

# Individual Publication Strategies and Collaboration in Authorship

## Analysis of Co-authorship Network from a Large University

Michał Bojanowski  
[m.bojanowski@uw.edu.pl](mailto:m.bojanowski@uw.edu.pl)  
[www.bojanorama.pl](http://www.bojanorama.pl)

ICM, University of Warsaw

Sunbelt XXXIV, St Pete Beach, FL, USA  
February 20, 2014

# Introduction

## Goal

Developing an understanding of collaboration and competition in science

Scientists collaborate and compete for

- Primacy in discovery
- Funding
- Attention

Scientific publications are

- Main medium of scientific communication (articles, books, etc.)
- Manifestations of collaboration
- Manifestations of authority/hierarchy relations

# Outline

1 Introduction

**2 Puzzle**

3 Data considerations

4 Co-authorship at University of Warsaw

5 Conclusions

# Collaboration desirable?

Collaboration in knowledge production (innovation) seems **globally** desirable because:

- Facilitates exchange of ideas
  - “Ideas [should] have sex!” (Ridley 2010)
  - crossover in evolution
- There are effects of scale (?)
  
- There is no anti-trust law in scientific research!
- So why there is no big global research commune?

# Collaboration individually rational?

Possible explanations:

- Are there communication / information barriers?
  - market frictions?
  - probably not in the Internet era
- Cooperation failure?
  - Trust problems
  - Risk of free-riding
  - Reputation
- Transaction costs?
  - It is difficult/costly to manage large research teams

Can we have a better behavioral understanding of knowledge production?

# Outline

- 1 Introduction
- 2 Puzzle
- 3 Data considerations**
- 4 Co-authorship at University of Warsaw
- 5 Conclusions

# Data sources

Main data source: bibliographic databases

**journal-based** publications enter the database because they appeared in specific journals.

- Web of Science, Scopus, thematic databases (e.g. DBLP), etc.
- “tie-driven”

**affiliation-based** publications enter the database because they were (co-)authored by scientists affiliated with specific institutions

- Institutional bibliographic systems (and their aggregators)
- “node-driven”

# Network boundary problem

## Journal-based sources

- can cover specific disciplines
- usually miss books etc.
- miss non-English publications

## Affiliation-based sources

- give fairly complete account for a specific population of scientists
- usually do not provide good information on non-affiliated coauthors

## Conjecture

To study collaboration and competition we need to focus on specific population of individuals, therefore affiliation-based sources seem more appropriate.



# Outline

- 1 Introduction
- 2 Puzzle
- 3 Data considerations
- 4 Co-authorship at University of Warsaw**
- 5 Conclusions

# Questions and data

## Research questions:

- What the are propensities to collaborate in different departments?
- To what extent can we observe closure in collaboration network?

## Data:

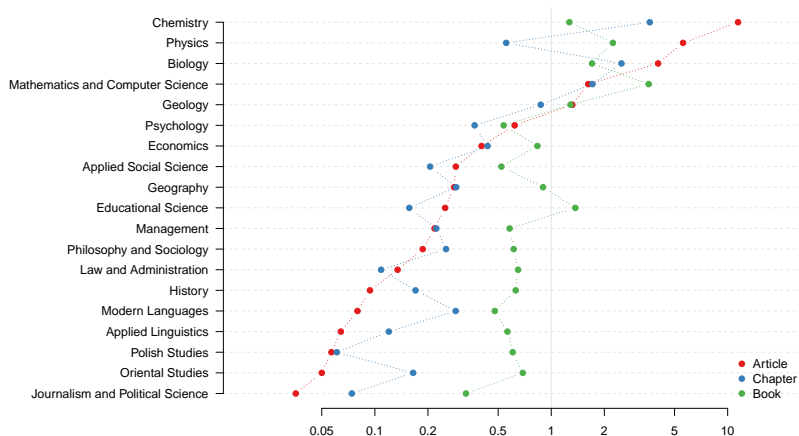
- Polish Scholarly Bibliography (PBN, <http://pbn.nauka.gov.pl>)
- Aggregate of repositories of Polish scientific institutions
- Currently, PBN system (and underlying data) *in statu nascendi*...
- Here data only on University of Warsaw (UW)

# UW in PBN: some summaries

- ~ 60000 works published in 2000–2011
  - Work = book, article, or chapter in edited volume
- ~ 22000 unique authors
- ~ 5300 affiliated with 19 departments of UW

# Collaborativeness

Odds-ratios whether publication has more than one author

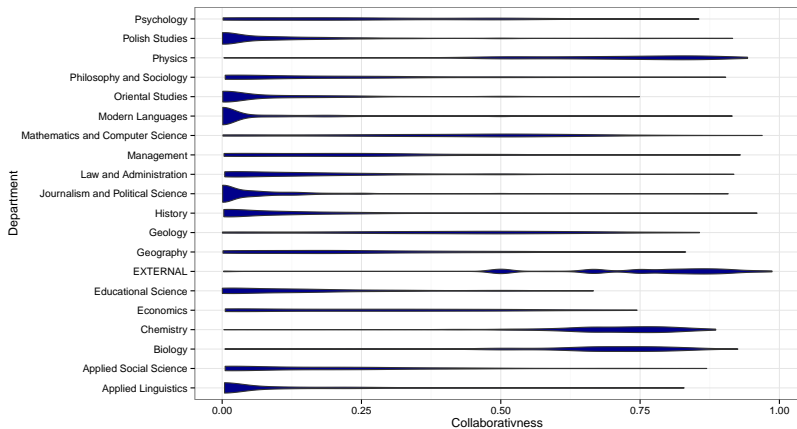


## Observations:

- Coauthorship more likely in natural sciences
- In social sciences and humanities books are more often co-authored than articles and chapters
- It is the other way around in Chemistry, Physics, and Biology

# Collaborativeness: index distributions

See Batagelj & Cernišek (2013)

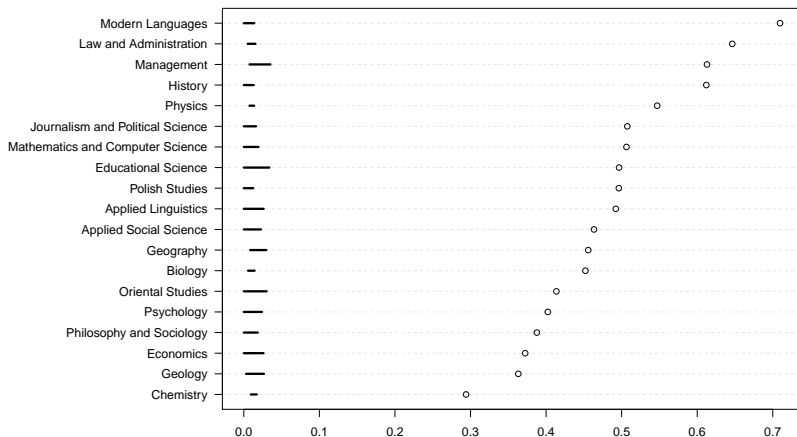


## Observations:

- In Chemistry and Biology almost all authors collaborate a lot
- In Physics, Geology, Mathematics/Computer science, there is much more variability
- In other disciplines collaboration is rather rare

# Transitivity within departments

CUG tests (cond. on dyad census)





# Outline

- 1 Introduction
- 2 Puzzle
- 3 Data considerations
- 4 Co-authorship at University of Warsaw
- 5 Conclusions**

# Conclusions

## In general

- Studies on collaboration in science (in publications) should be rather based on data from institutional repositories.

## Preliminary results on collaboration at UW:

- Big interdisciplinary differences
- More collaboration in natural science
- More collaboration not necessarily translates into more transitive networks (closure).

## Plans:

- Theory!
- Departmental homophily
- Dynamic analysis: e.g. “who is going to be your next collaboration partner?”