



# Modes of Collaboration in Modern Science

Staša Milojević

Indiana University

School of Library and Information Science

[smilojev@indiana.edu](mailto:smilojev@indiana.edu)



# Introduction

- Collaboration and coauthorship
- Collaboration network
  - Collaborators = all coauthors over some period
- Collaboration behavior of “stars” vs. collaboration behavior of **most** researchers
- An emerging interdisciplinary field of nanoscience as a case study



# Network studies of collaboration

- Network topologies
  - Small world
  - Scale-free
- Preferential attachment as an underlying process of network evolution
  - Especially for scale-free (“power law”) degree distribution



# Preferential attachment in collaboration networks

- Of new researchers to researchers who are already in the network
  - “New researchers more likely to connect to those who already have many collaborators”
- Among the existing researchers who have not collaborated previously
  - “New collaborations more likely among researchers with more mutual (common) collaborators” (Newman)
  - “New collaborations more likely among researchers with many collaborators altogether” (Barabasi)



- NanoBank, a digital library of:
  - 580,710 articles written by
  - 466,602 different authors published in
  - 8,300 journals from 1970 to 2006
- Articles have been selected from *ISI Science Citation Index*, *Social Science Citation Index*, and *Arts & Humanities Citation Index* using two methods:
  - search for 379 predefined terms
  - probabilistic procedure for the automatic identification of additional terms



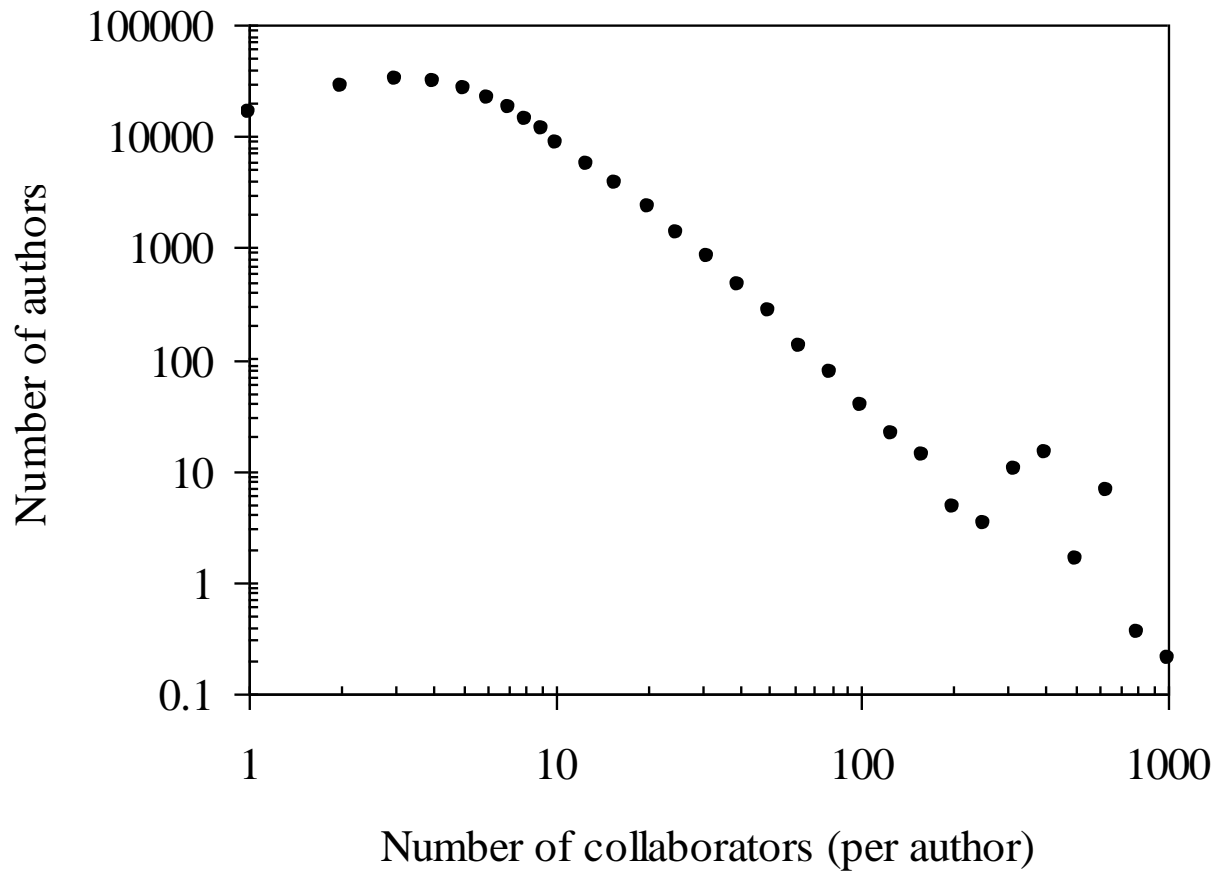
# Number of journals, articles and authors in NanoBank

<b>Number of</b>	<b>Total</b>	<b>1995-1999</b>	<b>2000-2004</b>
Journals	8,300	4,255	4,792
Articles	580,710	172,583	270,135
Authors	466,603	208,191	294,456

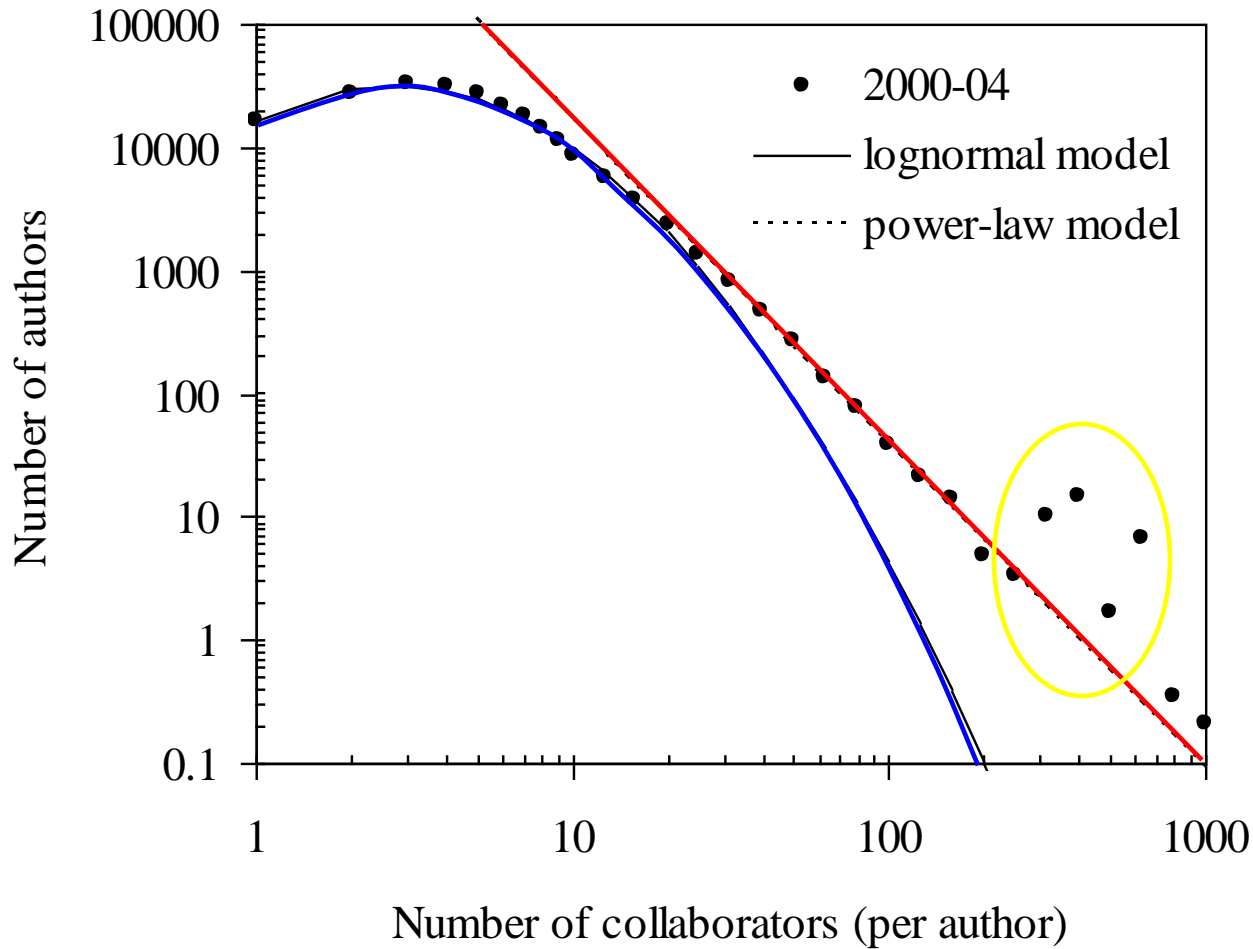
All analysis for 2000-2004 (5-yr period)

# Collaborator distribution

= node degree  
distribution



# Collaborator distribution

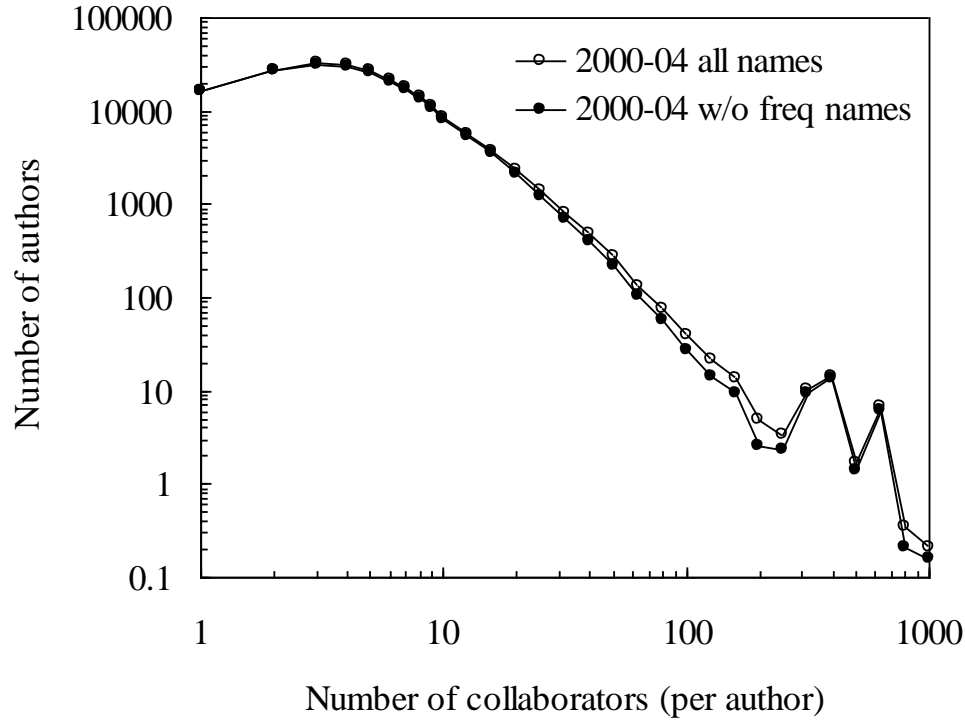


The log-normal and linear models fitted to the distribution of the number of collaborators of a given author (based on NanoBank database and restricted to the period 2000-04). Because of the logarithmic binning the number of authors (actually a probability) can have values less than one.



# Nature of “anomalous peak”

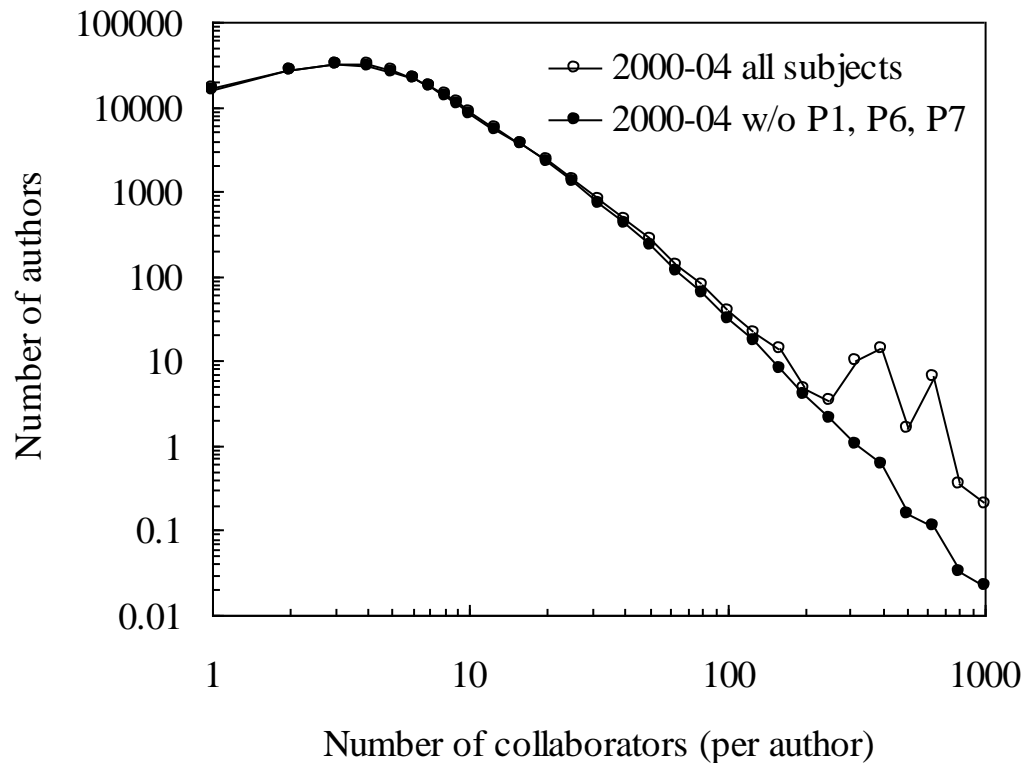
Is the “anomalous peak” the artifact of inability to distinguish different authors due to frequently occurring last names?



The distribution of the number of collaborators with 100 most frequent names removed.

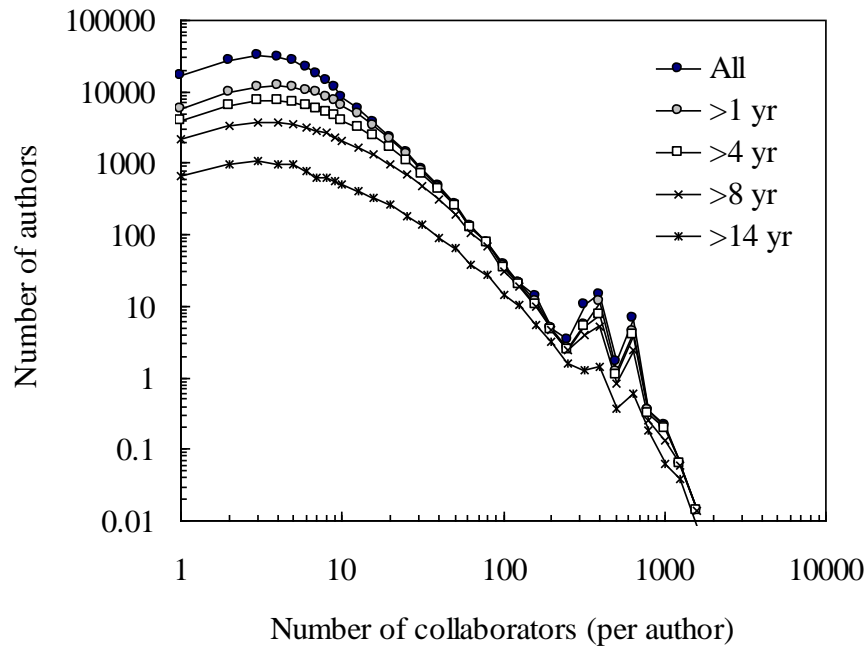
# Nature of “anomalous peak”

Is the “anomalous peak” the effect of hyperauthorship practices?

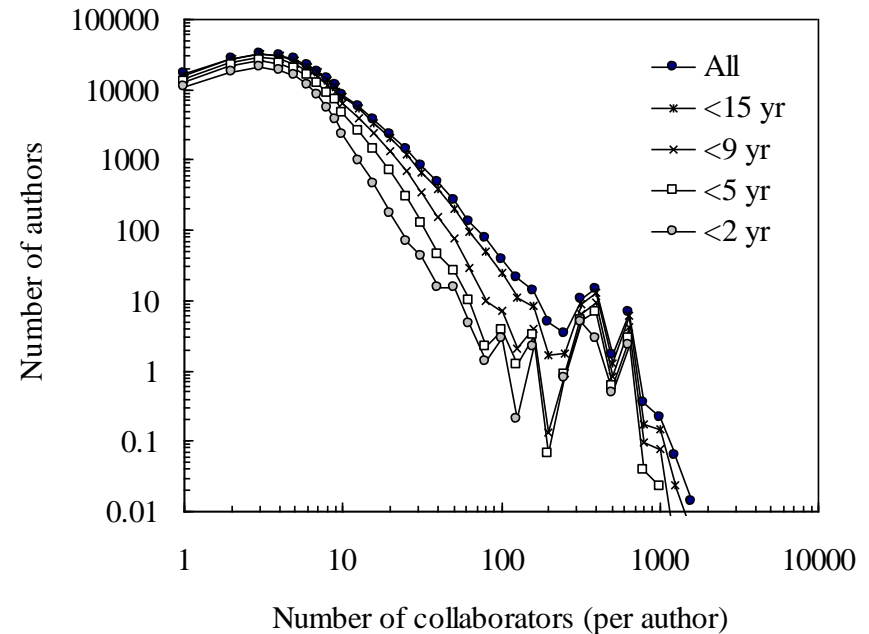


The distribution of the number of collaborators of a given author with three subfields (Multidisciplinary Physics (P1), Physics of Particles and Fluids (P6), Nuclear Physics (P7)) removed.

# Nature of the log-normal “hook”

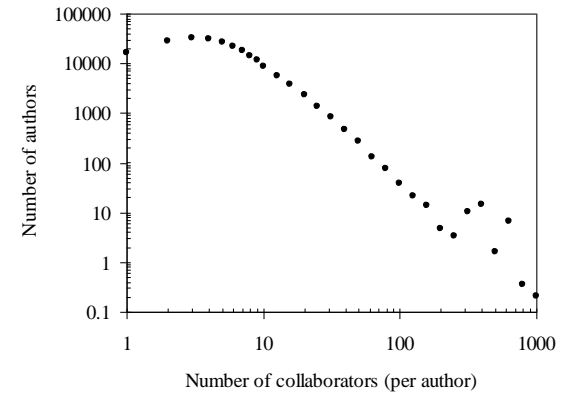
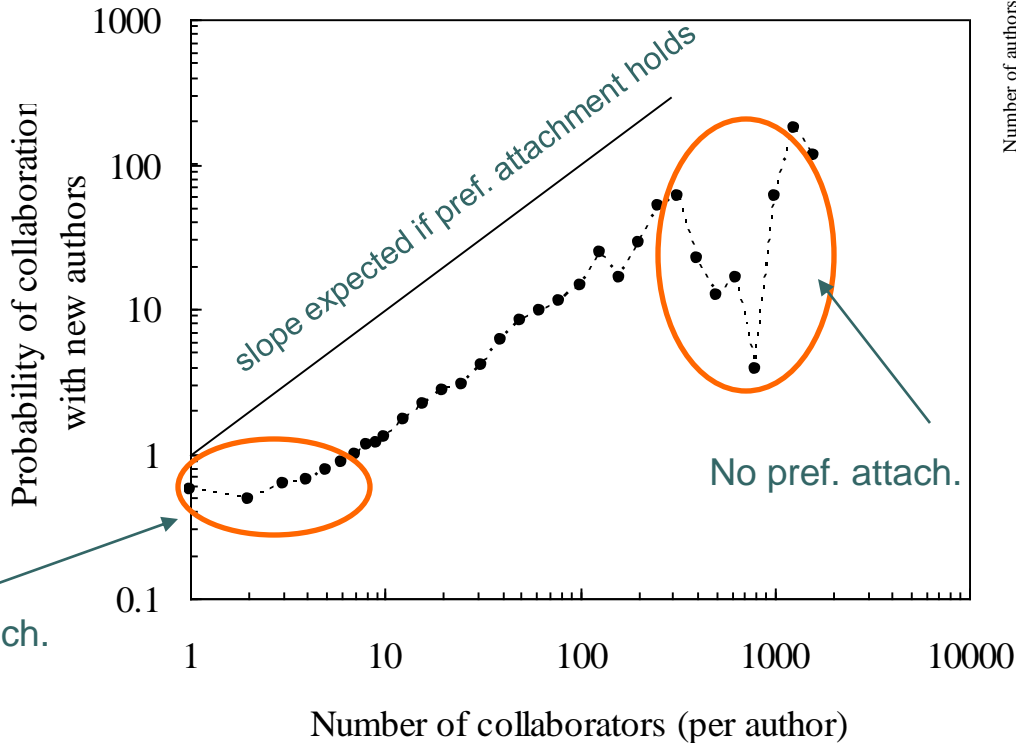


The effect on the collaborator distribution of the gradual removal of authors who have spent less than a certain time in the field, i.e., leaving those with more than 1, 4, 8 and 14 years spent in the field.



The effect on the collaborator distribution of the gradual removal of authors who have spent above certain time in the field (for period 2000-2004).

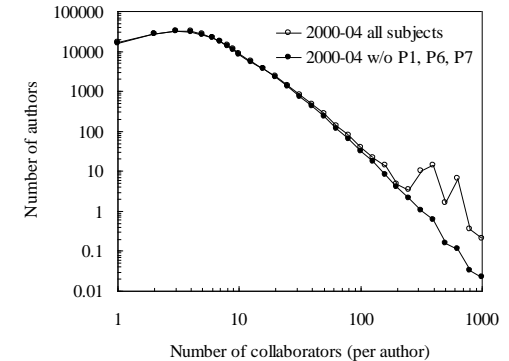
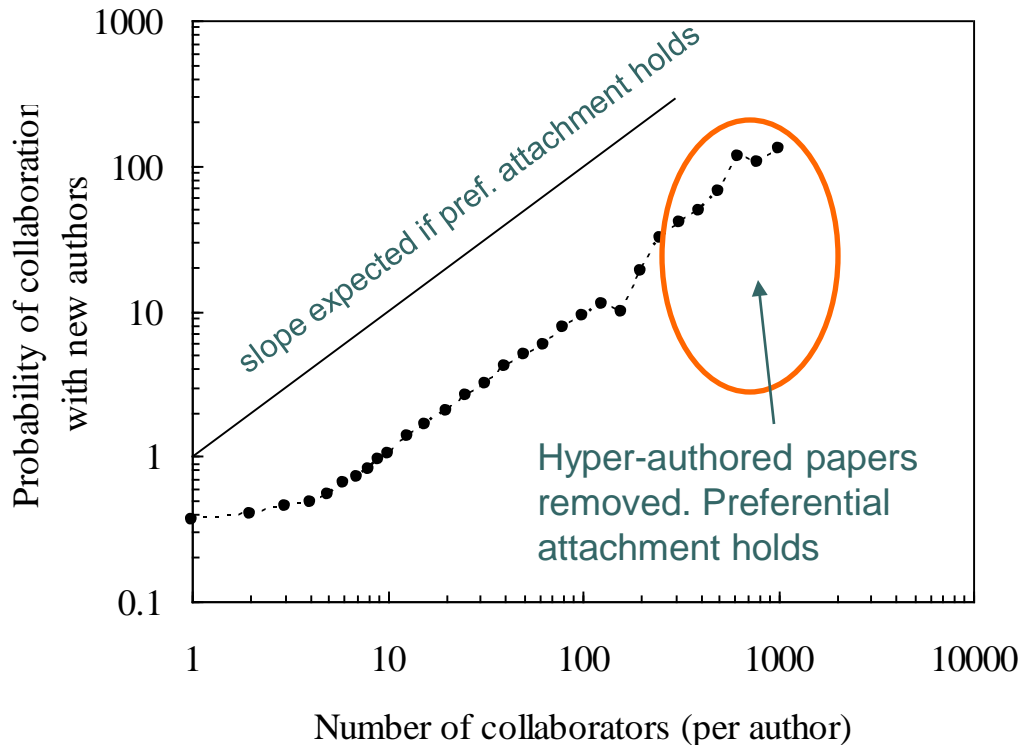
# Preferential attachment of new researchers to the existing researchers



“new researchers are more likely to connect to those who already have many collaborators”

Relative probability of collaboration with new authors (average number of new collaborators) as a function of the number of existing collaborators in log-log plot. The solid line has a slope of unity and represents linear preferential attachment.

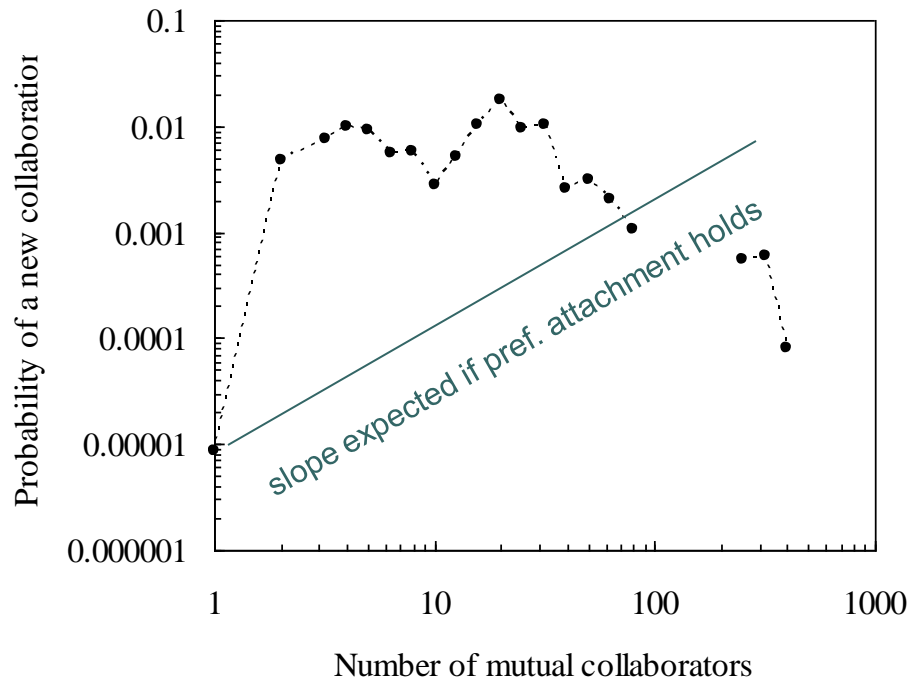
# Preferential attachment of new researchers to the existing researchers



“new researchers are more likely to connect to those who already have many collaborators”

Relative probability of collaboration with new authors (average number of new collaborators) as a function of the number of existing collaborators in log-log plot (now with papers with more than 200 authors excluded). The solid line has a slope of unity and represents linear preferential attachment.

# Preferential attachment among the existing researchers who have not collaborated previously

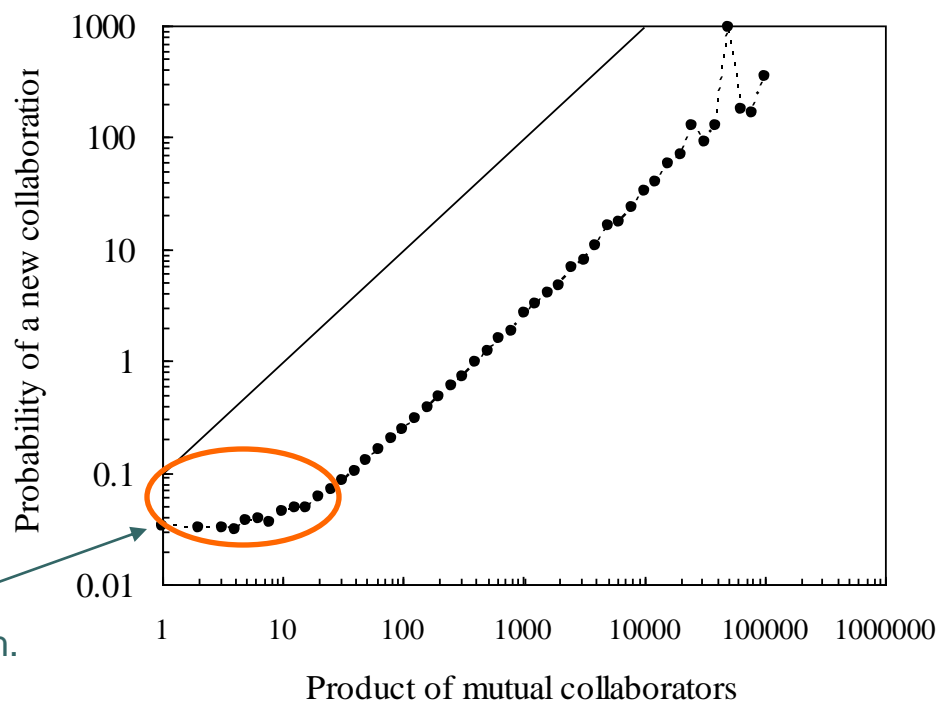


Newman formulation:

“new collaborations are more likely among researchers with more mutual (common) collaborators”

Probability of new collaboration among existing authors as a function of the number of mutual collaborators. Preferential attachment is **not** present.

# Preferential attachment among the existing researchers who have not collaborated previously

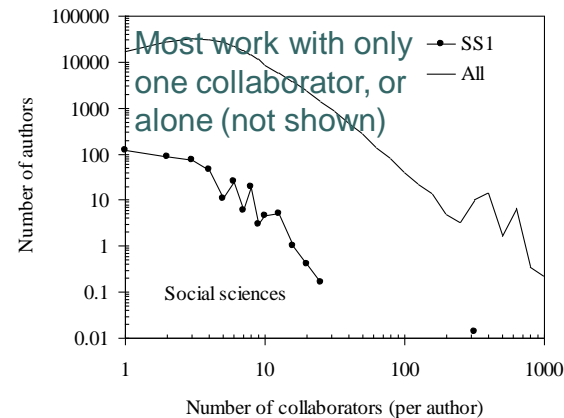
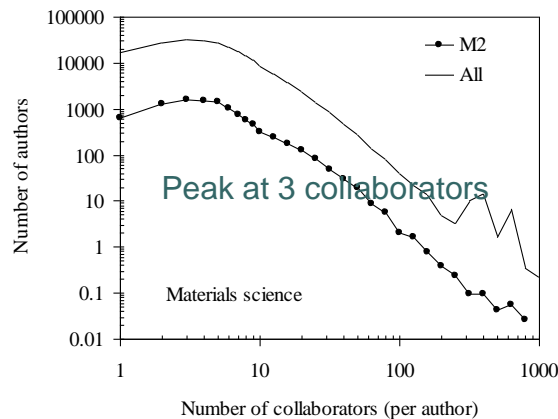
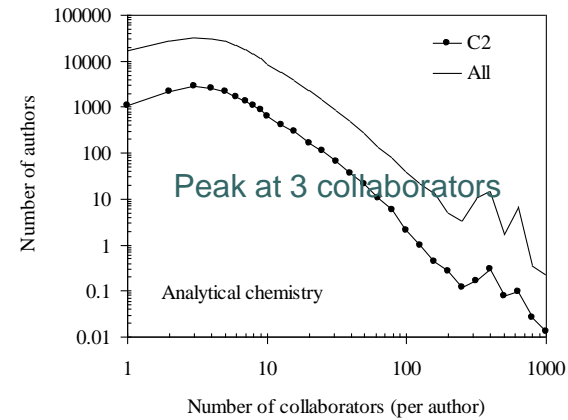
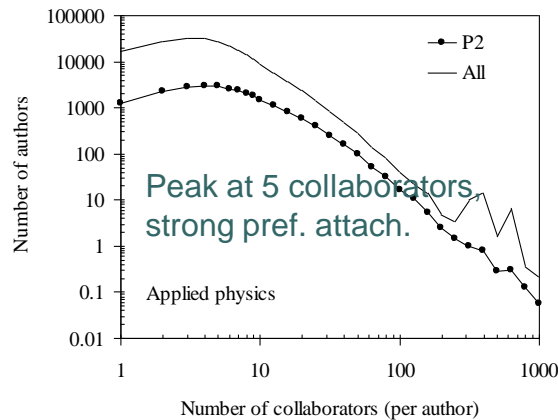


Barabasi formulation:

“new collaborations more likely among researchers with many collaborators altogether (product of the number of collaborators of both authors)”

Probability of a new collaboration among the existing authors as a function of the product of the number of collaborators of both authors (with papers with more than 200 authors excluded). Linear preferential attachment is shown as a solid line.

# Collaboration distribution is field dependent

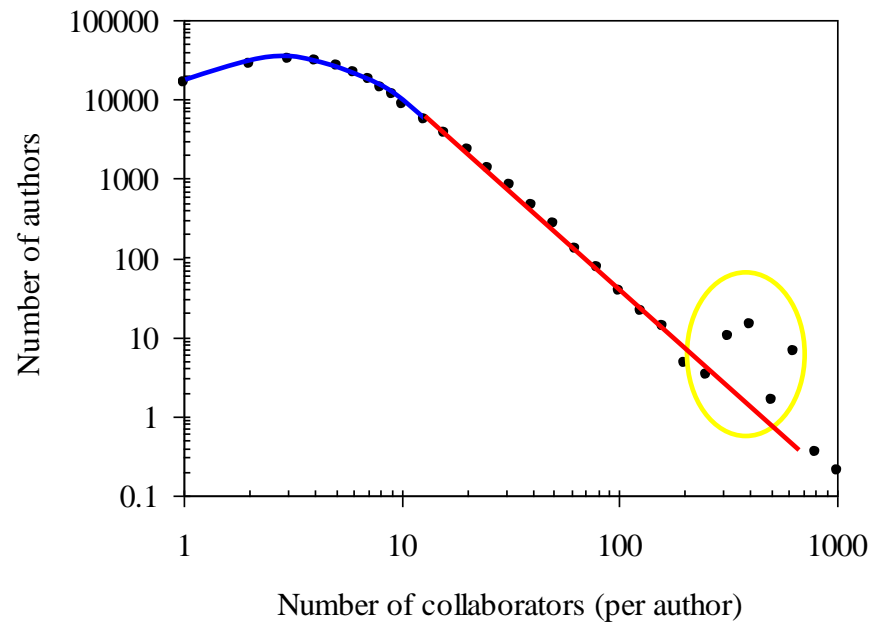


The distributions of the number of collaborators for 6 subfields (dotted line, Condensed Matter & Applied Physics (P2), Analytical Chemistry (C2), Multidisciplinary Materials Science (M2), Social Sciences (SS1), and the full distribution (line) for the period 2000-04. What is shown is the number of collaborators (from any subfield) of an author belonging to a given subfield.



# 3 types of collaboration modes

- The log-normal “hook”
- The power law tail
- The anomalous peak





# The log-normal “hook”

- Great majority (88%) of collaborating researchers
- Preferential attachment (=power law) does not operate
- Not merely the result of the amount of time spent in the field
- Reflects the “characteristic” (“optimal”) number of collaborators, which is not a power-law, but a variation on a normal distribution (has a peak)
- In nanotechnology the majority of scientists tend to coauthor papers with a small number of collaborators, rather than work alone or in very large groups
- The extent and peak of the “hook” is different from subfield to subfield



# Power law tail

- A reflection of the existence of scientific stars who attract a disproportionate number of collaborators through the process of preferential attachment
- New scientists are more likely to collaborate with scientists who already have a large number of collaborators
- The probability of new collaboration among the existing researchers who have not collaborated before does not depend on the number of mutual collaborators, but only on the total number of collaborators that each researcher has



# Coauthorship networks and hyperauthorship

- Considering hyperauthors as collaborators may be problematic
- Authors who work in very large teams, following hyperauthorship practices, do not benefit from preferential attachment



# Thank you!

Staša Milojević

Indiana University, Bloomington

SLIS

1320 E. 10th Street, LI Room 019

Bloomington, IN 47405-3907, USA

<http://www.slis.indiana.edu/faculty/smilojev/>  
[smilojev@indiana.edu](mailto:smilojev@indiana.edu)

**Acknowledgement:**

Certain data included herein are derived from Nanobank (Lynne G. Zucker and Michael R. Darby, Nanobank Data Description, release 1.0 (beta-test), Los Angeles, CA: UCLA Center for International Science, Technology, and Cultural Policy and Nanobank, Jan 17, 2007.) and from the Science Citation Index Expanded, Social Sciences Citation Index, and Arts & Humanities Citation Index of the Institute for Scientific Information®, Inc. (ISI®), Philadelphia, Pennsylvania, USA: © Copyright Institute for Scientific Information®, Inc. 2006. All rights reserved.