Social Networks & Recommendation Systems

VII. Probabilistic aspects of the complex networks.

Grzegorz Siudem

Warsaw University of Technology

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.

Before classes

Reminder

From SNARS_5:

- Properties of the ER graphs.

From SNARS_6:

- Mean-field approach to the BA model.

From other courses:

- generating functions approach in the combinatorics.

To think about:

- What do you think, which of the graphs: Erdős-Rényi or Barábasi-Albert is more vulnerable for intetional attacks and random failures? Why?

Lecture

Impressions after the derivations for the BA model?

Impressions after the derivations for the BA model?

Pros

- IT WORKS!

Impressions after the derivations for the BA model?

Pros

- IT WORKS!
- relatively simple (right?),

Impressions after the derivations for the BA model?

Pros

- IT WORKS!
- relatively simple (right?),
- short and based on the very basic tools.

Impressions after the derivations for the BA model?

Pros

- IT WORKS!
- relatively simple (right?),
- short and based on the very basic tools.

Cons

- unreal assumptions,

Impressions after the derivations for the BA model?

Pros

- IT WORKS!
- relatively simple (right?),
- short and based on the very basic tools.

Cons

- unreal assumptions,
- unverifiable approximations,

Impressions after the derivations for the BA model?

Pros

- IT WORKS!
- relatively simple (right?),
- short and based on the very basic tools.

Cons

- unreal assumptions,
- unverifiable approximations,
- lack of the mathematical precision.

Impressions after the derivations for the BA model?

Pros

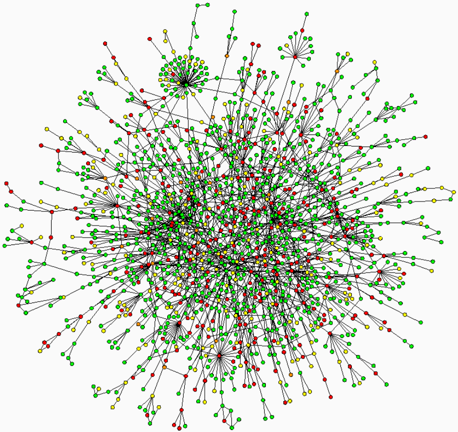
- IT WORKS!
- relatively simple (right?),
- short and based on the very basic tools.

Cons

- unreal assumptions,
- unverifiable approximations,
- lack of the mathematical precision.

Conclusion:
I stronlgy recommend reading Durret to those who are dissatisfied!

R. Durret



https://services.math.duke.edu/~rtd/RGD/RGD.html

Let's have a look to chapter 4.1.

Master equation

The equation describing the changes in the probability distribution over time

$$\frac{d\mathcal{P}_i}{dt} = \sum_j \mathcal{P}_j T_{j\to i} - \sum_j \mathcal{P}_j T_{i\to j},$$

which in the discrete version takes the following form

$$\mathcal{P}_i(t+1) - \mathcal{P}_i(t) = \sum_j \mathcal{P}_j(t) T_{j\to i} - \sum_j \mathcal{P}_j(t) T_{i\to j},$$

and this is what we will focus on.

Master equation for BA networks

We will follow

- chapter 4.1 in Durret's book,
- S.N. Dorogovstev, J.F.F. Mendes i A.N Samukhin, *Structure of growing networks with preferential linking*, Phys. Rev. Letters. 85, 4633–4636 (2000).

Master equation for BA networks

We will follow

- chapter 4.1 in Durret's book,
- S.N. Dorogovstev, J.F.F. Mendes i A.N Samukhin, *Structure of growing networks with preferential linking*, Phys. Rev. Letters. 85, 4633–4636 (2000).

Master equation takes the form:

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$

$$N_{m-1}(t) = 0.$$

Master equation for BA networks

We will follow

- chapter 4.1 in Durret's book,
- S.N. Dorogovstev, J.F.F. Mendes i A.N Samukhin, *Structure of growing networks with preferential linking*, Phys. Rev. Letters. **85**, 4633–4636 (2000).

Master equation takes the form:

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$

$$N_{m-1}(t) = 0.$$

Question:

Can you justify the components of this equation?

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$
$$N_{m-1}(t) = 0.$$

Interpretation:

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$
$$N_{m-1}(t) = 0.$$

Interpretation:

- $\frac{m(k-1)}{2mt} N_{k-1}(t)$ the number of vertices that has just promoted to the k-th level.

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$
$$N_{m-1}(t) = 0.$$

Interpretation:

- $\frac{m(k-1)}{2mt} N_{k-1}(t)$ the number of vertices that has just promoted to the k-th level.
- $\frac{mk}{2mt} N_k(t)$ the number of vertices that has just promoted to the $k+1$-st level.

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$
$$N_{m-1}(t) = 0.$$

Interpretation:

- $\frac{m(k-1)}{2mt} N_{k-1}(t)$ the number of vertices that has just promoted to the k-th level.
- $\frac{mk}{2mt} N_k(t)$ the number of vertices that has just promoted to the $k+1$-st level.
- δ_{km} newly added vertex.

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{m(k-1)}{2mt} N_{k-1}(t) - \frac{mk}{2mt} N_k(t) + \delta_{km}$$

$$N_{m-1}(t) = 0.$$

Interpretation:

- $\frac{m(k-1)}{2mt} N_{k-1}(t)$ the number of vertices that has just promoted to the k-th level.
- $\frac{mk}{2mt} N_k(t)$ the number of vertices that has just promoted to the $k+1$-st level.
- δ_{km} newly added vertex.

Question:

Is this solution mathematically exact?

Master equation for BA networks

Inaccuracies in the master equation approach:

- adding an edge does not change preferences (synchronous dynamics, mean-field theory reminiscence),

Master equation for BA networks

Inaccuracies in the master equation approach:

- adding an edge does not change preferences (synchronous dynamics, mean-field theory reminiscence),
- we allow for multiple connections.

Master equation for BA networks

Inaccuracies in the master equation approach:

- adding an edge does not change preferences (synchronous dynamics, mean-field theory reminiscence),
- we allow for multiple connections.

However, this approach works perfectly!

Why?

Master equation for BA networks

Inaccuracies in the master equation approach:

- adding an edge does not change preferences (synchronous dynamics, mean-field theory reminiscence),
- we allow for multiple connections.

However, this approach works perfectly!

Why?

Precision vs. simplicity of the method?

i.e. mathematicians vs. physicists...

Master equation for BA networks

$$N_k(t+1) - N_k(t) = \frac{k-1}{2t} N_{k-1}(t) - \frac{k}{2t} N_k(t) + \delta_{km}$$
$$N_{m-1}(t) = 0.$$

How to solve it?

`goto` Project;

Master equation for BA networks – solution

$$N_k(t+1) - N_k(t) = \frac{k-1}{2t} N_{k-1}(t) - \frac{k}{2t} N_k(t) + \delta_{km}.$$

$$N_{m-1}(t) = 0.$$

Solution:

$$\mathcal{P}(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Master equation for BA networks – solution

$$N_k(t+1) - N_k(t) = \frac{k-1}{2t} N_{k-1}(t) - \frac{k}{2t} N_k(t) + \delta_{km}.$$
$$N_{m-1}(t) = 0.$$

Solution:

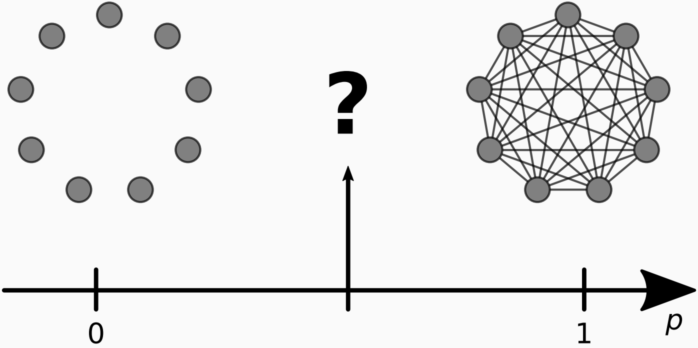
$$\mathcal{P}(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Question:

Argue that for $k \gg 1$ above results agrees with mean-field approach.

Percolation – ER graphs

Let us now return to the ER graphs



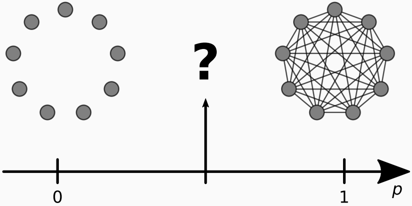
What is happening in the middle? We will follow sec. 4.3.2 in Fronczak and Fronczak book.

What is percolation?

wikipedia

Mathematical model of liquid leakage through porous material.

How does this relate to graphs?

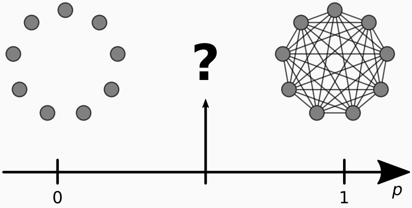


We ask when the graf is percolated

i.e. when percolation claster's size is equal to $N^* \propto N$.

Percolation – ER graphs

How does this relate to graphs?



We ask when the graf is percolated

i.e. when percolation claster's size is equal to $N^* \propto N$.

Attention!

We will approach in the physicist way. For more detailed approaches (sic!) see chapter 2 in Durret's book.

Percolation – ER graphs

Let us introduce distribution $\mathcal{Q}(k)$

$\mathcal{Q}(k)$ is the distribution of the degrees on the ends on the randomly selected edge.

Percolation – ER graphs

Let us introduce distribution $\mathcal{Q}(k)$

$\mathcal{Q}(k)$ is the distribution of the degrees on the ends on the randomly selected edge.

During project we will justify that $\mathcal{Q}(k) \propto k\mathcal{P}(k)$, thus

$$\mathcal{Q}(k) = \frac{k}{\langle k \rangle}\mathcal{P}(k).$$

Let us introduce distribution $\mathcal{Q}(k)$

$\mathcal{Q}(k)$ is the distribution of the degrees on the ends on the randomly selected edge.

During project we will justify that $\mathcal{Q}(k) \propto k\mathcal{P}(k)$, thus

$$\mathcal{Q}(k) = \frac{k}{\langle k \rangle}\mathcal{P}(k).$$

During Project we will prove that percolation threshold can be defined as

$$\sum_k k\mathcal{Q}(k) \geqslant 2,$$

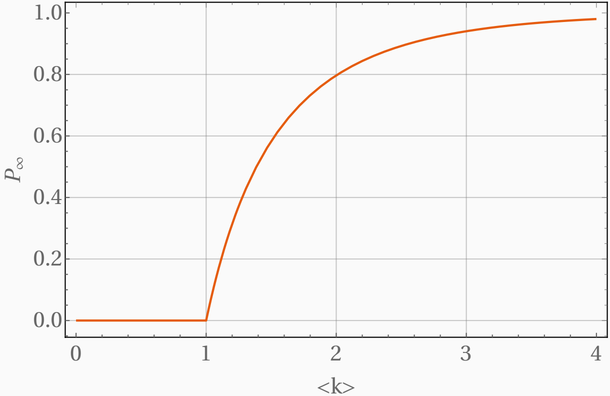
which is equivalent to $\langle k \rangle_{nn} = \frac{\langle k^2 \rangle}{\langle k \rangle} = 2$.

For ER graphs we obtain

$$\langle k \rangle = 1,$$

which means

$$p_c = \frac{1}{N}.$$



Thank you for your attention!

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.