

Scientometric indicators and collaboration network as a potential tool for *gift author* detection

June 9–11, 2016
Scotland, UK

T.S. Shinyaeva, Yu.Yu. Tarasevich

Astrakhan State University

June 10, 2016

Violations of the principles of publication ethics

- Plagiarism.

Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.

Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.

Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.

Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.
- Incorrect list of co-authors.

Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.
- Incorrect list of co-authors.



Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.
- Incorrect list of co-authors.



Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.
- Incorrect list of co-authors.



АНТИПЛАГИАТ



Violations of the principles of publication ethics

- Plagiarism.
- Multiple publication.
- 'Salami slicing'.
- Data fabrication and falsification.
- Incorrect list of co-authors.



Gift author

People who are listed as authors but who did not make a significant contribution to the research.

Hypotheses and Questions

Hypotheses

Hypotheses

- *Gift author* can be detected by analysis of objective data (co-authorship networks, scientometric indicators, etc.).

Hypotheses

- *Gift author* can be detected by analysis of objective data (co-authorship networks, scientometric indicators, etc.).
- This detection can be formalized and automated.

Hypotheses and Questions

Hypotheses

- *Gift author* can be detected by analysis of objective data (co-authorship networks, scientometric indicators, etc.).
- This detection can be formalized and automated.

Questions

Is it possible to develop quantitative scientometric criteria for determining risk groups? Namely,

Hypotheses and Questions

Hypotheses

- *Gift author* can be detected by analysis of objective data (co-authorship networks, scientometric indicators, etc.).
- This detection can be formalized and automated.

Questions

Is it possible to develop quantitative scientometric criteria for determining risk groups? Namely,

- dissertations in which all formal requirements are met, but the real scientific contribution of the applicant may be negligible;

Hypotheses and Questions

Hypotheses

- *Gift author* can be detected by analysis of objective data (co-authorship networks, scientometric indicators, etc.).
- This detection can be formalized and automated.

Questions

Is it possible to develop quantitative scientometric criteria for determining risk groups? Namely,

- dissertations in which all formal requirements are met, but the real scientific contribution of the applicant may be negligible;
- groups of authors which regularly published many articles with violations of the principles of publication ethics.

Open Data

Open Data

- We took in our consideration main publications of the Russian dissertations.

Open Data

- We took in our consideration main publications of the Russian dissertations.
- Open databases of dissertations (website <http://vak.ed.gov.ru/dis-list>, websites of organizations, where dissertations were defended).

Open Data

- We took in our consideration main publications of the Russian dissertations.
- Open databases of dissertations (website <http://vak.ed.gov.ru/dis-list>, websites of organizations, where dissertations were defended).
- CRIS of Astrakhan State University (<http://science.asu.edu.ru>)

Limitations

- In Russia, there is no complete database of publications related to dissertations.

Limitations

- In Russia, there is no complete database of publications related to dissertations.
- Main results of some thesis (defended before 2006) were published only in the conference papers, not in the journal articles.

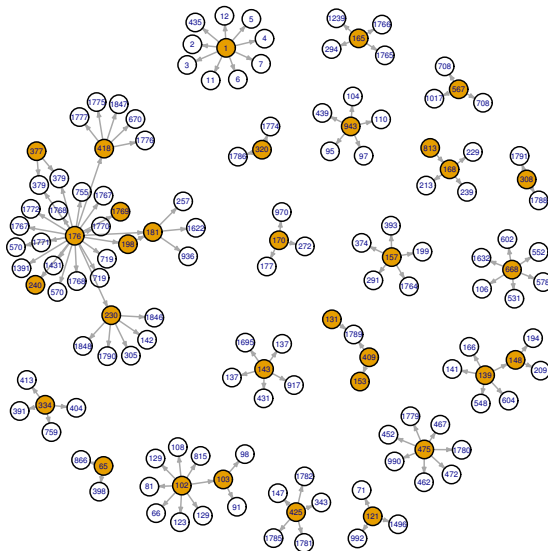
Limitations

- In Russia, there is no complete database of publications related to dissertations.
- Main results of some thesis (defended before 2006) were published only in the conference papers, not in the journal articles.
- Many old theses are not digitized yet.

- In Russia, there is no complete database of publications related to dissertations.
- Main results of some thesis (defended before 2006) were published only in the conference papers, not in the journal articles.
- Many old theses are not digitized yet.
- In Russia, a high fraction of publications in Humanities has no co-authors.

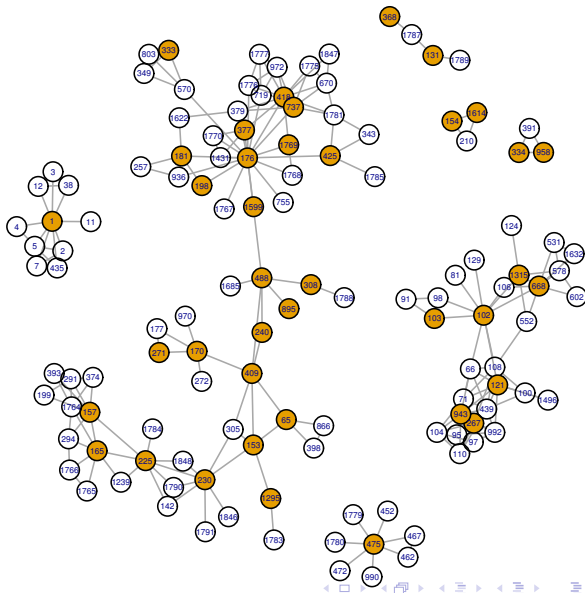
Relations between supervisors and PhD students.

143 dissertations

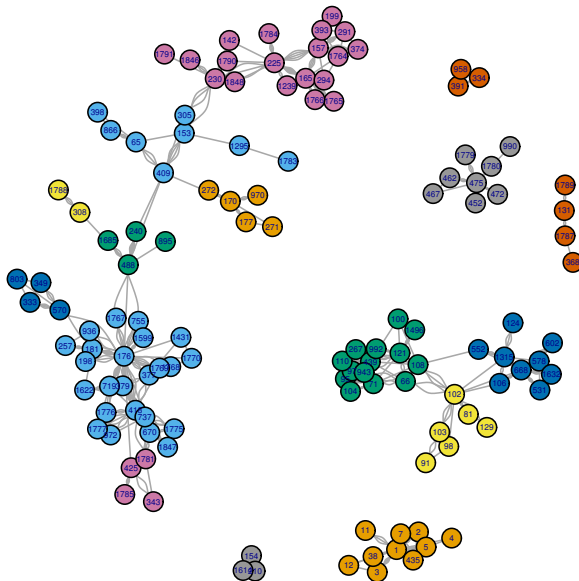


Network of co-authors

855 articles



Network of co-authors



Results

No.	N_{all}	$\langle NCA \rangle$	$\langle NP \rangle$	J_{ASU}	R_d	r	C	D
1	29	2.417	7.375	10.5%	0.745	-0.498	0.437	0.356
2	35	2.471	4.618	28.6%	0.01	-0.859	0.256	0.286
6	95	3.012	6.628	66.7%	0.255	0.083	0.006	0.6
16	30	3	6.7	67.8%	0.136	1	0	0.5
17	65	3.017	6.638	63.7%	0.314	-0.333	0.007	0.81
15	216	2.413	4.662	43.3%	0.011	-0.297	0.02	0.121
20	170	2.259	4.456	37%	0.012	-0.421	0.024	0.157

N_{all} is the total number of articles; $\langle NCA \rangle$ is the average number of articles co-authors; $\langle NP \rangle$ is the average length of articles (in pages); J_{ASU} is the number of articles in journals published by ASU; R_d is the average rating of dissertations; r is the assortativity coefficient; C is the network average clustering coefficient; D is the graph density.

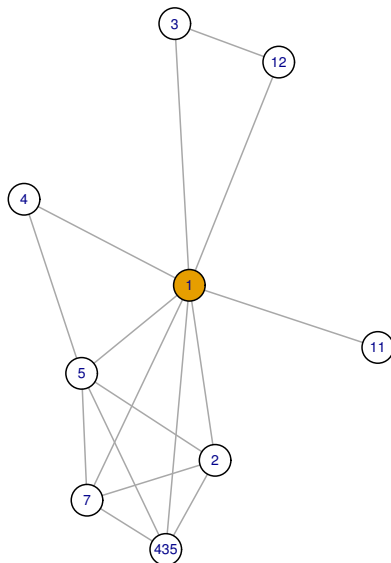
Results

Group No.1

9 dissertations

from 2005 to 2013

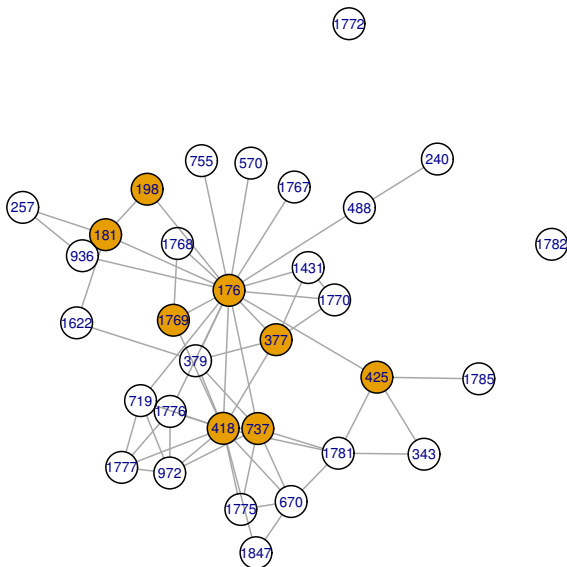
N_{all}	29
$\langle NCA \rangle$	2.417
$\langle NP \rangle$	7.375
J_{ASU}	10.5%
R_d	0.745
r	-0.498
D	0.356



Results

Group No.15
48 dissertations
from 1996 to 2015

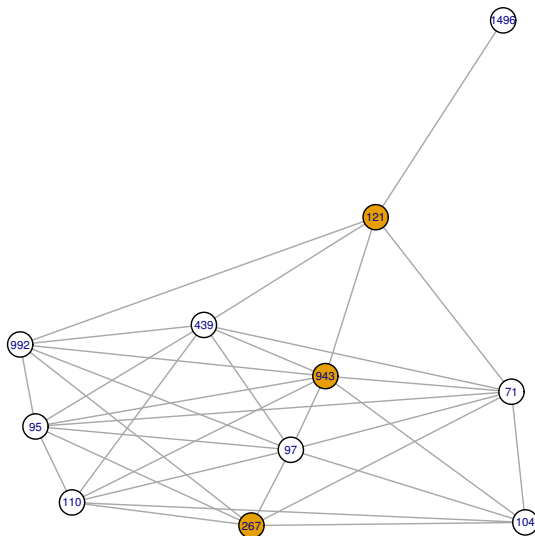
N_{all}	216
$\langle NCA \rangle$	2.413
$\langle NP \rangle$	4.662
J_{ASU}	43.3%
R_d	0.011
r	-0.297
D	0.121



Results

Group No.6
9 dissertations
from 2008 to 2014

N_{all}	95
$\langle NCA \rangle$	3.012
$\langle NP \rangle$	6.628
J_{ASU}	66.7%
R_d	0.255
r	0.083
D	0.6

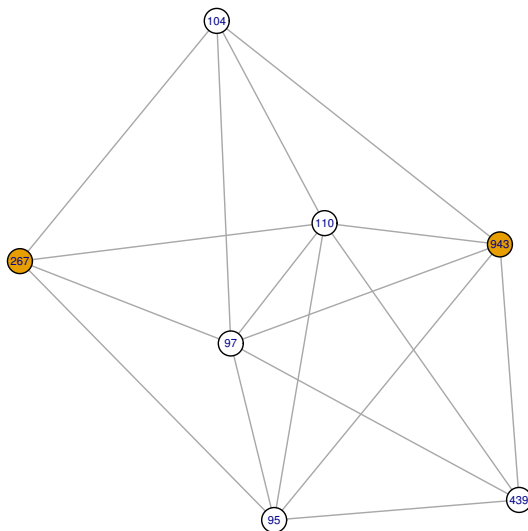


10.5 articles per dissertation!

Results

Group No.17
6 dissertations
from 2008 to 2012

N_{all}	65
$\langle NCA \rangle$	3.017
$\langle NP \rangle$	6.638
J_{ASU}	63.7%
R_d	0.314
r	-0.333
D	0.81



10.8 articles per dissertation!

Conclusions and Future Works

Future Works

Future Works

- Process more dissertations (dissertations from thesis factories which were determined by Dissernet).

Future Works

- Process more dissertations (dissertations from thesis factories which were determined by Dissernet).
- Consider other scientometric indicators, such as self-citations and citations by co-authors.

Future Works

- Process more dissertations (dissertations from thesis factories which were determined by Dissernet).
- Consider other scientometric indicators, such as self-citations and citations by co-authors.

Conclusions

Conclusions and Future Works

Future Works

- Process more dissertations (dissertations from thesis factories which were determined by Dissernet).
- Consider other scientometric indicators, such as self-citations and citations by co-authors.

Conclusions

- We hope that our research will give quantitative criteria for the detection of risk groups.

Thank you for your attention!

Rating

Dissertation's rating (R_d) and rating of dissertation's author (R_{da}) are calculated by the following formulas

$$R_d = \sum_{i=1}^n IF_i \frac{NP_i}{TNP_i}, \quad R_{da} = \sum_{i=1}^n IF_i \frac{NP_i}{TNP_i \cdot NCA_i},$$

where n is the number of publications concerning dissertation, IF_i is the journal impact factor in which i -th article was published, NP_i is the number of i -th article's pages, TNP_i is the typical article length (10), NCA_i is the number of co-authors of the i -th article.

When a journal has not impact factor, then similar indicator, i.e. Cites/Doc. (2 years), calculated by **Scimago** was used. Since a large number of journals published in Russia have neither impact factor (WoS) nor Cites/Doc. (Scimago), then $IF = 0.01$ for journals included in WoS or Scopus DB's, $IF = 0.005$ for journals included in Russian Science Index, $IF = 0.001$ for other journals.

Assortativity coefficient r is the Pearson correlation coefficient of degree between pairs of linked nodes.

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2},$$

where j_i , k_i are the degrees of the vertices at the ends of the i th edge, with $i = 1 \dots M$

Network average clustering coefficient

Network average clustering coefficient is measured as the average of the local clustering coefficients of all the vertices.

$$C = \frac{1}{n} \sum_{i=1}^n c_i,$$

where c_i is the local clustering coefficient.

$$c_i = \frac{2n_i}{k_i(k_i - 1)},$$

n_i is the number of links between neighbors of the i th node, k_i is the degree of the i th node. If $k_i \leq 1$ then $c_i = 0$.

Graph density

The density of a graph $G = (V, E)$ measures how many edges are in set E compared to the maximum possible number of edges between vertices in set V .

$$D = \frac{2m}{n(n-1)},$$

where m is the number of edges, n is the number of nodes.