



NATIONAL RESEARCH  
UNIVERSITY

# Co-author Recommender System

Ilya Makarov

Oleg Bulanov

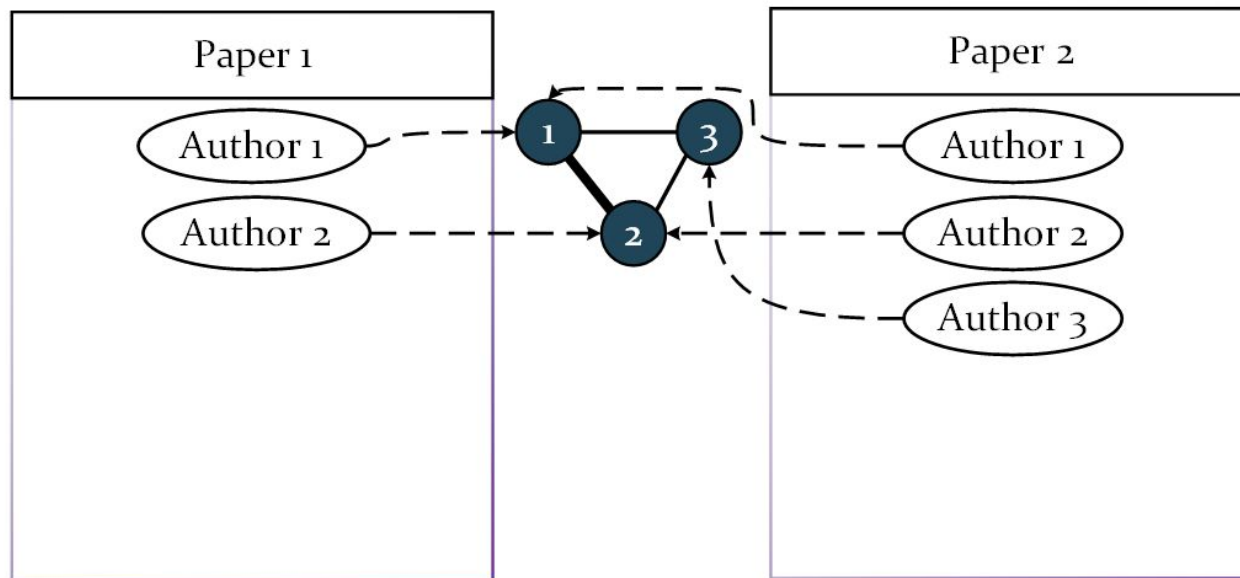
Leonid E. Zhukov

# Plan

1. Problem statement
2. Co-authorship building
3. Co-authorship network analysis
4. Recommender system building
5. Recommender system performance
6. Conclusion

# Co-authorship network

**Co-authorship network** – undirected graph which nodes represent authors and edges represent collaborations between authors



# Applicability

- University performance analysis
- Inefficient authors/departments detection
- Co-authors recommendations
- Best author of area searching

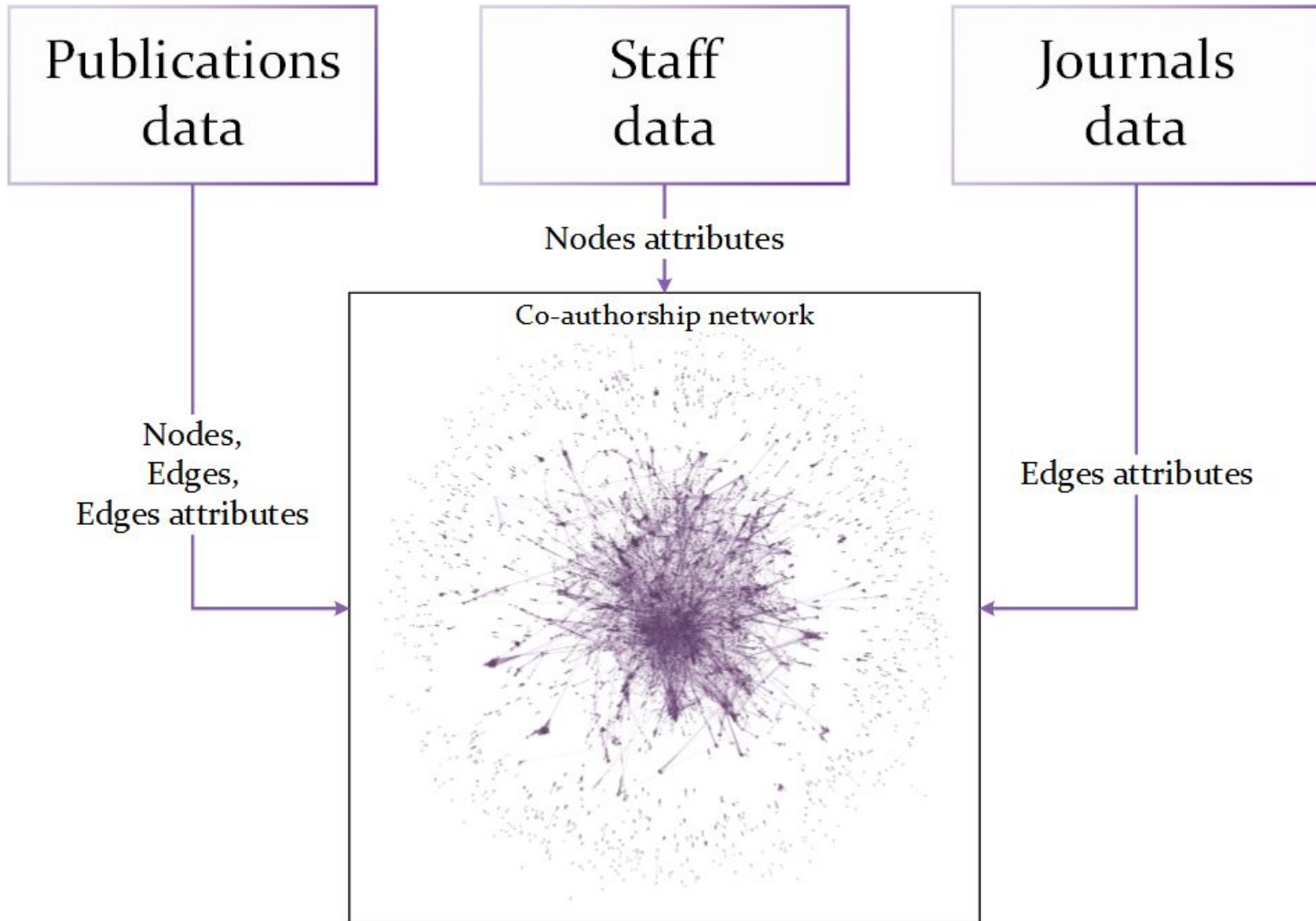
# Problem statement

**Goal:** analyze co-authorship of NRU HSE staff  
and build a recommender system

**Problems:**

- Build the co-authorship network
- Analyze the network
- Build the recommender system
- Analyze performance of the system

# Co-authorship network building



# Departments analysis

## We calculate:

- Number of authors
- Number of papers
- Number of papers in Q1, Q2

## For lecturers, senior lecturers, docents and professors:

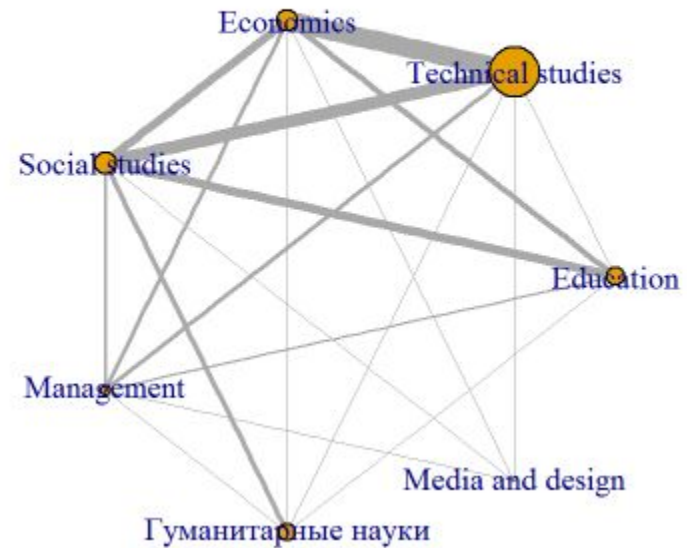
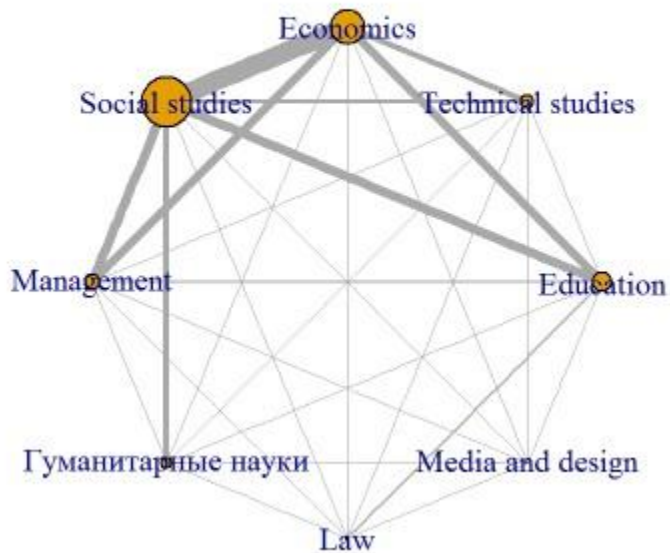
- Average number of papers
- Average marks 

$\min(15 \cdot N, 30)$	— lecturer, senior lecturer
$\min(10 \cdot N, 30)$	— docent
$\min(6 \cdot N, 30)$	— professor

# Areas interaction

Edge width - number of papers in collaboration  
Vert. size - number of papers in area

Edge width - number of papers in collaboration in Scopus  
Vert. size - number of papers in area in Scopus

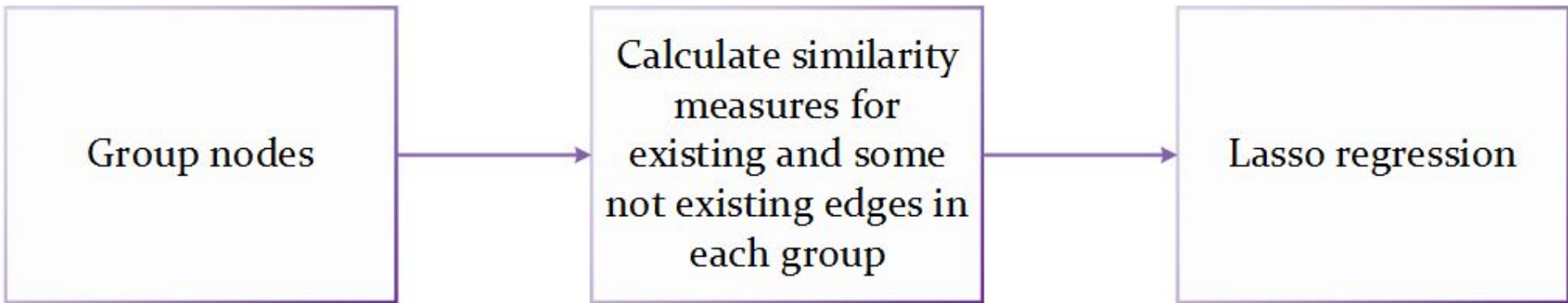




# Similarity measures

- **Number of common neighbors.**  $sim(v_i, v_j) = |N(v_i) \cap N(v_j)|$ , where  $N(v)$  — set of neighbors
- **Jacard coefficient**  $sim(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$
- **Adar coefficient**  $\sum_{v \in N(v_i) \cap N(v_j)} \frac{1}{\ln |N(v)|}$ .
- **Length of shortest path**
- **Norm similarity**  $sim(x, y) = \frac{1}{1 + ||x - y||}$ , where  $x, y$  characteristics of authors
- **Cosine similarity**  $sim(x, y) = \frac{(x, y)}{||x|| ||y||}$
- **Number of common journal areas**

# Recommender system learning scheme



# Group method

group := nodes of department

group := group  $\cup$  nodes of similar departments

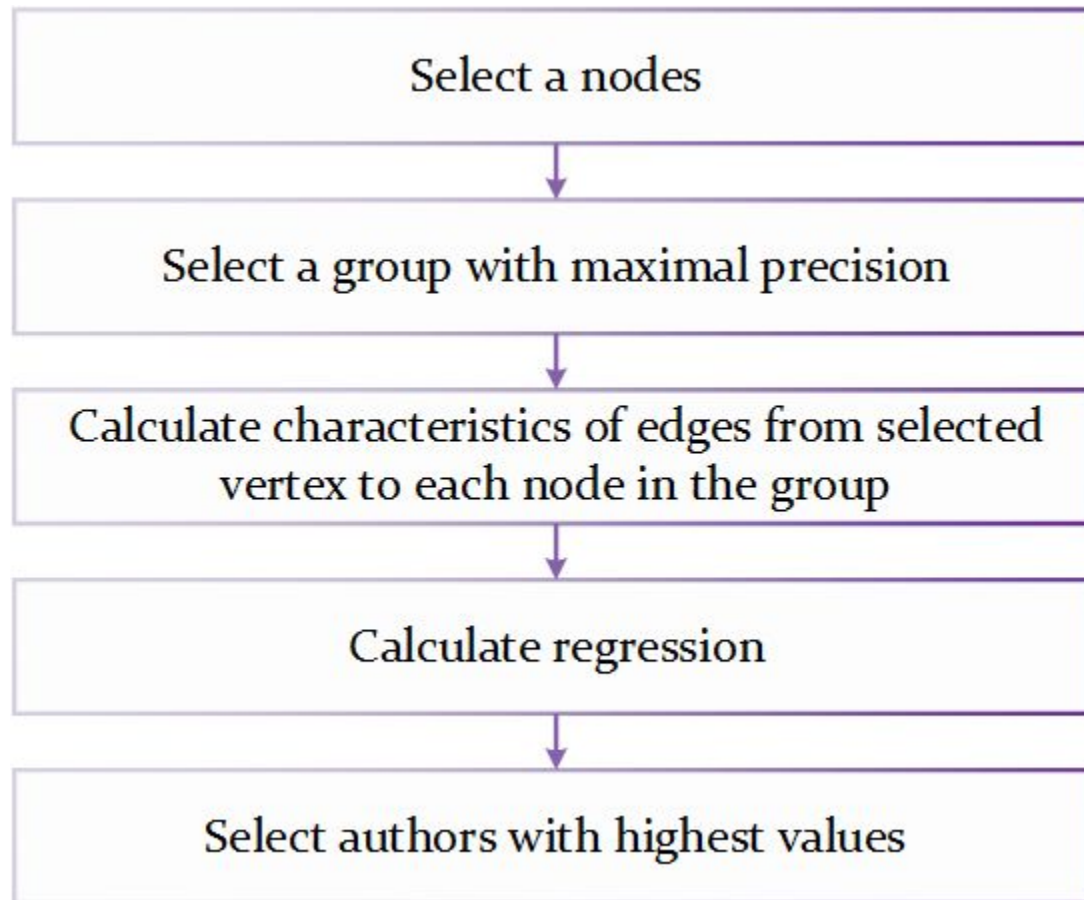
group := group  $\cup$  nodes of authors with similar publications

Remove not hse authors

group := group  $\cup$  nodes of clusters that contain nodes from group

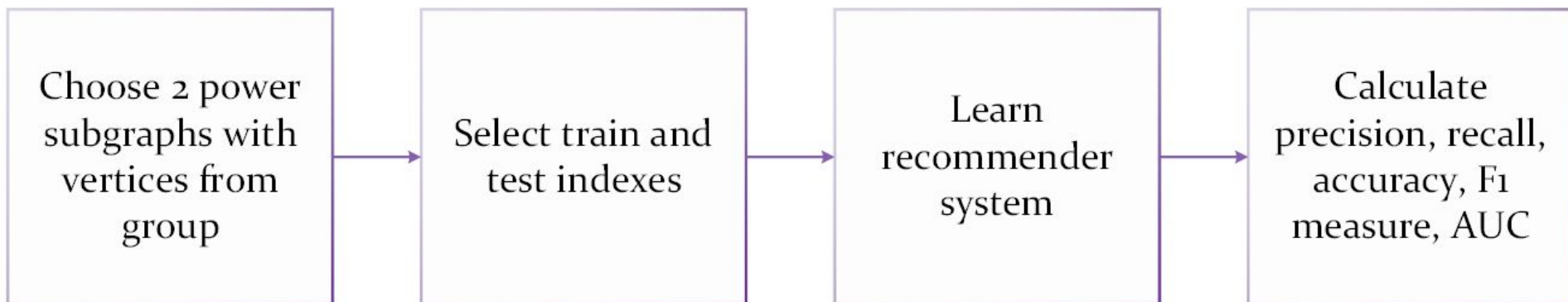
group := hse authors from group

# Making recommendation



# Testing recommender system

Power subgraph – subgraph with edges weight greater than some value.



# Recommender system performance

	Precision	Recall	Accuracy	F1 measure	AUC
Train	0,916	0,991	0,947	0,950	<b>0,991</b>
Test	<b>0,901</b>	0,868	0,873	0,870	0,924

# Conclusion

- Some patterns of departments efficiency were figured out
- Recommender system was built and showed nice performance

**THANK YOU FOR THE ATTENTION!**