the distances between nodes.
- $\gamma = 3$: Some key models produce γ=3, so the result is of particular importance for them. This was first derived by Bollobas and collaborators for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

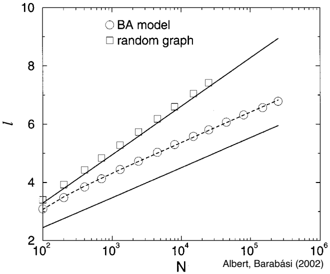
Small World

- $\gamma > 3$: The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

$$\langle l \rangle = \frac{\ln N}{\ln\ln N}$$

Ultra Small World network

Bollobás, Riordan (2001)



Albert, Barabási (2002)

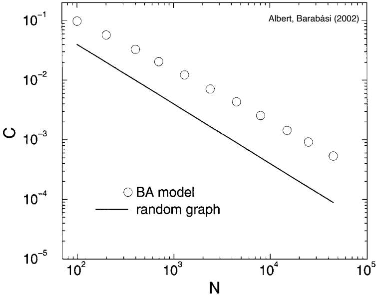
The Barabási-Albert Model

Clustering coefficient:

- The clustering coefficient decreases with the system size as

$$\overline{C} = \frac{(m-1)(\ln N)^2}{8N} \sim \frac{(\ln N)^2}{N}$$

- It is 5 times more than for random graphs (not large enough).



The Barabási-Albert Model

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small
WS small-world networks	exponential	short	large
BA scale-free networks	power-law	short	Rather small

The Barabási-Albert Model

Preferential attachment is a very simple mechanism, but it implies arriving nodes have complete knowledge of the existing network's degree distribution.

- Vary attachment kernel
- Vary mechanisms:
  - Add edge deletion
  - Add node deletion
  - Add edge rewiring
- Deal with directed vs. undirected networks

The Barabási-Albert Model

Preferential attachment is a very simple mechanism, but it implies arriving nodes have complete knowledge of the existing network's degree distribution.

- Vary attachment kernel
- Vary mechanisms:
  - Add edge deletion
  - Add node deletion
  - Add edge rewiring
- Deal with directed vs. undirected networks

The Barabási-Albert Model

Preferential attachment is a very simple mechanism, but it implies arriving nodes have complete knowledge of the existing network's degree distribution.

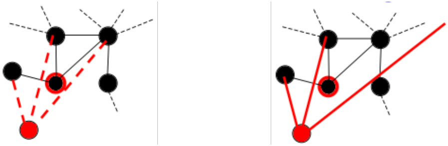
- Vary attachment kernel
- Vary mechanisms:
  - Add edge deletion
  - Add node deletion
  - Add edge rewiring
- Deal with directed vs. undirected networks

Other Models

The Vertex-copying Model

Motivation: citation network or WWW where links are often copies; local explanation to preferential attachment

1. Take a small seed network
2. Pick a random vertex
3. Make a copy of it
4. With probability p, move each edge of the copy to a random vertex
5. Repeat 2-4 until desired network size N



scale-free with $\gamma \geq 3$

The Holme-Kim Model

Motivation: more realistic clustering coefficient

1. Take a small seed network
2. Create a new vertex with m degree
3. Connect the first of the m edges to existing vertices with a probability proportional to their degree k
4. With probability p, connect the next edge to a random neighbour of the vertex of step 3; otherwise do step 3 again
5. Repeat 2-4 until desired network size N

Other Models

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small
WS small-world networks	exponential	short	large
BA scale-free networks	power-law	short	Rather small
Other models	power-law	short	Large

End Notes

- "All models are wrong, but some are useful."
- The Erdös-Rényi and configuration models are used as reference models in a large number of applications
- The Watts-Strogatz and Barabási-Albert models are more "making a point" type models: simple processes can explain non-trivial properties of real-life networks
- Correlation is not causation!

End Notes

- "All models are wrong, but some are useful."
- The Erdös-Rényi and configuration models are used as reference models in a large number of applications
- The Watts-Strogatz and Barabási-Albert models are more "making a point" type models: simple processes can explain non-trivial properties of real-life networks
- Correlation is not causation!

End Notes

- "All models are wrong, but some are useful."
- The Erdös-Rényi and configuration models are used as reference models in a large number of applications
- The Watts-Strogatz and Barabási-Albert models are more "making a point" type models: simple processes can explain non-trivial properties of real-life networks
- Correlation is not causation!

End Notes

- “All models are wrong, but some are useful.”
- The Erdös-Rényi and configuration models are used as reference models in a large number of applications
- The Watts-Strogatz and Barabási-Albert models are more “making a point” type models: simple processes can explain non-trivial properties of real-life networks
- Correlation is not causation!