Wrocław University of Technology

Data Science Group

datasciencegroup.pl

Computational Network Science (Obliczeniowa nauka o sieciach)

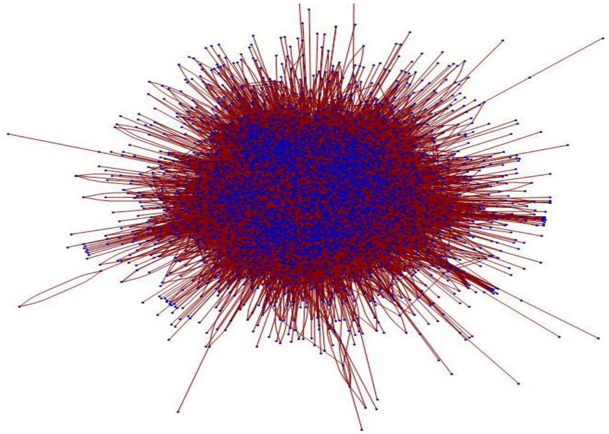
Przemysław Kazienko

Wroclaw University of Science and Technology, Poland

http://www.ii.pwr.edu.pl/~kazienko/

Poznań, April 22, 2016

Social Network Group
@ Wroclaw University of Technology
http://www.zsi.pwr.wroc.pl/~sna/



Outline

- Network Science and Complex Systems
- Computational Network Science
  - Why
  - Tasks
  - Collective classification & its open issues

Complex systems & complex networks

- Behind each **complex system** there is a **network**, that defines the interactions between the component

- We will never understand complex system unless we map out and understand the **networks behind them**

Central quantities in network science

Degree distribution: **P(k)**

Path length: **<*d*>**

Clustering coefficient: $C_i = \frac{2e_i}{k_i(k_i - 1)}$

Współczynnik gronowania

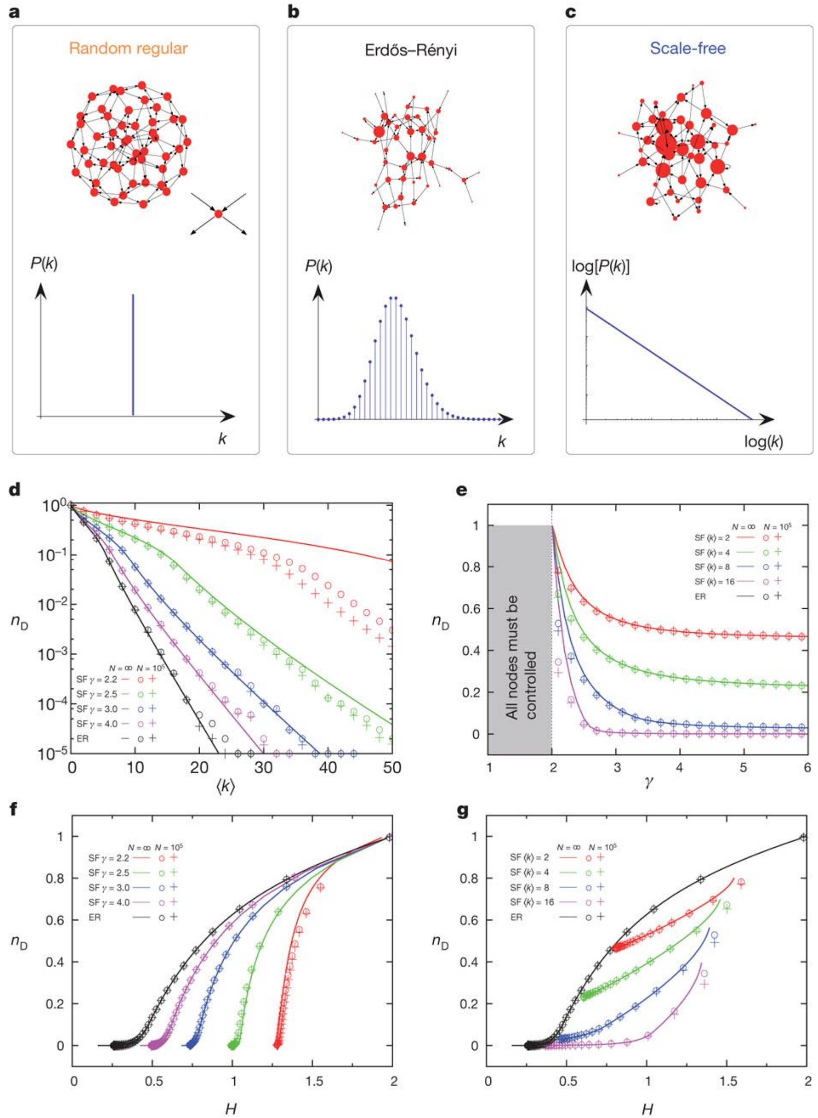
Network Science: Break-through Discovery

Structure matters!!!

The impact
of network structure
on the number
of **driver nodes**

Controllability

Y-Y Liu *et al.* *Nature* **473**, 167-173 (2011)



Wrocław University of Technology

Network Science: problems considered

- Influence of network structures on
  - Resistance/robustness
    - power grids,
    - failure cascades
  - Controllability
  - Diffusion processes
    - Epidemics, Immunology
    - Spreading of opinions, competiting opinions
- Brain networks
- Network reduction
- Dynamics of structures

Computer Science

\+ Network Science

= Computational Network Science

Why computational network science is important?

CNS: Tasks 1

1. Training models from data for reasoning
   - **Relational ML** - collective classification
     - **node** classification
   - **Link classification:**
     - link prediction and/or resilience/survival
   - Entity resolution
     - clustering for networks

CNS: Tasks 2 & 3

2. Network(s) extraction from data
   – Data structuring
   – Behavioral data quantification
   – Extraction of relationships from events
   – Data integration and fusion
     - Multiple sources integration, external and internal
     - Network creation from **unstructured data**, e.g.texts
   – Information networks (Jiawei Han)

3. **High Performance Computing - HPC**

Collective classification

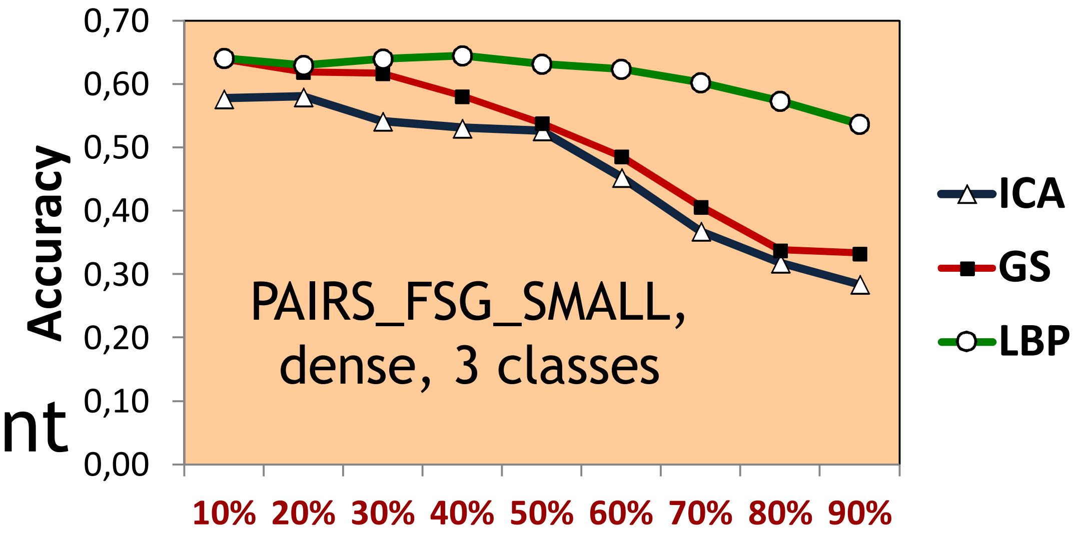
- **No iid** limitation like in classical ML
  (independent and identically distributed)

- **Classification** of nodes and links based on the **structure** – relationships **ONLY**

Problem 1: Sparsely Labelled Networks

- Problem
  - accuracy depends on the amount of class label **information available at inference**

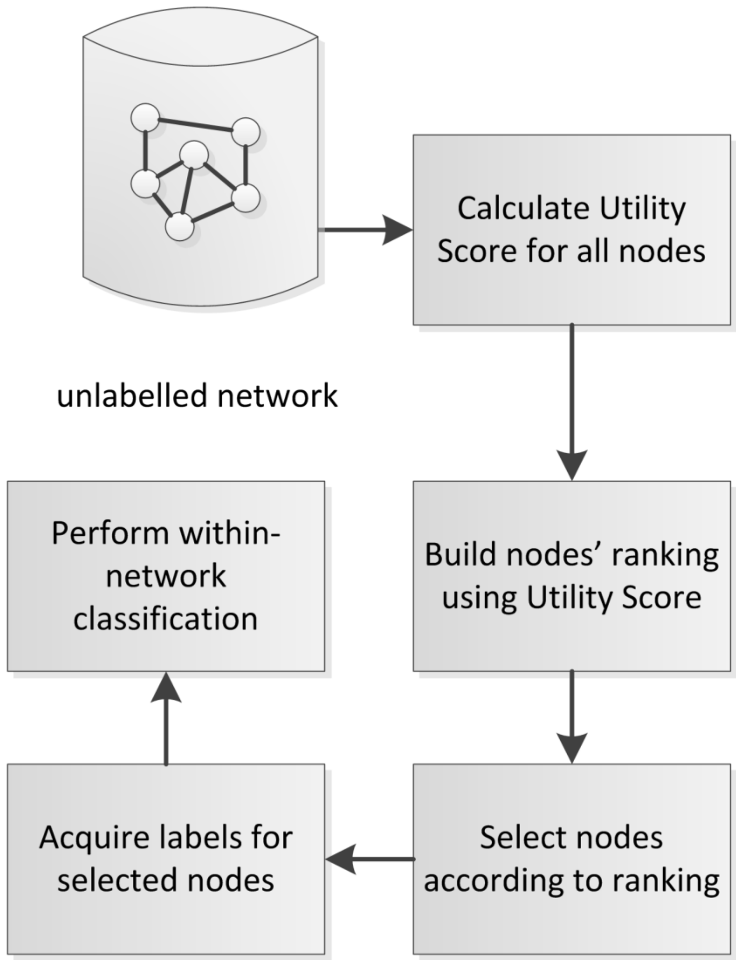
- Possible solution
  - semi-supervised collective classication
  - or label-dependent classication



Kajdanowicz T., Kazienko P., Janczak M.: *Collective Classification Techniques: an Experimental Study*. **ADBIS 2012 - MCSD 2012** - Mining Complex and Stream Data, New Trends in Databases and Information Systems, 185, Springer, 2012, 99-108.

Problem 3: Active Learning and Inference

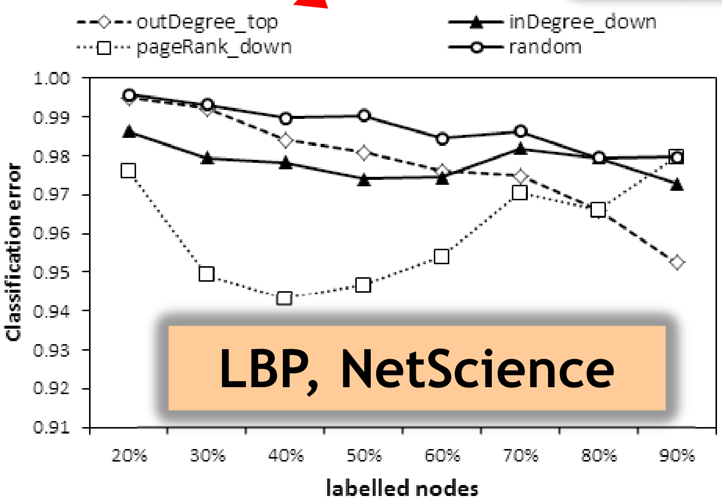
- Problem:
  - Identify **the sample of nodes** to acquire their labels
- Solution:
  - Methods for **active** inference in collective classication

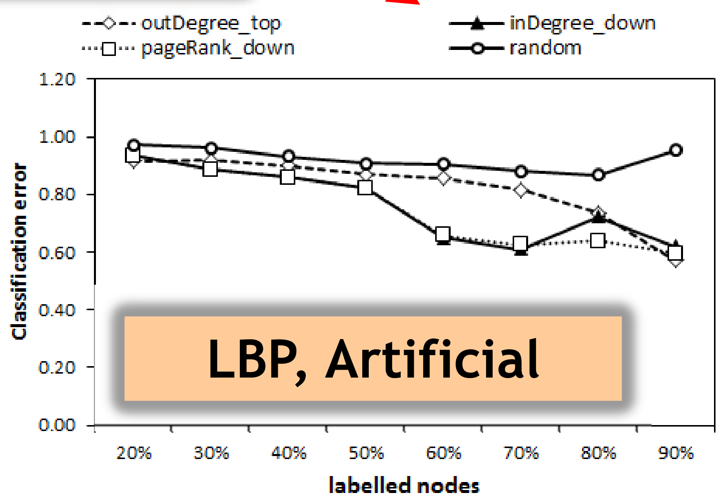


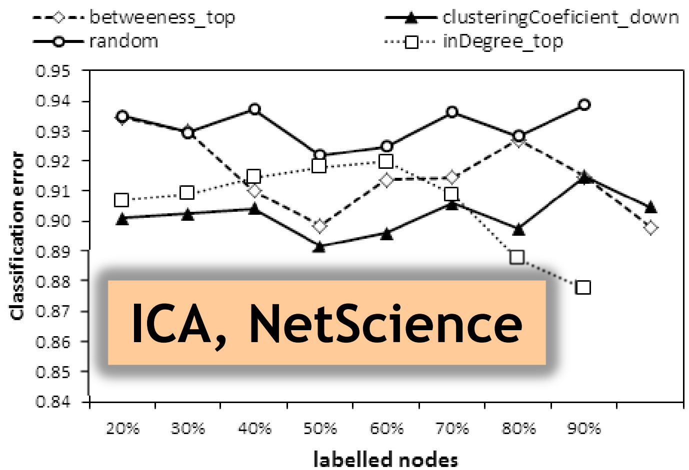
Wrocław University of Technology

Problem 3: Active Le nce

Node selection strategies (measures)







Kajdanowicz T., Michalski R., Musial-Gabrys K., Kazienko P.: *Learning in Unlabelled Networks - An Active Learning and Inference Approach*. **AI Communications**, 29(1), 2016, 123-148, also **ASONAM 2013**

Problem 4: Inference for Huge Networks

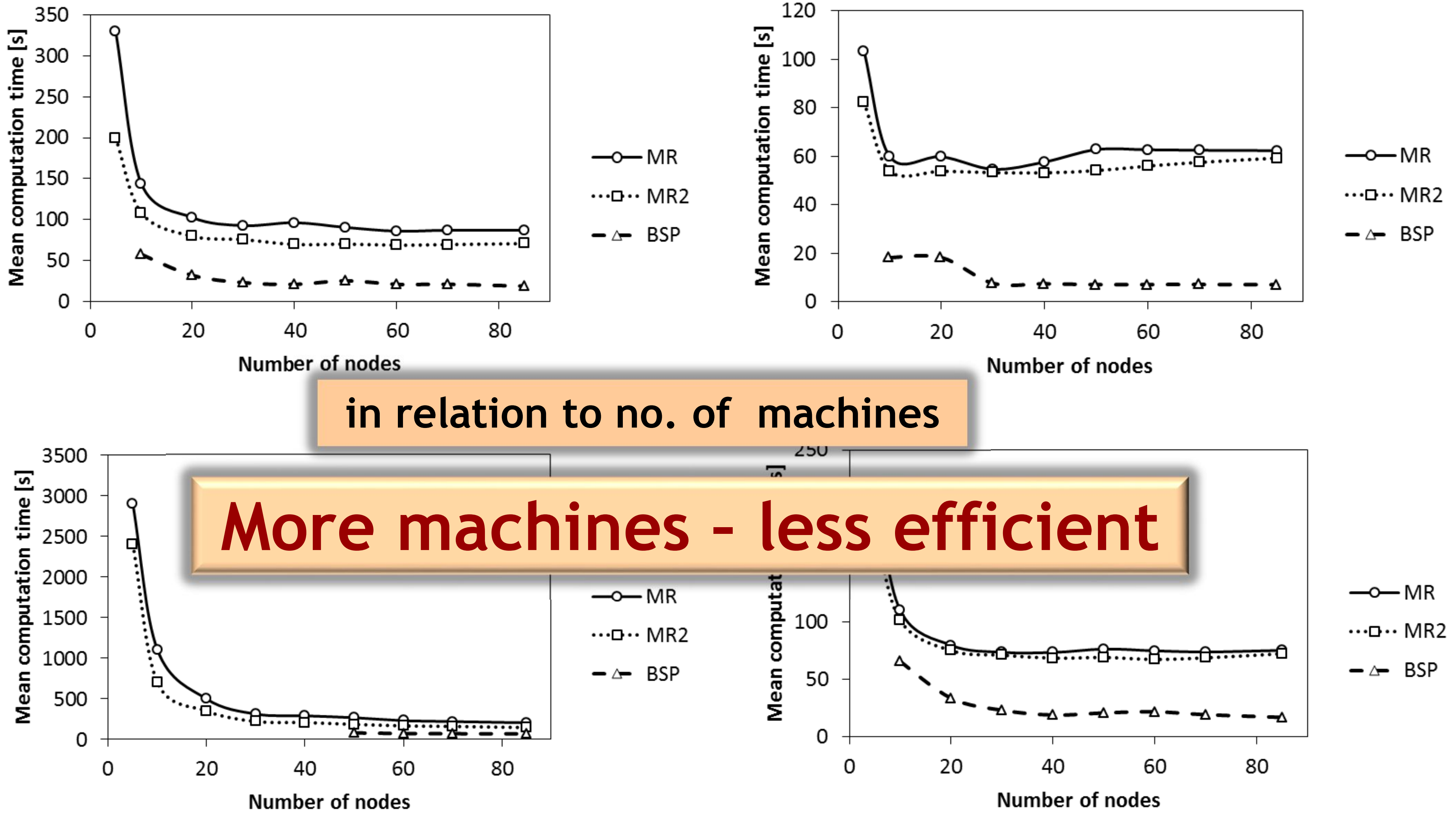
- Problem:
  - Big data needs parallel processing
  - **MapReduce** requires data **redistribution** at every iteration (Hadoop itself)

- Solution:
  - **Sparkling Graph** (WrUT)
  - **Bulk Synchronous Parallel (BSP)** (Giraph on Hado
  - **Directed Acyclic Graph** (TEZ on Hadoop)

Efficient Inference for Huge Networks: MapReduce vs. BSP

Wrocław University of Technology

Social Network Group
@ Wroclaw University of Technology



in relation to no. of machines

More machines - less efficient

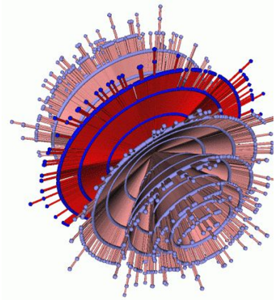
Kajdanowicz T., Kazienko P., Indyk W.: *Parallel Processing of Large Graphs*. **Future Generation Computer Systems**, 32, 2014, 324-337.

Wrocław University of Technology

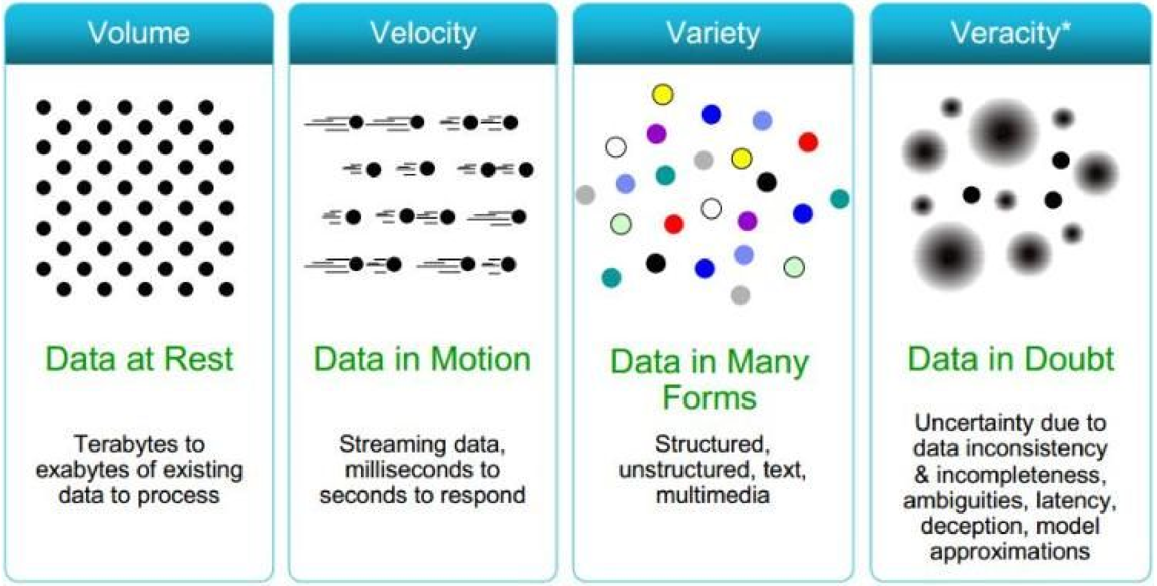
Other Problems in Collective Classication

Social Network Group
@ Wroclaw University of Technology

6. Learing for **temporal** / dynamic networks
7. **Incremental** relational learning
   – also for streams
8. Concept **drift** – model adaptation
9. Classification for **multimodal** networks
10. Lifted Graphical models [Kim15]



Big Data and Complex Networks



+ Complexity

Network processing is complex

„Over 80% of our data is from text/natural language/social media, unstructured, noisy, dynamic, unreliable, …, but ***interconnected****"*

Jiawei Han*, Wrocław, January 2016*

http://www.datasciencecentral.com/profiles/blogs/data-veracity

Wrocław University of Technology

Selected papers

1. Sen P., Namata G., Bilgic M., Getoor L., Galligher B., Eliassi-Rad T.: Collective Classification in Network Data. AI Magazine, 29 (3), 2008, 93-106.
2. Ahmed N., Neville J., Kompella R.: Network Sampling: From Static to Streaming Graphs. ACM Transactions on Knowledge Discovery from Data, 8, 2014.
3. Kimming A., Mihalkova L., Getoor L.: Lifted Graphical Models: A Survey. Machine Learning, 2015, 99 (1), 1-45.

4. Kajdanowicz T., Michalski R., Musial-Gabrys K., Kazienko P.: Learning in Unlabelled Networks - An Active Learning and Inference Approach. AI Communications, 29(1), 2016, 123-148, also ASONAM 2013
5. Kajdanowicz T.: Relational Classification Using Random Walks in Graphs. New Generation Computing, 2015, 33(4), 409-424
6. Kajdanowicz T., Kazienko P.: Collective Classification. in Encyclopedia of Social Network Analysis and Mining, Vol. 1. Springer, 2014, 144-156.
7. Kazienko P., Kajdanowicz T.: Collective Classification, Structural Features. in Encyclopedia of Social Network Analysis and Mining, Vol. 1. Springer, 2014, 156-168.
8. Indyk W., Kajdanowicz T., Kazienko P.: Relational Large Scale Multi-label Classification Method for Video Categorization. Multimedia Tools and Applications, 2013, 65 (1), 63-74.
9. Kazienko P., Kajdanowicz T.: Label-dependent Node Classification in the Network. Neurocomputing, 75 (1), 2012, 199-209
10. Michalski R., Kajdanowicz T., Bródka P., Kazienko P.: Seed Selection for Spread of Influence in Social Networks: Temporal vs. Static Approach. New Generation Computing, 32 (3-4), 2014, 213-235
11. Kajdanowicz T., Kazienko P., Indyk W.: Parallel Processing of Large Graphs. Future Generation Computer Systems, 32, 2014, 324-337.
12. Kajdanowicz T., Indyk W., Plamowski S., Kazienko P.: MapReduce Approach to Relational Influence Propagation in Complex Networks. Pattern Analysis and Applications, 17(4), 2014, 739-746
13. Kajdanowicz, T., Indyk, W., Kazienko, P., Kukuł J.: Comparison of the Efficiency of MapReduce and Bulk Synchronous Parallel Approaches to Large Network Processing. ICDM 2012, DaMNet 2012 - The Second IEEE ICDM Workshop on Data Mining in Networks, IEEE, 2012, 218-225.

Przesłanie końcowe

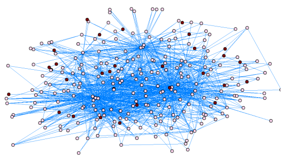
1. ‚Nauka o sieciach’ zainspirowała inne nauki…

2. Informatyka to matematyka XXI w.

3. Wiele danych *big data* ma charakter sieciowy

4. Relacje są wszędzie
   => przetwarzanie relacyjne to przyszłość

…więc myśl sieciowo

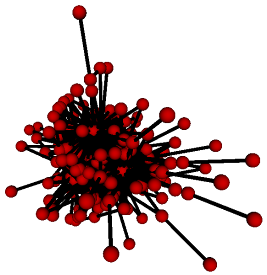


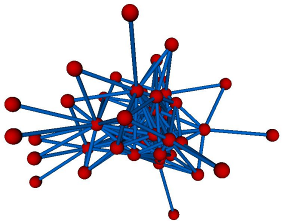
Pytanie na odchodne...

ENG: ***Data Science***

PL: ***Danologia***

?





engine

engine.pwr.edu.pl

European research centre of **N**etwork intelli**G**ence for **IN**novation **E**nhancement

Thank you for your attention!

Q & A ?

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 316097 [ENGINE].

Wrocław University of Technology