Social Networks & Recommendation Systems

I. Mathematical background: graph theory refresher.

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

Short reminder from the previous courses

Discrete Mathematics (see e.g. R.J.Wilson, *Introduction to Graph Theory*)

- What are graphs? What do they composed of?
- What types of graphs do you know?
- How can graphs be represented in computer's memory?

Probability Theory (see e.g. W.Feller, *An Introduction to Probability Theory and its Applications*)

- What are the differences between the binomial, Poisson, and Cauchy distributions?
- What does mean that distribution has *a fat tail*?

Algorithmic

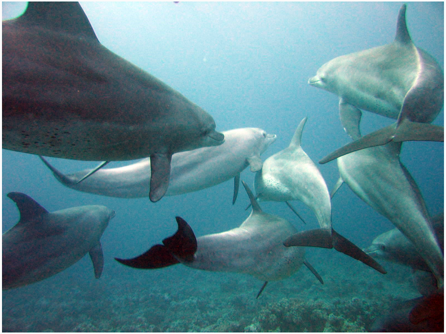
- What kind of graph searching algorithms do you know?

Lecture

What are complex networks?

Complex Networks:

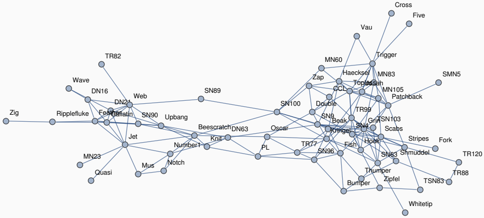
- Real Data,
- Graph Theory,
- Mathematical Models,
- Computer Science's Techniques.



Indo-Pacific Bottlenose Dolphin, Tursiops aduncus. Red Sea. [Wiki]

Complex Networks:

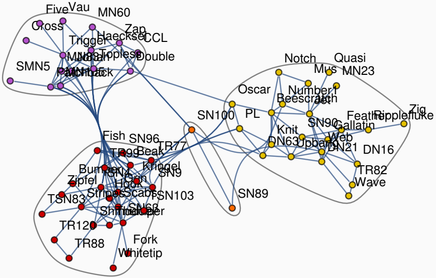
- Real Data,
- Graph Theory,
- Mathematical Models,
- Computer Science’s Techniques.



An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand. [WolframMathematica]

Complex Networks:

- Real Data,
- Graph Theory,
- Mathematical Models,
- Computer Science’s Techniques.

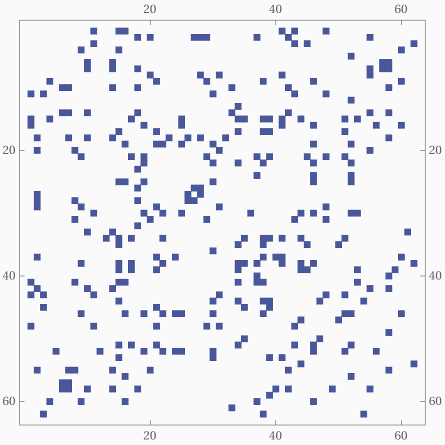


An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand. [WolframMathematica]

What are complex networks?

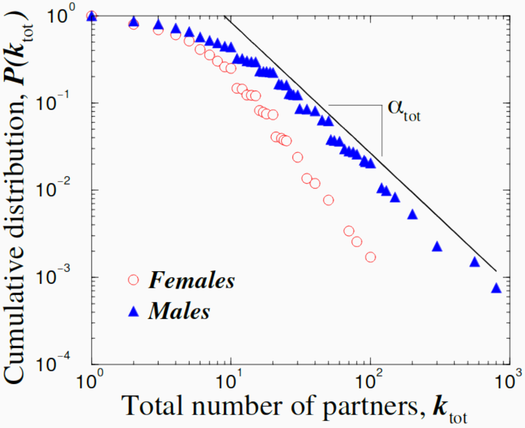
Complex Networks:

- Real Data,
- Graph Theory,
- Mathematical Models,
- Computer Science's Techniques.



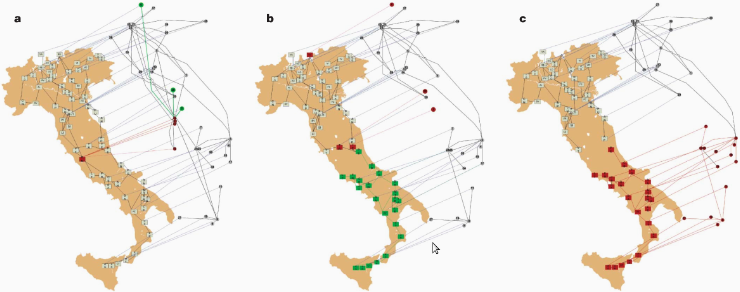
An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand. [WolframMathematica]

Motivation: networks are sexy!



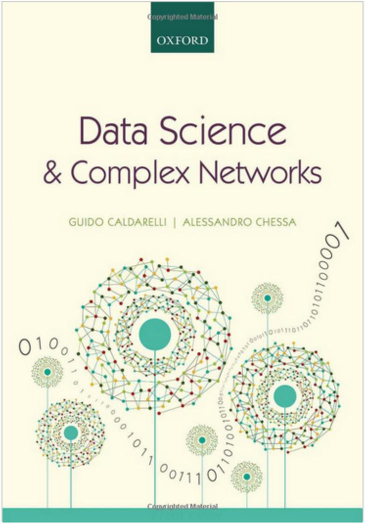
Degree distribution of the sexual contacts network [F. Liljeros i in. Nature **411**, 907–908 (2001)]

Motivation: real-life applications



Coupled networks: communication and electrical [S.V Buldyrev i in. Nature **464**, 1025–1028, (2010)]

Motivation: networks + data



Course content

Introduction

1. Mathematical background.
2. Historical overview of the complex network science.
3. Real networks: their analysis and visualization.

- Reminder and extension of the knowledge of graph theory.
- Problems with the power law distributions.
- Methods of representation of graphs in computers memory.
- Introduction to the software environments.

Course content

Introduction

1. Mathematical background.
2. Historical overview of the complex network science.
3. Real networks: their analysis and visualization.

- What is the complex network science?
- Review of the databases of networks.
- Import and preliminary analysis of complex networks.

Course content

Introduction

1. Mathematical background.
2. Historical overview of the complex network science.
3. Real networks: their analysis and visualization.

- Typical properties of the complex networks.
- Power laws in nature.
- Graph visualization algorithms.

Course content

Models and analysis of networks

4. Network’s metrics overview.
5. Static random graphs.
6. Evolving networks.
7. Probabilistic aspects of complex networks.
8. Hierachical, layered and temporal networks.

- Detection and analysis of power distributions in data.
- Vertices correlation measures - assortativity vs. disassortativity.
- Clustering coefficients and measures of centrality.
- Erdős and Bacon numbers, Hirsch index and other specific network metrics.

Course content

Models and analysis of networks

4. Network's metrics overview.
5. Static random graphs.
6. Evolving networks.
7. Probabilistic aspects of complex networks.
8. Hierachical, layered and temporal networks.

- Erdős-Rényi random graphs.
- Watts-Strogatz model.
- Configuration model.
- Networks with given hamiltonian.

Course content

Models and analysis of networks

4. Network's metrics overview.
5. Static random graphs.
6. Evolving networks.
7. Probabilistic aspects of complex networks.
8. Hierachical, layered and temporal networks.

- Matthew effect.
- Barabási-Albert model.
- Mean field approach.

Course content

Models and analysis of networks

4. Network's metrics overview.
5. Static random graphs.
6. Evolving networks.
7. Probabilistic aspects of complex networks.
8. Hierachical, layered and temporal networks.

- (More) strict approach to the BA model.
- Percolation in the ER graphs.
- Intentional attacks and random failures in networks.

Course content

Models and analysis of networks

4. Network's metrics overview.
5. Static random graphs.
6. Evolving networks.
7. Probabilistic aspects of complex networks.
8. Hierachical, layered and temporal networks.

- Hierarchical networks.
- Layered networks.
- Multiplex networks.
- Temporal networks.

Course content

Dynamics on the networks

9. Random walks.
10. Community detection.
11. Agent Based Models.

- Reminder of the Markov processes.
- Diffusion on the networks.
- PageRank

Course content

Dynamics on the networks

9. Random walks.
10. Community detection.
11. Agent Based Models.

- What are communities in networks?
- How to detect them?

Course content

Dynamics on the networks

9. Random walks.
10. Community detection.
11. Agent Based Models.

- Agent based models.
- Social interactions models.
- Epidemic models
- Spreading of news, gossips, memes etc.

Course content

Applications

12. Social Networks.
13. Recommendation Systems
14. Individual project.
15. Individual project

- Analysis of the social networks.
- Dynamical processes in social networks.

Course content

Applications

12. Social Networks.
13. Recommendation Systems
14. Individual project.
15. Individual project

- How to **optimally** choose a movie for the evening?

Course content

Applications

12. Social Networks.
13. Recommendation Systems
14. Individual project.
15. Individual project

- Topics to be selected from a prepared list or (recommended!) to be proposed by yourself.

Project content

1.-13. Small project - illustrations for the lecture.

14.-15. Big individual project.

- We work together and discuss the consecutive steps of the solution.
- Projects are examples for the lecture.

Project content

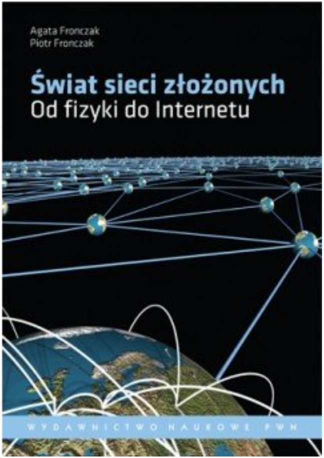
1.-13. Small project - illustrations for the lecture.

14.-15. Big individual project.

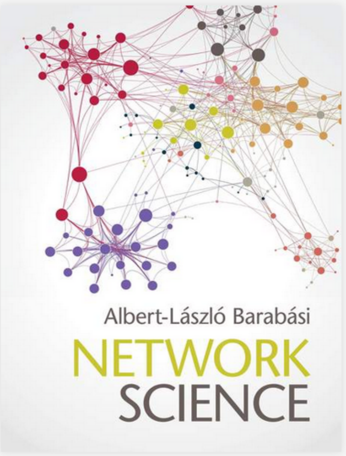
- Topics to be selected from a prepared list or (recommended!) to be proposed by yourself.
- No lectures - consultation classes.

Suggested literature

- Resources on the homepage of the cours (primary source)
- Świat sieci złożonych. Od fizyki do Internetu, A.Fronczak and P.Fronczak (source highly recommended, unfortunatelly only in Polish)
- Data Science and Complex Networks, G. Caldarelli i A. Chessa (source highly recommended)
- Random Graph Dynamics, R. Durrett (source highly recommended)
- Network Science, A.-L. Barabási (source highly recommended)
- Networks. An Introduction, M.E.J. Newman
- Foundation of data science
- ...

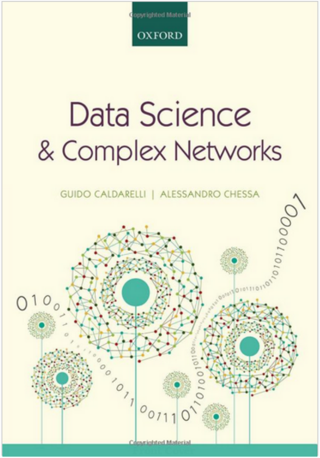


The book is very accessible, almost popular, but at the same time a very good textbook. The first and probably the only one in Polish.

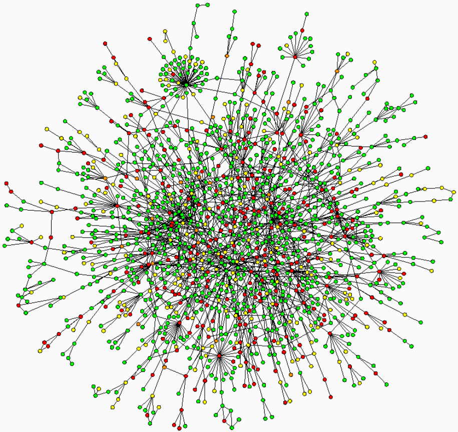


`http://networksciencebook.com/`

Very good introduction to complex network science. Available for free online.


Copyrighted Material
OXFORD
Data Science
& Complex Networks
GUIDO CALDARELLI | ALESSANDRO CHESSA
Copyrighted Material

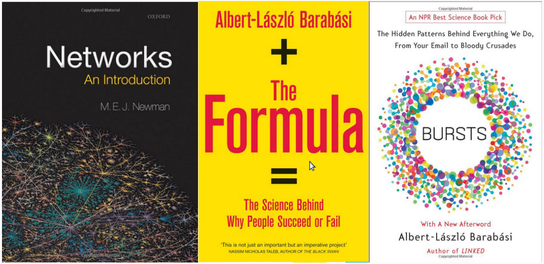
R. Durret



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

A serious mathematical approach to network science.

Other books


Copyrighted Material
OXFORD
Networks
An Introduction
M. E. J. Newman
Albert-László Barabási
+
The
Formula
=
The Science Behind
Why People Succeed or Fail
'This is not just an important but an imperative project'
NASSIM NICHOLAS TALEB, AUTHOR OF THE BLACK SWAN
Copyrighted Material
An NPR Best Science Book Pick
The Hidden Patterns Behind Everything We Do,
From Your Email to Bloody Crusades
BURSTS
With A New Afterword
Albert-László Barabási
Author of LINKED
Copyrighted Material

Graph Theory refresher

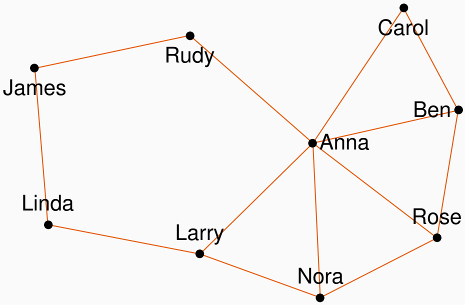
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

Edges are un-oriented pairs of vertices.



Example from `ExampleData[]` Wolfram Mathematica:

A social network of a school swimming team.

Graph Theory refresher

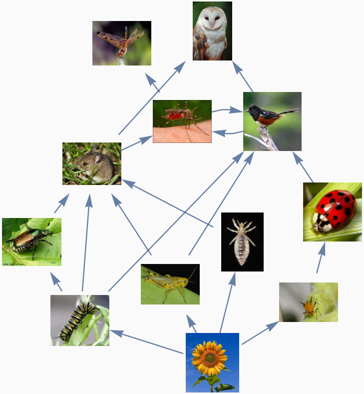
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

Edges are oriented pairs of vertices.



Example from `ExampleData[]` Wolfram Mathematica:

A simple food web.

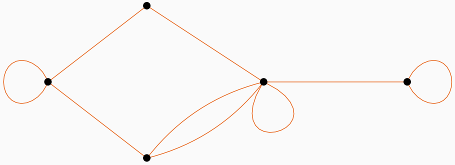
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

We allows multiple edges to have the same pair of endpoints as well as self-loops.



Example.

Graph Theory refresher

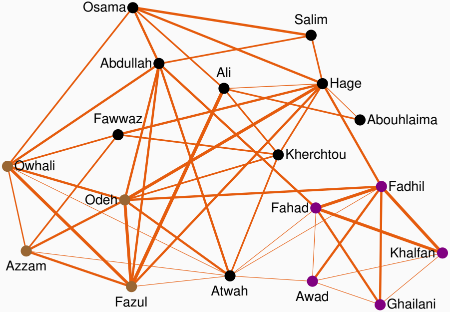
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

Every edge has assigned a number (the weight).



Example from `ExampleData[]` Wolfram Mathematica:

Terrorist network linked to the 1998 bombings of the US embassies in Kenya and Tanzania.

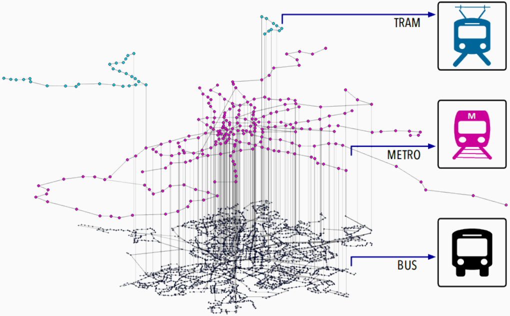
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

Edges belong to the different layers.



A. Aleta and Y. Moreno, Annual Review of Condensed Matter Physics 10:1, 45-62, (2019)

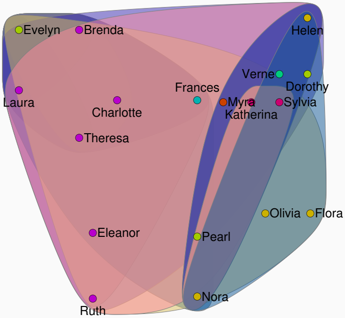
Definition

Graph is a pair (V, E), where V is non-empty set of vertices and E is a set of edges.

Types of graphs

- Simple Graphs.
- Directed Graph.
- Multigraphs.
- Weighted Graphs.
- Layered Graphs.
- (Hipergraphs)

We allows for non-binary relations.



Example from `ExampleData[]` Wolfram Mathematica:

The network of southern women social club.

Graph Theory Refresher – methods of graph representation

Adjacency matrix

Adjacency matrix of graph G with N vertices is a $N \times N$ matrix $A = [a_{ij}]$ such that

$$a_{ij} = \text{strength of the edge between} i \text{ and } j.$$

Adjacency list

Adjacency list of graph G is a list of its edges.

Graph theory refresher

Vertex degree

Degree of the vertex k_i is a number of the edges between i and other vertices

$$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji}.$$

Vertex degree – directed graphs

$$k_i^{in} \neq k_i^{out},$$

$$k_i^{in} = \sum_{j=1}^{N} a_{ij}, \quad k_i^{out} = \sum_{j=1}^{N} a_{ji}.$$

Graph theory refresher

Path

Path in the graph $G = (V, E)$ is a tupple of edges $\{\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_n, v_{n+1}\}\}$, where every $v_i \in V$, and $\{v_i, v_{i+1}\} \in E$ for $i = 1, 2, \ldots, n$. **Length of the path** is equal to n.

Distance in graph

Distance $d(i, j)$ between vertex i and vertex j in the graph G is the length of the shortest path between i and j.

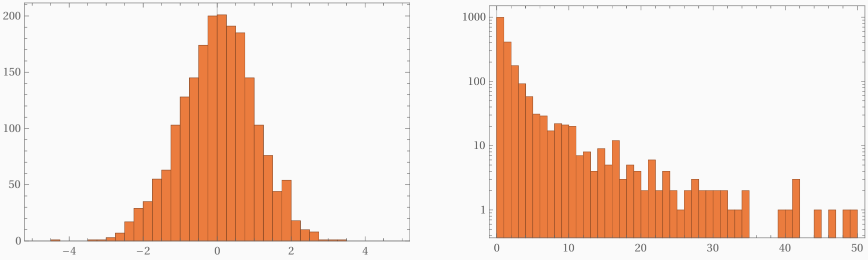
Graph connectivity

Graph is **connected** when for every pair of its vertices exists connecting them path.

In directed graphs we distinguish between weak nad strong connectivity.

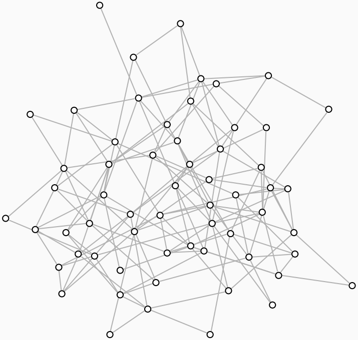
Question: What are differences?

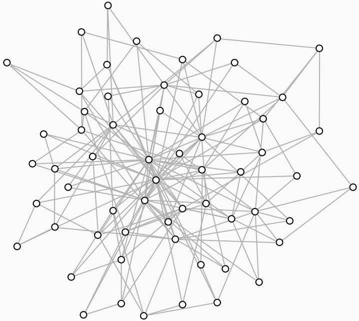
Fat tails

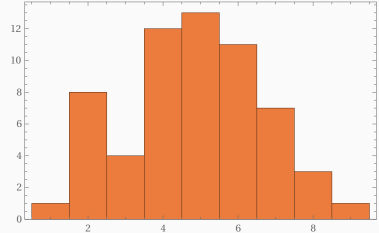


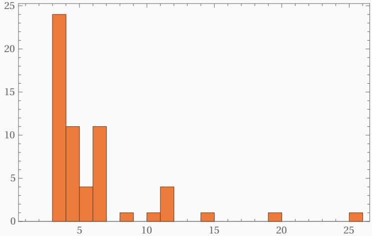
In the case of fat-tailed distributions, our intuitions fail!

Fat tails in networks – epidemic threshold

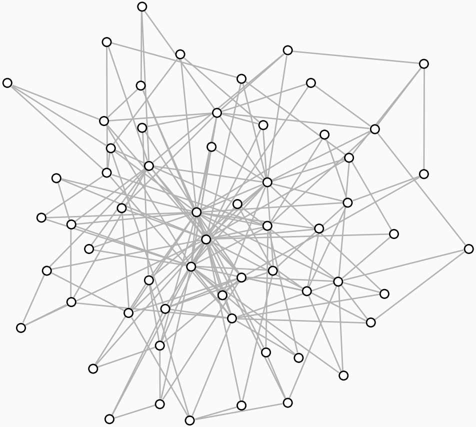




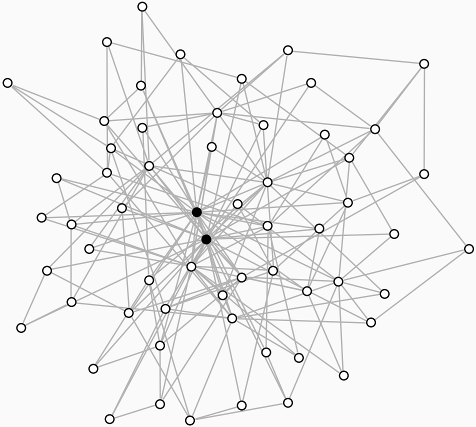




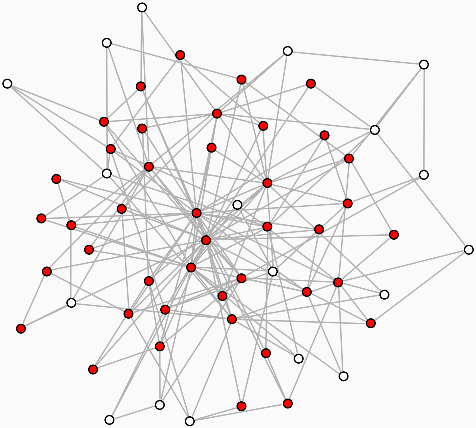
Fat tails in networks – epidemic threshold



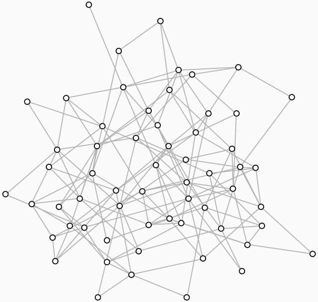
Fat tails in networks – epidemic threshold



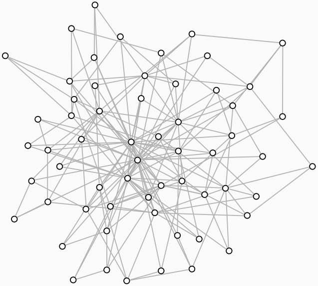
Fat tails in networks – epidemic threshold



Fat tails in networks – epidemic thresholdi



$$\lambda_c = \frac{1}{\langle k \rangle} > 0$$



$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \to 0$$

More of the epidemic models on lecture 11.

Network models vs. dynamics models

Network models (5-7,9,10)

- Deterministic vs. random,
- Static vs. evolving,

Dynamics on networks (11,12)

- diffusion (AKA random walk),
- opinion's spreading,
- random failures and intentional attacks,
- epidemics.

During project classes you will use

Python/R

- according to your preferences.

Summary

Homework

Suggest graph data for the next class.

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.