Social Networks & Recommendation Systems

VI. Evolving networks.

Grzegorz Siudem

Warsaw University of Technology

MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.

Before classes

Reminder

From SNARS_2:

- The history of the Barábasi and Alberty model.

Other topics:

- binomial distribution $\mathbb{P}(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$,
- Chapman-Kolmogorov equation,
- methods for solving recurrence equations.

Lecture

Historical background: reminder.

Known at this moment graph models did not have real networks' properties

- no fat-tailed distributions, but also
- low clustering coefficients,
- lack of scale-free property.

A.-L. Barabási, R. Albert, Emergence on scaling in random networks, Science, 286:509-512, 1999.

Historical background: reminder.

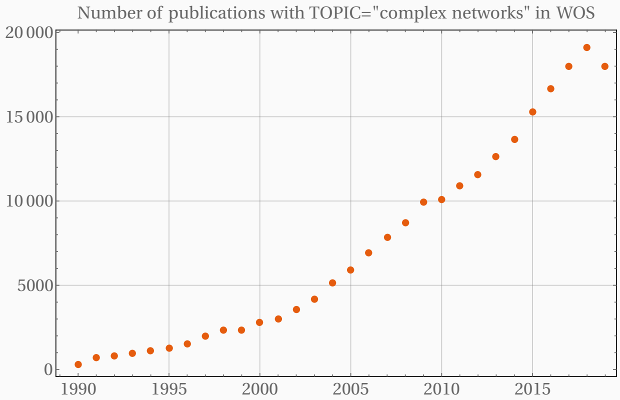
Known at this moment graph models did not have real networks' properties

- no fat-tailed distributions, but also
- low clustering coefficients,
- lack of scale-free property.

A.-L. Barabási, R. Albert, Emergence on scaling in random networks, Science, 286:509-512, 1999.

Some people argue that this date begins the complex networks.
Is it right?

Development of the complex networks science



Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.

Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.
- we connect edges to the previous vertices due to the preferential attachment rule.

Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.
- we connect edges to the previous vertices due to the preferential attachment rule.

Both rules are reasonable:

World wide web was an inspiration for the authors:

Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.
- we connect edges to the previous vertices due to the preferential attachment rule.

Both rules are reasonable:

World wide web was an inspiration for the authors:

- this networks growth.

Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.
- we connect edges to the previous vertices due to the preferential attachment rule.

Both rules are reasonable:

World wide web was an inspiration for the authors:

- this networks growth.
- one observe Matthew effect.

Empirical justification

BA construction algorithm based on two assumptions:

- the network evolves i.e. in every time step we add new vertex and connected edges.
- we connect edges to the previous vertices due to the preferential attachment rule.

Both rules are reasonable:

World wide web was an inspiration for the authors:

- this networks growth.
- one observe Matthew effect.

The above also applies to more network...

Examples?

BA Algorithm – description

- At $t = 0$ we start with complete graph with $m_0 \geqslant 1$ vertices.

BA Algorithm – description

- At $t = 0$ we start with complete graph with $m_0 \geqslant 1$ vertices.
- At every of the consecutive time steps we add new vertex with $m \leqslant m_0$ edges according to preferential attachment rule

$$\Pi(k_i) \propto k_i.$$

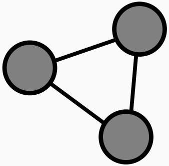
BA Algorithm – description

- At $t = 0$ we start with complete graph with $m_0 \geqslant 1$ vertices.
- At every of the consecutive time steps we add new vertex with $m \leqslant m_0$ edges according to preferential attachment rule

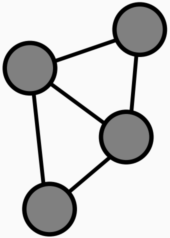
$$\Pi(k_i) \propto k_i.$$

- The procedure ends at any time $t = N$.

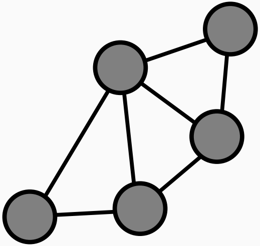
BA Algorithm – visualization



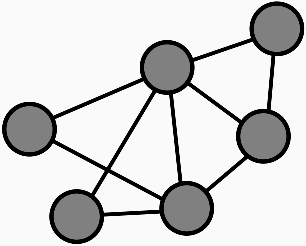
BA Algorithm – visualization



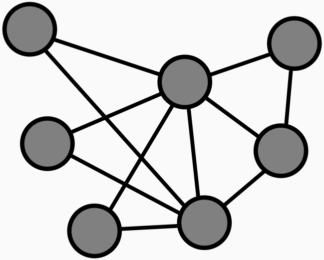
BA Algorithm – visualization



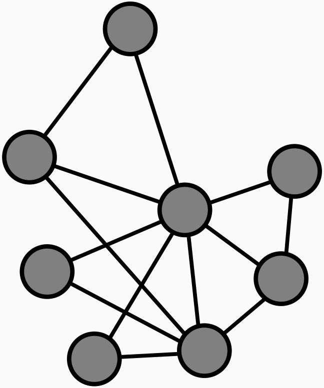
BA Algorithm – visualization



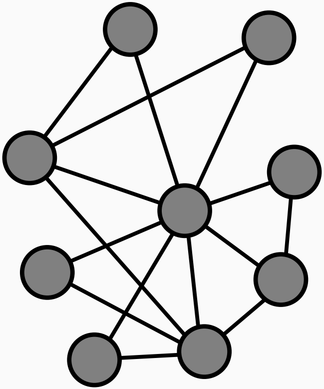
BA Algorithm – visualization



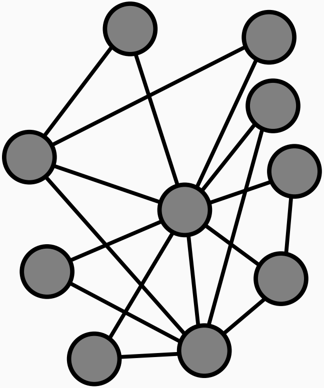
BA Algorithm – visualization



BA Algorithm – visualization



BA Algorithm – visualization



after a while...

BA Algorithm – visualization

BA Algorithm – analytical solution

Number of edges and vertices

$$N = t + m_0 \approx t,$$

$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

BA Algorithm – analytical solution

Number of edges and vertices

$$N = t + m_0 \approx t,$$

$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Preferential attachment rule

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

Why?

BA Algorithm – analytical solution

Number of edges and vertices

$$N = t + m_0 \approx t,$$
$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Preferential attachment rule

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

Why?

Common approaches to the problem:

- continuous-time method (in the mean-field approach),
- master equation.

Number of edges and vertices

$$N = t + m_0 \approx t,$$
$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Preferential attachment rule

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

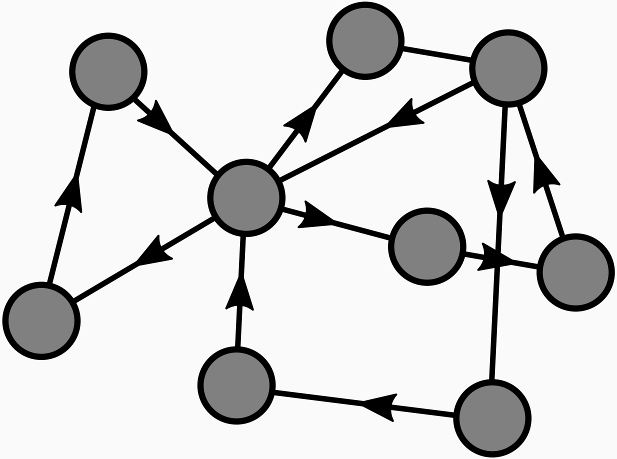
Why?

Common approaches to the problem:

- continuous-time method (in the mean-field approach),
- master equation.

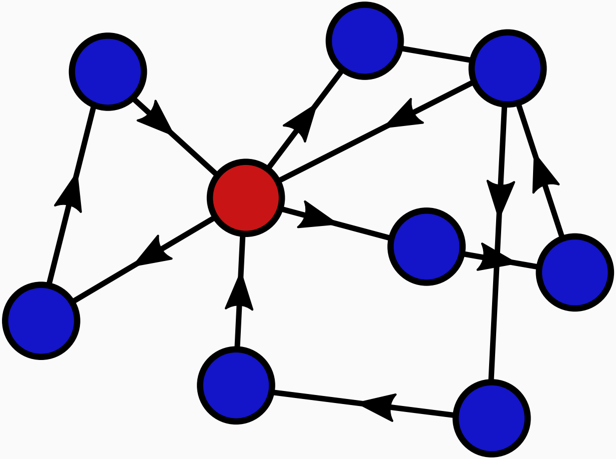
How would you approach to solve this problem?

Mean-field approach – as physicists do



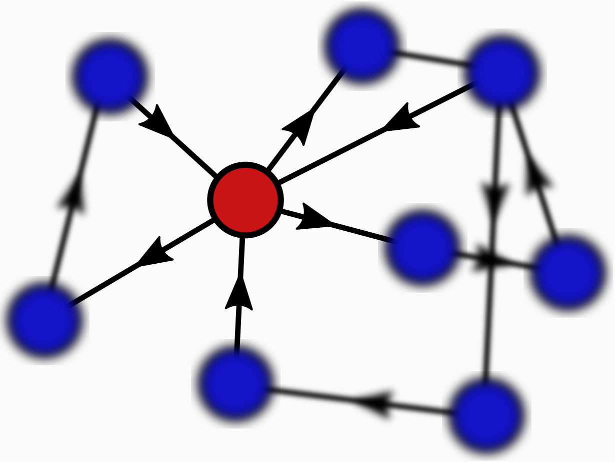
Statistical physics problems are often difficult because they are *tangled*...

Mean-field approach – as physicists do



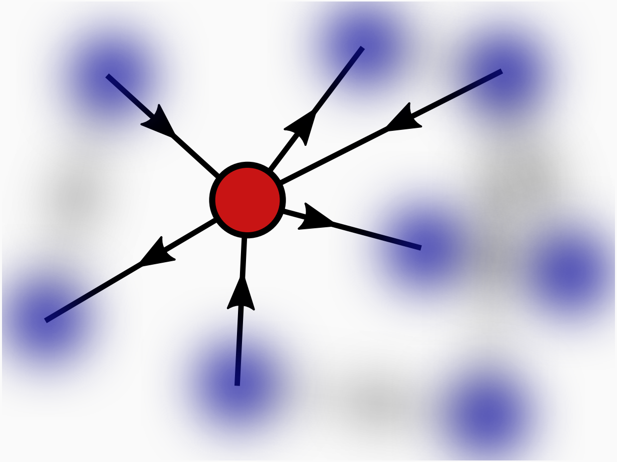
How to untangle them?

Mean-field approach – as physicists do



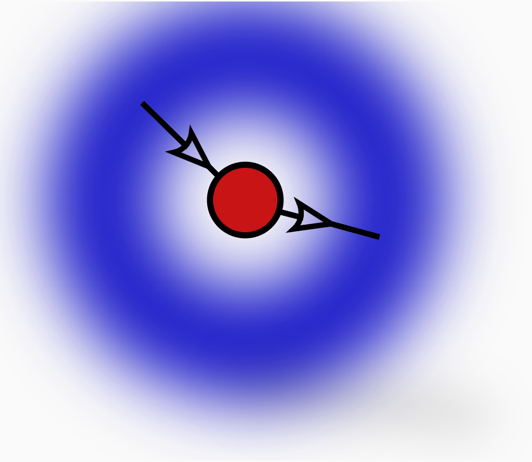
Let us simplify the problem...

Mean-field approach – as physicists do



... and instead tangled fields consider *mean-field*.

Mean-field approach – as physicists do



We get a similar system, but easier to solve!

How to apply mean-field for BA algorithm?

The expected value replaces the random variable:

- Let k_i denotes *expected* (mean) degree of i-th vertex.
- So we allow non-integer values!

How to apply mean-field for BA algorithm?

The expected value replaces the random variable:

- Let k_i denotes *expected* (mean) degree of i-th vertex.
- So we allow non-integer values!

Continuous time:

- We assume that the new edges are attached continuously and not discreetly.
- Though new edges appear at discrete moments t_i.

How to apply mean-field for BA algorithm?

The expected value replaces the random variable:

- Let k_i denotes *expected* (mean) degree of i-th vertex.
- So we allow non-integer values!

Continuous time:

- We assume that the new edges are attached continuously and not discreetly.
- Though new edges appear at discrete moments t_i.

Mean field:

- New edges are distributed independently of each other.

Continuous time approach

Vertex degree changes

In line with the assumptions made, the degrees come from the binomial distribution:

$$\frac{dk_i}{dt} = \sum_{l=0}^{m} l \binom{m}{l} [\Pi(k_i)]^l [1 - \Pi(k_i)]^{m-l} = m\Pi(k_i) = \frac{k_i}{2t}$$

$$k_i(t_i) = m,$$

where t_i is the time of connection of the i-th vertex.

Continuous time approach

Vertex degree changes

In line with the assumptions made, the degrees come from the binomial distribution:

$$\frac{dk_i}{dt} = \sum_{l=0}^{m} l \binom{m}{l} \left[\Pi(k_i)\right]^l \left[1 - \Pi(k_i)\right]^{m-l} = m\Pi(k_i) = \frac{k_i}{2t}$$

$$k_i(t_i) = m,$$

where t_i is the time of connection of the i-th vertex.

We solve differential equation

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Continuous time approach – continuation

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Let us consider distribution $\mathcal{P}(k_i)$

$$\mathcal{P}(k_i) = T(t_i)\left|\frac{dk_i}{dt_i}\right|^{-1},$$

where $T(t_i)$ is the density distribution of times t_i

$$T(t_i) = \frac{1}{t}$$

Why?

Continuous time approach – continuation

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Let us consider distribution $\mathcal{P}(k_i)$

$$\mathcal{P}(k_i) = T(t_i)\left|\frac{dk_i}{dt_i}\right|^{-1},$$

where $T(t_i)$ is the density distribution of times t_i

$$T(t_i) = \frac{1}{t}$$

Why?

Combining the three equations above leads to:

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

Let's derive it!

Importance of the assumptions

Barábas-Albert algorithm has only two assumptions:

- network growths,
- new edges are added due to the preferential attachment rule.

Can any of these assumptions be ignored?

Let’s check it!

Random connections (A model)

With the mean-field approach obtain degree distribution in the evolving network in which

- we add new vertex at every time step.
- edges are distributed randomly i.e.

$$\Pi(k_i) = \frac{1}{t + m_0} \approx \frac{1}{t}.$$

Random connections (A model)

With the mean-field approach obtain degree distribution in the evolving network in which

- we add new vertex at every time step.
- edges are distributed randomly i.e.

$$\Pi(k_i) = \frac{1}{t + m_0} \approx \frac{1}{t}.$$

Solution:

Differential equation of the form

$$\frac{dk_i}{dt} = \frac{m}{t},$$

has solution $k_i(t) = m \ln\left(\frac{t}{t_i}\right) + m$, which leads to $\mathcal{P}(k) = \frac{e}{m} e^{-k/m}$.

Network with fixed size (B model)

With the mean-field approach (as far as possible!) determine the degree distribution of the network in which

- The number of vertices is from start constant and equal to N.
- The edges are distributed with preferential attachment rule.

Network with fixed size (B model)

With the mean-field approach (as far as possible!) determine the degree distribution of the network in which

- The number of vertices is from start constant and equal to N.
- The edges are distributed with preferential attachment rule.

Solution:

Following differential equation

$$\frac{dk_i}{dt} = \frac{N-1}{N}\frac{k_i}{2t} + \frac{1}{N},$$

has the solution of the form

$$k_i(t) = \frac{2(N-1)}{N(N-2)}t \approx \frac{2}{N}t,$$

but how to get the distribution from this?

Summary

Homework

Read *Personal Introduction* in A.-L. Barabási, *Network Science*
`http://networksciencebook.com/chapter/0`

European Funds
Knowledge Education Development

Warsaw University of Technology

European Union
European Social Fund

MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.

Thank you for your attention!