Structure and dynamics of complex networks

March 26, 2020

THE BARABÁSI-ALBERT MODEL

Random graphs vs real networks

	E-R	W-S
- sparseness, $\langle k \rangle$?	**OK**	**OK**
-the small-world effect, $\langle \ell \rangle$?	**OK**	**OK**
-large local clustering coeff., $\langle C \rangle$?	**NO!**	**OK**
-scale-free $p(k)$?	**NO!**	**NO!**

Random graphs vs real networks

	E-R	W-S
- sparseness, $\langle k \rangle$?	**OK**	**OK**
-the small-world effect, $\langle \ell \rangle$?	**OK**	**OK**
-large local clustering coeff., $\langle C \rangle$?	**NO!**	**OK**
-scale-free $p(k)$?	**NO!**	**NO!**

How to generate a **scale-free** random graph in a simple way?

The Barabási-Albert model

Growing networks...

- Erdős-Rényi model: fixed N, (static)
- Watts-Strogatz model: fixed N, (static)
- Real networks?

The Barabási-Albert model

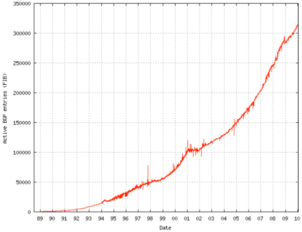
Growing networks...

Actor network

Number of movies in IMDB

Herr II, Bruce W., Ke, Weimao, Hardy, Elisha, and Börner, Katy. (2007) Movies and Actors: Mapping the Internet Movie Database. In Conference Proceedings of 11th Annual Information Visualization International Conference (IV 2007), Zurich, Switzerland, July 4-6, pp. 465-469.

Internet



Growth of the Internet routing table

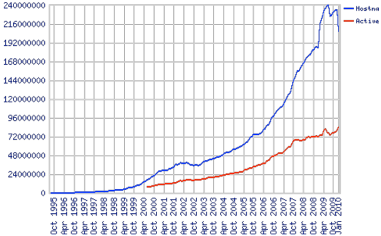
http://www.trainsignaltraining.com/ccna-ipv6

(from the slides of A.-L. Barabási)

The Barabási-Albert model

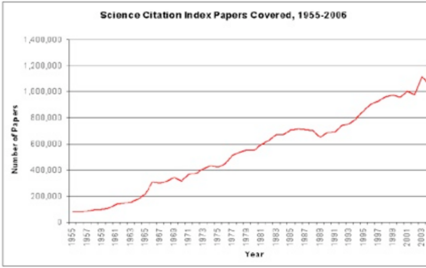
Growing networks...

WWW



http://website101.com/define-ecommerce-web-terms-definitions/

Scientific Publications



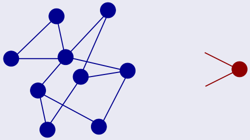
http://www.kk.org/thetechnium/archives/2008/10/the_expansion_o.php

(from the slides of A.-L. Barabási)

The Barabási-Albert model

The Barabási-Albert model

- Growing model: one new node in one step with m new links.



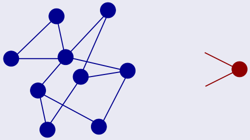
- The other end of the new links is attached according to the preferential attachment rule, i.e., we choose from the 'old' nodes with probabilities proportional to their degree:

$$\mathcal{P}(\text{choosing node } i) \equiv \mathcal{P}_i \sim k_i$$

The Barabási-Albert model

The Barabási-Albert model

- Growing model: one new node in one step with m new links.



- The other end of the new links is attached according to the preferential attachment rule, i.e., we choose from the 'old' nodes with probabilities proportional to their degree:

$$\mathcal{P}(\text{choosing node } i) \equiv \mathcal{P}_i \sim k_i$$

We are going to examine the B-A model first with the help of simulations.

We are going to examine the B-A model first with the help of simulations.

- Please take a look at the 2 accompanying videos showing a growing B-A network and an E-R graph, where we also add links one by one (in random order).

The Barabási-Albert model

Simulations

We are going to examine the B-A model first with the help of simulations.

- Please take a look at the 2 accompanying videos showing a growing B-A network and an E-R graph, where we also add links one by one (in random order).

→ Important differences are apparent already at such small system sizes: we can observe nodes that are likely candidates for becoming HUBS in the long run in the B-A video.

The Barabási-Albert model
Simulations

We are going to examine the B-A model first with the help of simulations.

- Please take a look at the 2 accompanying videos showing a growing B-A network and an E-R graph, where we also add links one by one (in random order).

→ Important differences are apparent already at such small system sizes: we can observe nodes that are likely candidates for becoming HUBS in the long run in the B-A video.

- Next, take a look at the empirical degree distribution of B-A networks generated using Networkx in the accompanying jupyter notebook.

We are going to examine the B-A model first with the help of simulations.

- Please take a look at the 2 accompanying videos showing a growing B-A network and an E-R graph, where we also add links one by one (in random order).

→ Important differences are apparent already at such small system sizes: we can observe nodes that are likely candidates for becoming HUBS in the long run in the B-A video.

- Next, take a look at the empirical degree distribution of B-A networks generated using Networkx in the accompanying jupyter notebook.

→ It looks scale-free!

- The scale-free $p(k)$ in the simulation results looks really satisfying.

The Barabási-Albert model
Simulations

- The scale-free $p(k)$ in the simulation results looks really satisfying.
- But can we also derive in more analytic terms that this model leads to a scale-free network?

- The scale-free $p(k)$ in the simulation results looks really satisfying.
- But can we also derive in more analytic terms that this model leads to a scale-free network?

$\rightarrow$ Actually, it can be proven also on pure analytic grounds that the B-A model is scale-free.

- The scale-free $p(k)$ in the simulation results looks really satisfying.
- But can we also derive in more analytic terms that this model leads to a scale-free network?

→ Actually, it can be proven also on pure analytic grounds that the B-A model is scale-free.

- However, here we are going to discuss a more simple (approximate) analytic derivation of the $p(k)$.

Main stages of the forthcoming calculation:

1) Calculation of the $k_i(t)$, corresponding to the degree of node i at time step t, given that node i appeared as the new node at a given time step t_i.
2) Based on the formula obtained for $k_i(t)$, the derivation of $p(k)$.

OK, so the first step is to study the time evolution of the degree for a given node in this model. Before diving into the formulas, a simple observation:

OK, so the first step is to study the time evolution of the degree for a given node in this model. Before diving into the formulas, a simple observation:

Due to the model definition, the node degree for any node i in any time step

- can increase if the newcomer node is actually choosing i to connect to,
- or it may remain the same (if the newcomer node is choosing someone else).

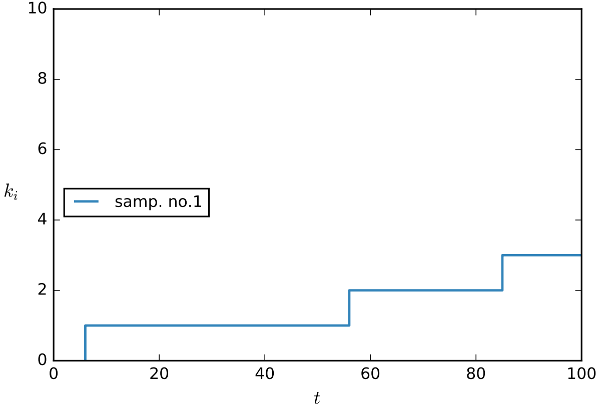
OK, so the first step is to study the time evolution of the degree for a given node in this model. Before diving into the formulas, a simple observation:

Due to the model definition, the node degree for any node i in any time step

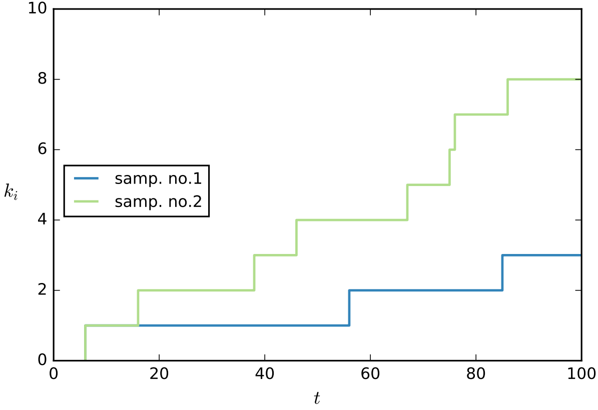
- can increase if the newcomer node is actually choosing i to connect to,
- or it may remain the same (if the newcomer node is choosing someone else).

Let us first take a look at the degree as a function of time in simulation results!

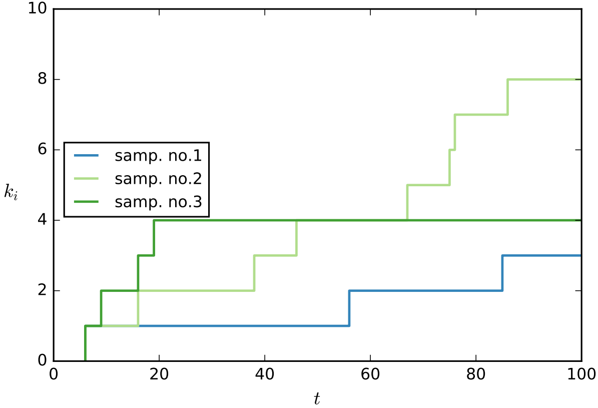
The degree of a particular node as a function of t in simulations:



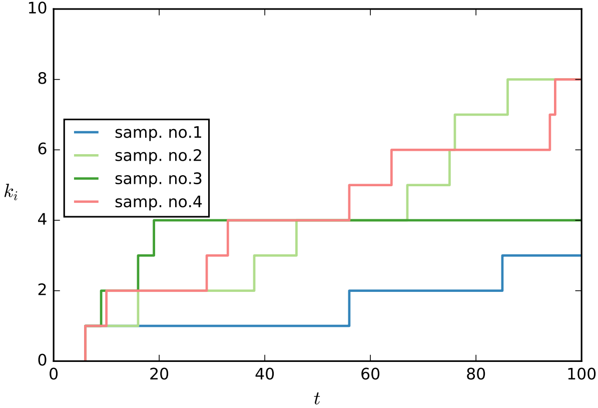
The degree of a particular node as a function of t in simulations:



The degree of a particular node as a function of t in simulations:

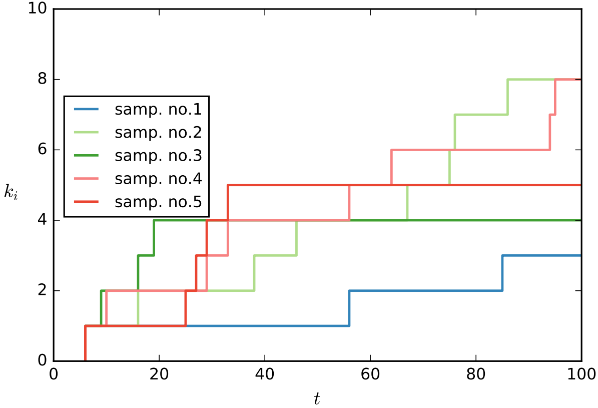


The degree of a particular node as a function of t in simulations:

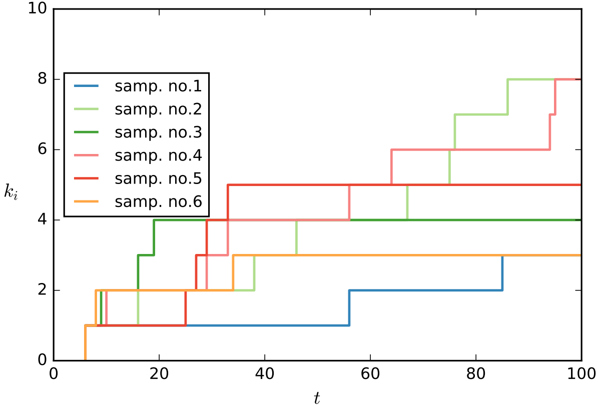


The Barabási-Albert model
Time dependent degree

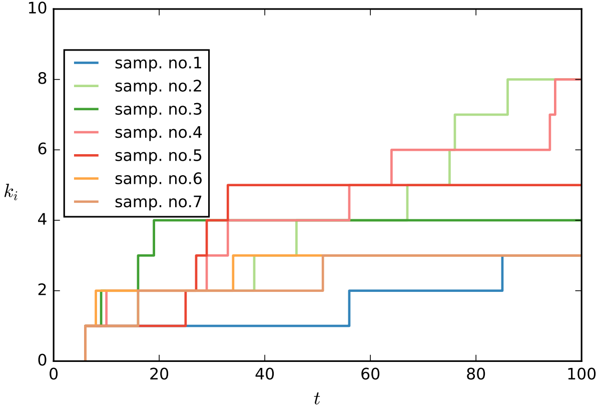
The degree of a particular node as a function of t in simulations:



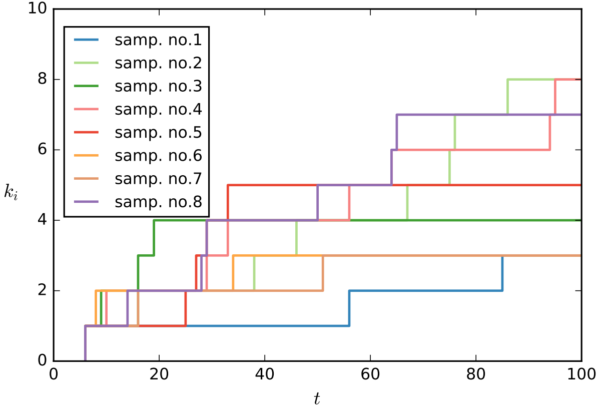
The degree of a particular node as a function of t in simulations:



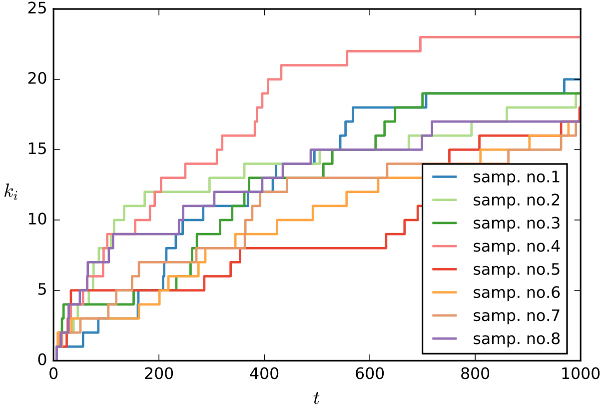
The degree of a particular node as a function of t in simulations:



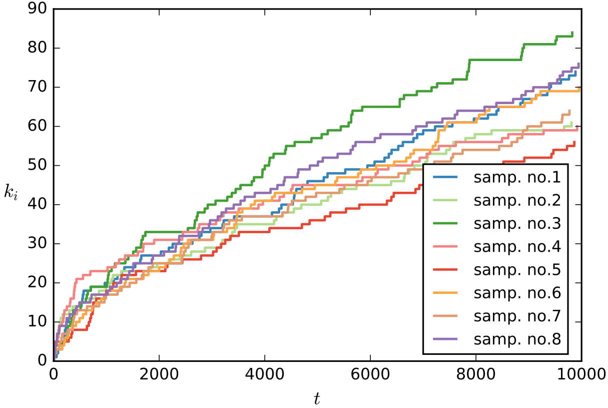
The degree of a particular node as a function of t in simulations:



The degree of a particular node as a function of t in simulations:

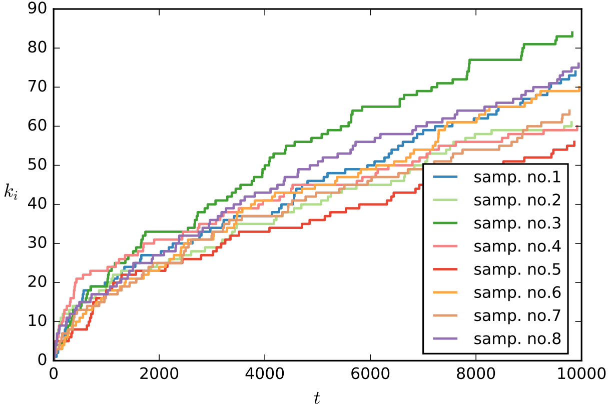


The degree of a particular node as a function of t in simulations:



The Barabási-Albert model
Time dependent degree

The degree of a particular node as a function of t in simulations:



We can see a clear trend. Next, we are going to derive an analytic formula for the average $k_i(t)$!

The Barabási-Albert model
Analytic $p(k)$

- For large t the number of nodes and links is

$$N \simeq t,$$
$$M \simeq mt.$$

- The probability of choosing i:

$$\mathcal{P}_i = \frac{k_i}{\sum_j k_j}.$$

→ The approximate change of k_i in a time step:

$$\frac{\Delta k_i}{\Delta t} \simeq m\mathcal{P}_i \quad \to \quad \frac{\partial k_i}{\partial t} = m\frac{k_i}{\sum_j k_j}.$$

Since the sum of the degrees : $\sum_j k_j = 2M = 2mt$,

$$\to \frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

- For large t the number of nodes and links is

$$N \simeq t,$$
$$M \simeq mt.$$

- The probability of choosing i:

$$\mathcal{P}_i = \frac{k_i}{\sum_j k_j}.$$

→ The approximate change of k_i in a time step:

$$\frac{\Delta k_i}{\Delta t} \simeq m\mathcal{P}_i \quad \rightarrow \quad \frac{\partial k_i}{\partial t} = m\frac{k_i}{\sum_j k_j}.$$

Since the sum of the degrees : $\sum_j k_j = 2M = 2mt$,

$$\rightarrow \frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

- For large t the number of nodes and links is

$$N \simeq t,$$
$$M \simeq mt.$$

- The probability of choosing i:

$$\mathcal{P}_i = \frac{k_i}{\sum_j k_j}.$$

$\rightarrow$ The approximate change of k_i in a time step:

$$\frac{\Delta k_i}{\Delta t} \simeq m\mathcal{P}_i \quad \rightarrow \quad \frac{\partial k_i}{\partial t} = m\frac{k_i}{\sum_j k_j}.$$

Since the sum of the degrees : $\sum_j k_j = 2M = 2mt$,

$$\rightarrow \frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

- For large t the number of nodes and links is

$$N \simeq t,$$
$$M \simeq mt.$$

- The probability of choosing i:

$$\mathcal{P}_i = \frac{k_i}{\sum_j k_j}.$$

→ The approximate change of k_i in a time step:

$$\frac{\Delta k_i}{\Delta t} \simeq m\mathcal{P}_i \quad \rightarrow \quad \frac{\partial k_i}{\partial t} = m\frac{k_i}{\sum_j k_j}.$$

Since the sum of the degrees : $\sum_j k_j = 2M = 2mt$,

$$\rightarrow \frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

- Solving the diff. eq.:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad \rightarrow \quad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \quad \rightarrow \quad \int \frac{dk_i}{k_i} = \int \frac{dt}{2t}$$

$$\rightarrow \quad \ln k_i = \frac{1}{2} \ln t + \text{const.} \quad \rightarrow \quad k_i(t) = ct^{\frac{1}{2}}$$

- How to determine the constant c?

At $t = t_i$, (the appearance of node i): $k_i = m$, leading to

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}.$$

- Thus, the degree of any node is increasing as $k(t) \sim t^{\beta}$, where $\beta = 1/2$ is the so-called dynamical exponent.

The Barabási-Albert model
Dynamical exponent

- Solving the diff. eq.:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad \rightarrow \quad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \quad \rightarrow \quad \int \frac{dk_i}{k_i} = \int \frac{dt}{2t}$$

$$\rightarrow \quad \ln k_i = \frac{1}{2}\ln t + \text{const.} \quad \rightarrow \quad k_i(t) = ct^{\frac{1}{2}}$$

- How to determine the constant c?

At $t = t_i$, (the appearance of node i): $k_i = m$, leading to

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}.$$

- Thus, the degree of any node is increasing as $k(t) \sim t^{\beta}$, where $\beta = 1/2$ is the so-called dynamical exponent.

- Solving the diff. eq.:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad \to \quad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \quad \to \quad \int \frac{dk_i}{k_i} = \int \frac{dt}{2t}$$

$$\to \quad \ln k_i = \frac{1}{2} \ln t + \text{const.} \quad \to \quad k_i(t) = c t^{\frac{1}{2}}$$

- How to determine the constant c?

  At $t = t_i$, (the appearance of node i): $k_i = m$, leading to

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}}.$$

- Thus, the degree of any node is increasing as $k(t) \sim t^{\beta}$, where $\beta = 1/2$ is the so-called dynamical exponent.

- Solving the diff. eq.:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad \to \quad \frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \quad \to \quad \int \frac{dk_i}{k_i} = \int \frac{dt}{2t}$$

$$\to \quad \ln k_i = \frac{1}{2}\ln t + \text{const.} \quad \to \quad k_i(t) = ct^{\frac{1}{2}}$$

- How to determine the constant c?

  At $t = t_i$, (the appearance of node i): $k_i = m$, leading to

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}.$$

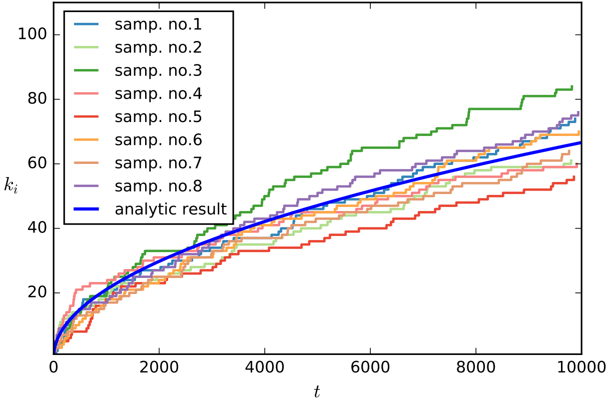
- Thus, the degree of any node is increasing as $k(t) \sim t^{\beta}$, where $\beta = 1/2$ is the so-called dynamical exponent.

The Barabási-Albert model

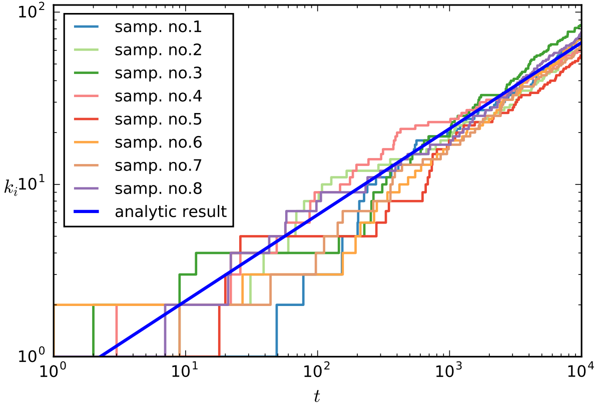
Dynamical exponent

Comparing the analytic result to the simulations:

Comparing the analytic result to the simulations:



Comparing the analytic result to the simulations:



- How to calculate the degree distribution from $k_i(t)$?

$\rightarrow$ The easiest way is by considering the cumulative distribution first.

The Barabási-Albert model

Analytic $p(k)$

- The cumulative degree distribution:

$$P(k) \equiv \mathcal{P}(k_i < k) = \mathcal{P}(m(t/t_i)^{1/2} < k) = \mathcal{P}(t/t_i < (k/m)^2) = \mathcal{P}(t_i/t > (m/k)^2).$$

- The lengths of the time steps:

$$P(k) = 1 - \left(\frac{m}{k}\right)^2$$

$$\rightarrow \mathbf{p(k)} = 2m^2\mathbf{k^{-3}}$$

- The cumulative degree distribution:

$$P(k) \equiv \mathcal{P}(k_i < k) = \mathcal{P}(m(t/t_i)^{1/2} < k) = \mathcal{P}(t/t_i < (k/m)^2) = \mathcal{P}(t_i/t > (m/k)^2).$$

- The lengths of the time steps:

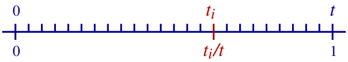
$$P(k) = 1 - \left(\frac{m}{k}\right)^2$$
$$\rightarrow \mathbf{p(k)} = 2m^2\mathbf{k}^{-3}$$

The Barabási-Albert model
Analytic $p(k)$

- The cumulative degree distribution:

$$P(k) \equiv \mathcal{P}(k_i < k) = \mathcal{P}(m(t/t_i)^{1/2} < k) = \mathcal{P}(t/t_i < (k/m)^2) = \mathcal{P}(t_i/t > (m/k)^2).$$

- The lengths of the time steps:



$$P(k) = 1 - \left(\frac{m}{k}\right)^2$$

$$\rightarrow \mathbf{p(k)} = 2m^2\mathbf{k^{-3}}$$

- The cumulative degree distribution:

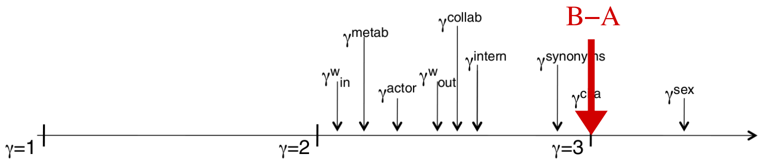
$$P(k) \equiv \mathcal{P}(k_i < k) = \mathcal{P}(m(t/t_i)^{1/2} < k) = \mathcal{P}(t/t_i < (k/m)^2) = \mathcal{P}(t_i/t > (m/k)^2).$$

- The lengths of the time steps:

$$P(k) = 1 - \left(\frac{m}{k}\right)^2$$
$$\rightarrow \mathbf{p(k)} = 2m^2\mathbf{k}^{-3}$$

- The degree distribution is **SCALE-FREE** with $\gamma = 3$!

The Barabási-Albert model



<k²> diverges | <k²> finite

<k> diverges | <k> finite

Ultra small world behavior | Small world

Regime full of anomalies… | The scale-free behavior is relevant | Behaves like a random network

(from the slides of A.-L. Barabási)

Average distance in scale-free networks

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

$\gamma \le 2$ and $2 < \gamma < 3$: Ultra Small World

$\gamma > 3$: Small World

- What is the **clustering coefficient** in the B-A model?

- What is the **clustering coefficient** in the B-A model?

→ We can calculate it in a rather simple way. As a first step, let us express the probability for nodes i and j, (introduced at time steps t_i and t_j) to be connected!

- What is the **clustering coefficient** in the B-A model?

→ We can calculate it in a rather simple way. As a first step, let us express the probability for nodes i and j, (introduced at time steps t_i and t_j) to be connected!

  - The only moment when they can connect is when the later coming node is actually the new node. Let's assume that i comes first and j comes later.

- What is the **clustering coefficient** in the B-A model?

→ We can calculate it in a rather simple way. As a first step, let us express the probability for nodes i and j, (introduced at time steps t_i and t_j) to be connected!

  - The only moment when they can connect is when the later coming node is actually the new node. Let's assume that i comes first and j comes later.
  - The degree of node i is given by $k_i(t) = m(\frac{t}{t_i})^{1/2}$

- What is the **clustering coefficient** in the B-A model?

→ We can calculate it in a rather simple way. As a first step, let us express the probability for nodes i and j, (introduced at time steps t_i and t_j) to be connected!

  - The only moment when they can connect is when the later coming node is actually the new node. Let's assume that i comes first and j comes later.
  - The degree of node i is given by $k_i(t) = m(\frac{t}{t_i})^{1/2}$
  - The probability of a connection to j introduced at $t = t_j$:

$$\mathcal{P}(i-j) = m\frac{k_i}{2mt} = \frac{k_i}{2t} = \frac{m\left(\frac{t_j}{t_i}\right)^{\frac{1}{2}}}{2t_j} = \frac{m}{2}(t_it_j)^{-\frac{1}{2}}.$$

- As the next step, let us write down the expected number of links between the neighbours of a node l, (introduced at t_l), at the end of the node generation process $t = N$:

$$\begin{aligned}
n_l &= \frac{1}{2}\sum_{t_i=1}^{N}\sum_{t_j=1}^{N}\mathcal{P}(l-i)\mathcal{P}(l-j)\mathcal{P}(i-j) \\
&\simeq \frac{1}{2}\int_1^N dt_i \int_1^N dt_j \mathcal{P}(l-i)\mathcal{P}(l-j)\mathcal{P}(i-j) \\
&= \frac{m^3}{16}\int_1^N dt_i \int_1^N dt_j (t_l t_i)^{-\frac{1}{2}}(t_l t_j)^{-\frac{1}{2}}(t_i t_j)^{-\frac{1}{2}} \\
&= \frac{m^3}{16 t_l}\int_1^N dt_i \frac{1}{t_i}\int_1^N dt_j \frac{1}{t_j} = \frac{m^3}{16 t_l}(\ln N)^2 .
\end{aligned}$$

- The number of "edge places" between the neighbors of l at the end $(t = N)$:

$$\frac{k_l(k_l - 1)}{2} \simeq \frac{k_l^2}{2} = \frac{1}{2}\left[m\left(\frac{t}{t_l}\right)^{\frac{1}{2}}\right]^2 = \frac{m^2 N}{2t_l}$$

- Thus, the clustering coefficient of l:

$$C_l = \frac{m^3}{16t_l}(\ln N)^2 \frac{2t_l}{m^2 N} = \frac{m(\ln N)^2}{8N}.$$

- Since we choose a general l, and C_l does not depend on l, our result for C in the B-A model:

$$C = \frac{m(\ln N)^2}{8N}.$$

$\langle C \rangle = \frac{m(\ln N)^2}{8N}$ → **Is this good or bad?**

→ Let's compare it to the Erdős-Rényi model:

- Erdős-Rényi: $\langle C \rangle \simeq \frac{\langle k \rangle}{N} \sim \frac{1}{N}$
- Barabási-Albert: $\langle C \rangle = \frac{m(\ln N)^2}{8N} \sim \frac{(\ln N)^2}{N}$

Slower decay with N, however, no decay at all in real networks:

$\langle C\rangle = \frac{m(\ln N)^2}{8N}$ → **Is this good or bad?**

→ Let's compare it to the Erdős-Rényi model:

- Erdős-Rényi: $\langle C\rangle \simeq \frac{\langle k\rangle}{N} \sim \frac{1}{N}$
- Barabási-Albert: $\langle C\rangle = \frac{m(\ln N)^2}{8N} \sim \frac{(\ln N)^2}{N}$

Slower decay with N, however, no decay at all in real networks:

$\langle C\rangle = \frac{m(\ln N)^2}{8N}$ → **Is this good or bad?**

→ Let's compare it to the Erdős-Rényi model:

- Erdős-Rényi: $\langle C\rangle \simeq \frac{\langle k\rangle}{N} \sim \frac{1}{N}$
- Barabási-Albert: $\langle C\rangle = \frac{m(\ln N)^2}{8N} \sim \frac{(\ln N)^2}{N}$

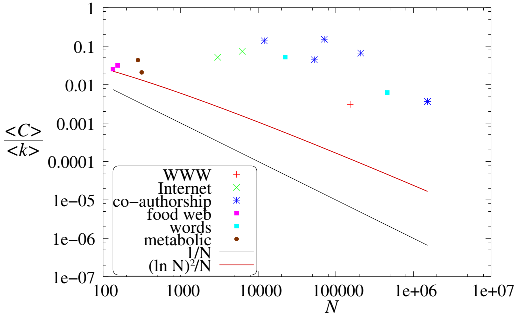
Slower decay with N, however, no decay at all in real networks:

$\langle C\rangle = \frac{m(\ln N)^2}{8N}$ → **Is this good or bad?**

→ Let's compare it to the Erdős-Rényi model:

- Erdős-Rényi: $\langle C\rangle \simeq \frac{\langle k\rangle}{N} \sim \frac{1}{N}$
- Barabási-Albert: $\langle C\rangle = \frac{m(\ln N)^2}{8N} \sim \frac{(\ln N)^2}{N}$

Slower decay with N, however, no decay at all in real networks:



The Barabási-Albert model

Summary

- **Growth** and **preferential attachment**,
- Dynamical exponent: $k(t) \sim t^{\beta}$, $\beta = \frac{1}{2}$,
- **SCALE-FREE**: $p(k) \sim k^{-3}$, $\gamma = \mathbf{3}$,
- Small-world: $\langle l \rangle \sim \frac{\ln N}{\ln \ln N}$,
- $C \sim \frac{(\ln N)^2}{N}$.

Variations of the B-A model

Remaining challenges, problems?

Variations of the B-A model

Remaining challenges, problems:

- Original B-A: $\gamma = 3$.

  $\updownarrow$

  In real networks $2 \leq \gamma \leq 3$.

- Original B-A: oldest nodes have highest degree.

  $\updownarrow$

  Not always in real systems, e.g., Google.

- Low clustering coefficient.

- A simple solution to the problem of **tunable** γ **exponent** is given by one of the practice exercise, involving a **uniform additional fitness** beside the degree.

B-A model and fitness

To enable **newcomers overtake old nodes** in the B-A model we need a further **multiplicative fitness**:

- We introduce a fitness parameter η for the nodes, drawn from some distribution $\rho(\eta)$.
- The probability for node i to acquire new links:

$$\mathcal{P}_i \sim \eta_i k_i \quad \rightarrow \quad \mathcal{P}_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

→ The time evolution of the degree now depends also on the fitness:

$$k_i(t) \sim t^{\beta(\eta_i)}, \quad \beta(\eta) = c\eta$$

- This also means that now a later coming node with higher fitness (and higher β exponent) can actually overtake older but less fit nodes.

G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**,5632 (2001)

B-A model and fitness

To enable **newcomers overtake old nodes** in the B-A model we need a further **multiplicative fitness**:

- We introduce a fitness parameter η for the nodes, drawn from some distribution $\rho(\eta)$.
- The probability for node i to acquire new links:

$$\mathcal{P}_i \sim \eta_i k_i \quad \rightarrow \quad \mathcal{P}_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

→ The time evolution of the degree now depends also on the fitness:

$$k_i(t) \sim t^{\beta(\eta_i)}, \quad \beta(\eta) = c\eta$$

- This also means that now a later coming node with higher fitness (and higher β exponent) can actually overtake older but less fit nodes.

G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**,5632 (2001)

B-A model with high clustering

- P. Holme and B. J. Kim: **extra triad formation steps**.

  - Add a new node according to the P.A. rule.
  - After each step, with probability p also form m triangles with new node and randomly chosen neighbours of the nodes the new node was just attached.

B-A model with high clustering

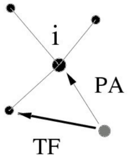
- P. Holme and B. J. Kim: **extra triad formation steps**.

  - Add a new node according to the P.A. rule.
  - After each step, with probability p also form m triangles with new node and randomly chosen neighbours of the nodes the new node was just attached.

B-A model with high clustering

- P. Holme and B. J. Kim: **extra triad formation steps**.

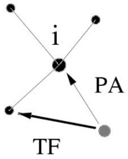
  - Add a new node according to the P.A. rule.
  - After each step, with probability p also form m triangles with new node and randomly chosen neighbours of the nodes the new node was just attached.

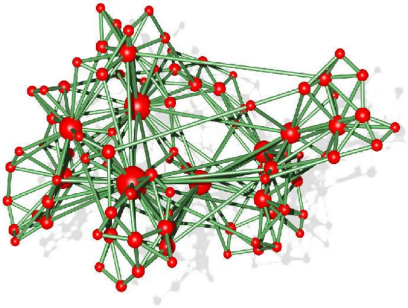


B-A model with high clustering

Network models

- P. Holme and B. J. Kim: **extra triad formation steps**.
  - Add a new node according to the P.A. rule.
  - After each step, with probability p also form m triangles with new node and randomly chosen neighbours of the nodes the new node was just attached.

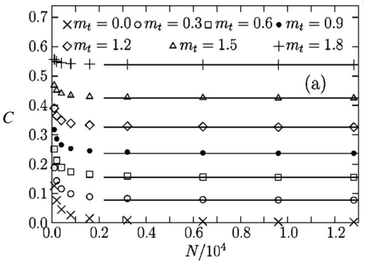


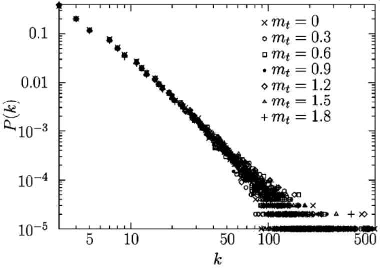


HK graph with p = 0.9

Scale-free model with high clustering

- Holme-Kim model:





$\rightarrow$ tunable clustering coefficient!

(The larger p we set, the larger is $\langle C \rangle$ in the resulting network).

P. Holme and B. J. Kim, *Phys. Rev. E* **65**, 026107 (2002)

Further reading on the Barabási-Albert model (not compulsory):

- Network science book by A.-L. Barabási, chapter 5: `http://networksciencebook.com/chapter/5`
- Wikipedia: `https://en.wikipedia.org/wiki/Barabasi-Albert_model`

THE CONFIGURATION MODEL

The configuration model
Motivation

- What if we do not care about the **mechanism** leading to scale-free $p(k)$ and are interested only in the **actual form** of the degree distribution?

The configuration model
Motivation

- What if we do not care about the **mechanism** leading to scale-free $p(k)$ and are interested only in the **actual form** of the degree distribution?
- How to generate a random network with arbitrary $p(k)$?

The configuration model

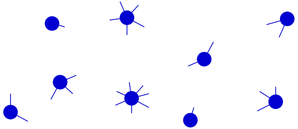
The configuration model

- The goal is to generate a network with N nodes and a given $p(k)$ corresponding to a random sample from all possible realisations.
- We first draw N times from $p(k)$ to obtain the degree sequence.
- Then the “half links” are joined to connect the nodes with each other.

The configuration model

The configuration model

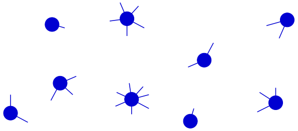
- The goal is to generate a network with N nodes and a given $p(k)$ corresponding to a random sample from all possible realisations.
- We first draw N times from $p(k)$ to obtain the degree sequence.
- Then the “half links” are joined to connect the nodes with each other.



The configuration model

The configuration model

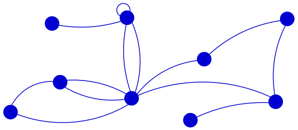
- The goal is to generate a network with N nodes and a given $p(k)$ corresponding to a random sample from all possible realisations.
- We first draw N times from $p(k)$ to obtain the degree sequence.
- Then the “half links” are joined to connect the nodes with each other.



The configuration model

The configuration model

- The goal is to generate a network with N nodes and a given $p(k)$ corresponding to a random sample from all possible realisations.
- We first draw N times from $p(k)$ to obtain the degree sequence.
- Then the "half links" are joined to connect the nodes with each other.



The configuration model

- If you want a simple graph (without self-connections and multi-edges), connecting the nodes is the hard part...
- How would you start?

The configuration model

- The usual tactics is to start the connection process at the large degree nodes.

The configuration model

- The usual tactics is to start the connection process at the large degree nodes.
- However, this way the obtained graph cannot be regarded as completely random sample...

The configuration model

- The usual tactics is to start the connection process at the large degree nodes.
- However, this way the obtained graph cannot be regarded as completely random sample...

$\rightarrow$ Randomisation is needed!

Randomisation

- If we keep only N and M fixed during the randomisation, we end up with an Erdős-Rényi graph in the end.

Randomisation

- If we keep only N and M fixed during the randomisation, we end up with an Erdős-Rényi graph in the end.
- The aim is to get rid of any extra correlation coming from the node connection process, but also preserve the degree of the nodes.

Randomisation

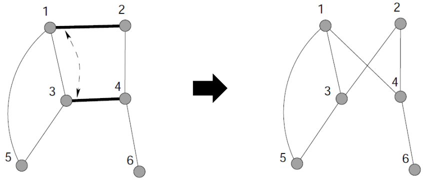
- If we keep only N and M fixed during the randomisation, we end up with an Erdős-Rényi graph in the end.
- The aim is to get rid of any extra correlation coming from the node connection process, but also preserve the degree of the nodes.
- How can we do that?

Randomisation

Link randomisation

An intuitive idea is link randomisation:

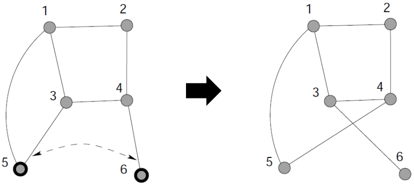
- in every step, choose 2 links at random,
- and swap one end of the links.
- (Of course, we have to check that we do not introduce self connections or multi edges before actually carrying out the swap).



Randomization
Node-randomization

Another possibility is node randomisation:

- in every step, choose 2 nodes at random,
- and swap one link on the nodes.
- (Again, we have to check that we do not introduce self connections or multi edges before actually carrying out the swap).



Randomisation

How long do we have to keep on swapping the links?

Randomisation

How long do we have to keep on swapping the links?

→ A general rule of the thumb is that the number of swaps should be high enough so that on average every link was swapped a few times, i.e., a the number of swaps should be something like 3 or 5 times the number of links.

Randomisation and the configuration model

Now that we can randomise networks in a way that the node degrees remain the same, comparing any given real network to its configuration model counterpart has become very easy:

Randomisation and the configuration model

Now that we can randomise networks in a way that the node degrees remain the same, comparing any given real network to its configuration model counterpart has become very easy:

We just **randomise a copy** of the original network (again, the number of carried out rewirings should be a few times the number of links), and the **resulting graph can be viewed as a sample from the configuration model**, where $p(k)$ **is exactly the same** as in the original network!

Randomisation and the configuration model

Now that we can randomise networks in a way that the node degrees remain the same, comparing any given real network to its configuration model counterpart has become very easy:

We just **randomise a copy** of the original network (again, the number of carried out rewirings should be a few times the number of links), and the **resulting graph can be viewed as a sample from the configuration model**, where $p(k)$ **is exactly the same** as in the original network!

$\rightarrow$ The configuration model is also a very important random 'baseline' to which we can compare a given real network.

Randomisation

- Any other application of randomisation apart from the configuration model?

Measuring significance by randomization

- Real networks are not completely random nor completely regular...
- To what extent is a given feature (e.g., $\langle \ell \rangle$, C, k_{nn}, etc.) is due to some intrinsic structure, and to what extent is it random?

$\rightarrow$ Let's randomise the network!