



With a little help from your friends: The impact of social networks on publication success

Frank Schweitzer

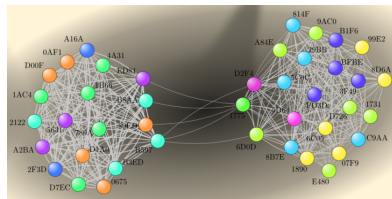
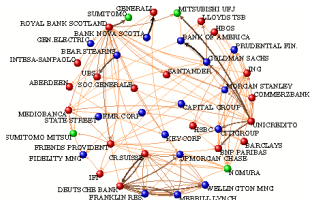
In collaboration with: *G. Casiraghi, V. Nanumyan, E. Sarigöl, I. Scholtes*

Chair of Systems Design at ETH Zurich

■ Main Research Areas

■ Economic Networks & Social Organizations

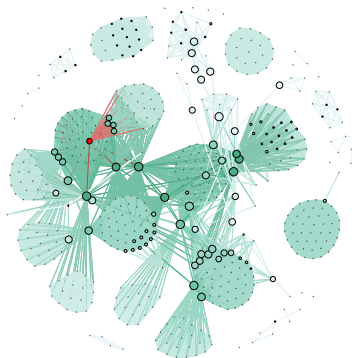
- e.g. ownership networks, R&D networks, financial networks, ...
- e.g. online communities, OSS projects, animal societies, ...



■ Methodological Approach: Data Driven Modeling

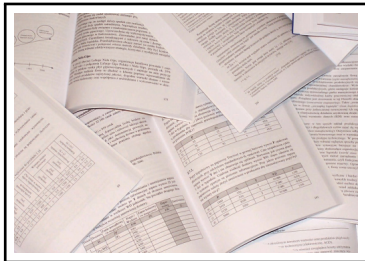
- **economic databases:** ORBIS, Bloomberg, patent databases
- **online data:** user interaction, communication records, blogs

Collaboration of scientists



- **agent:** individual scientist, **network:** co-authorship network
- **data:**
 - APS (1895-2004): 226.724 authors, 1,567.084 collaborations
 - MSAS (1996-2008): 160.891 authors, 5,324.330 collaborations

Science = Publications?

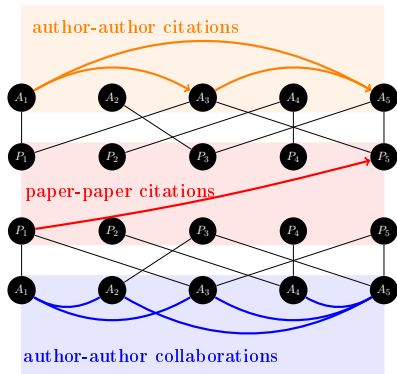


Follow-up mistakes:

- if you believe in:
Science \equiv Publications
you also believe in:
Success \equiv Citations
- common misunderstanding:
IF \equiv importance
 - IF = size of community
 - 1st order strategy to increase IF:
publish *less*

Get your paper into the top 10% in your field
(measured by citations)

Multi-layer network: Scientific collaborations



3 different projections

1 Collaboration network

- relations between *authors*
- undirected, evolves in time

2 Citation network

- relation between *papers*
- *directed*, *temporal order* (!)

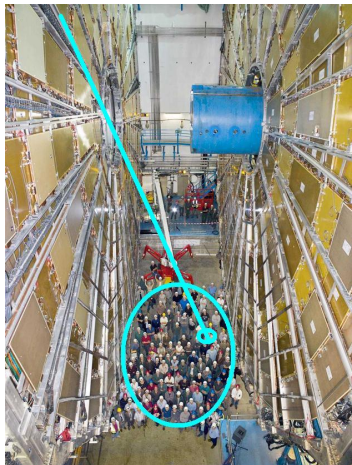
3 Author-author citations

- do we cite papers or authors?

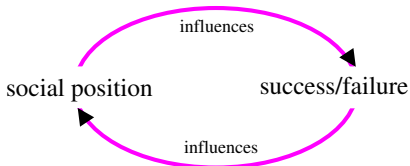
Checkout our multiplex network visualization tool!

multinets.io

Does your social position really matter?



- Quality matters ... but
 - paper should become known
 - *authors* have to become known



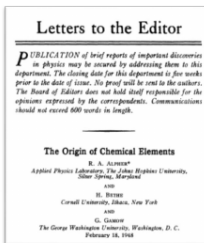
Experiment

Can social position at year t predict success of a paper at year $t + 5$?

Dynamic collaboration network



α
Ralph **Alpher**

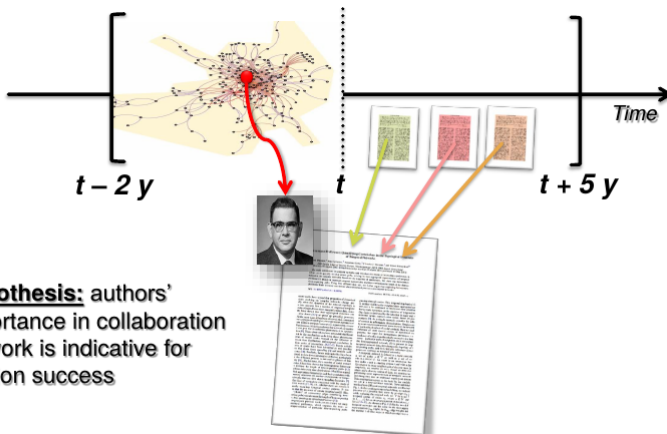


γ
George **Gamow**



β
Hans **Bethe**

Network Position and Success



Hypothesis: authors' importance in collaboration network is indicative for citation success

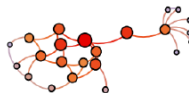
E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, F. Schweitzer: Predicting Scientific Success Based on Coauthorship Networks, EPJ Data Science (2014)

How to quantify network position?

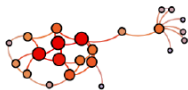
- We apply 9 different metrics (all time resolved)
 - 1 *centralization*: how many agents can be removed
 - 2 *weighted k-core decomposition*: vulnerability against cascades
 - 3 *algebraic connectivity*: identify potential break points



degree centrality



closeness centrality

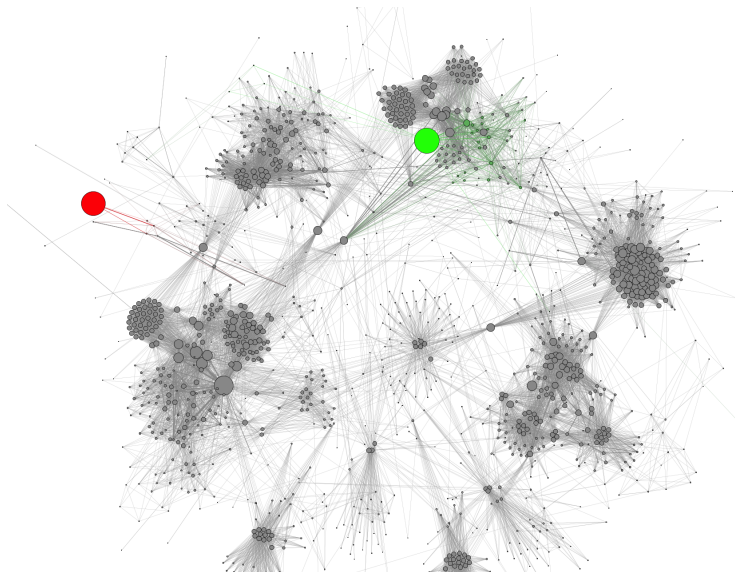


eigenvector centrality

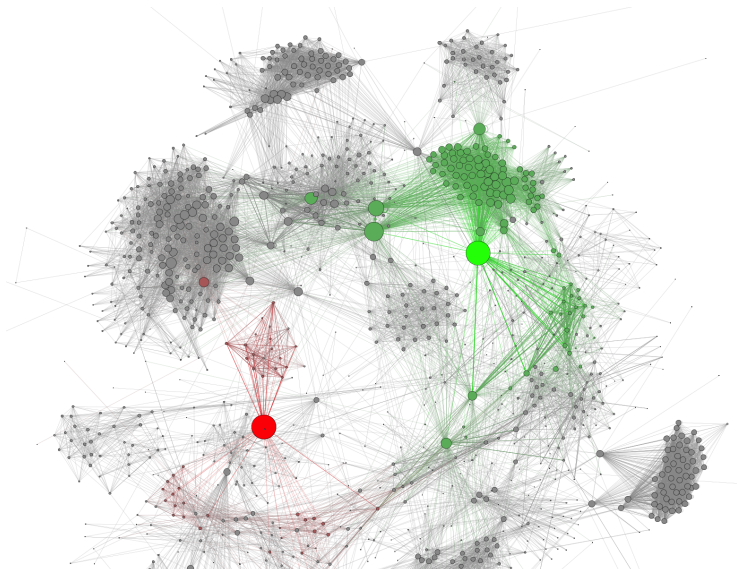


betweenness centrality

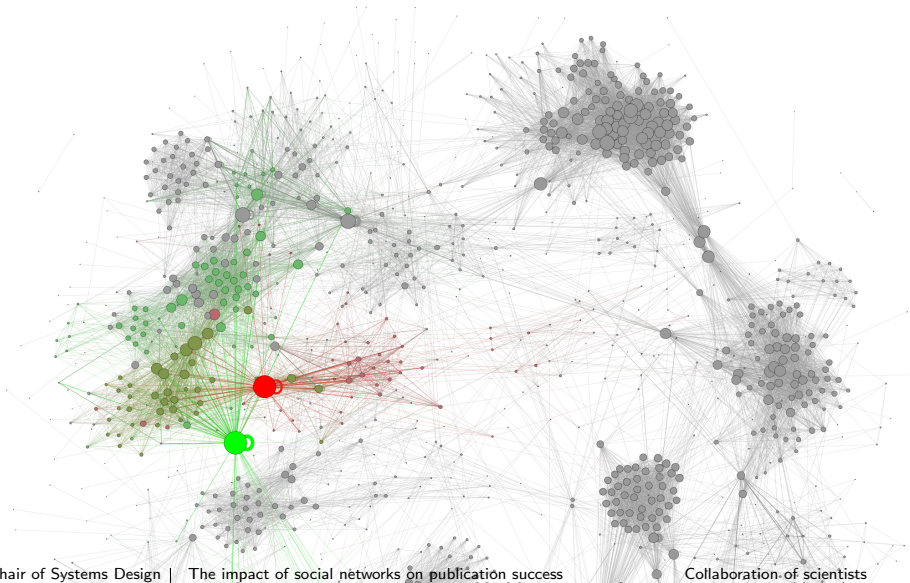
Coauthorship Network Physics (1999-2000)



Coauthorship Network Physics (2002-2003)



Coauthorship Network Physics (2005-2006)

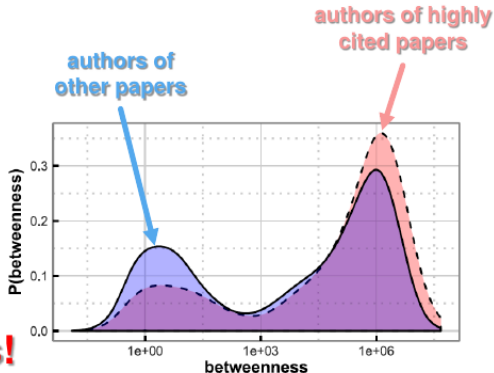


Centrality and citation success

data: ~ 108,000
computer science
publications (1996-2008)
from MS Academic
Search

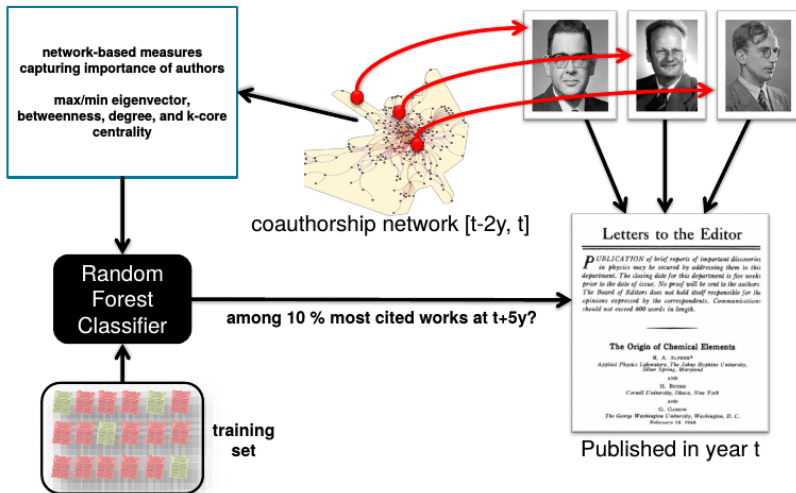
Hypothesis: authors' importance in collaboration network is indicative for citation success

YES, it is!



E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, F. Schweitzer: Predicting Scientific Success Based on Coauthorship Networks, EPJ Data Science (2014)

Prediction machine for the top 10% papers



Social position correlates with future success

	Nr.Publications	Precision	Recall
Chemistry	58129	53.9%	11.9%
Computer Science	125061	55.8%	21.2%
Biology	59871	60.6%	16.6%
Physics	52008	49.0%	8.4%

■ Bad precision? Not really ...

- *precision*: (correctly classified)/(classified)
- A random guess leads to a precision of 10% \Rightarrow **5 - 6 times better**

■ Bad recall? Thank god!

- *recall*: (classified)/(existing)
- (CS) For only 2.580 out of 12.000 papers the social position of the authors already determines their success
- 1.440 are correctly identified
- There is room for nobodys to become champions

Unknown Champions?

How to use these insights? ⇒ Better understanding of “success”

1 Reveal impact of social network on success

- Intuition is right! But now, it can be quantified.
- *Surprise:* Role of the social network

Unknown Champions?

How to use these insights? ⇒ Better understanding of “success”

1 Reveal impact of social network on success

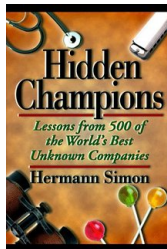
- Intuition is right! But now, it can be quantified.
- *Surprise*: Role of the social network

2 Discover success without social network

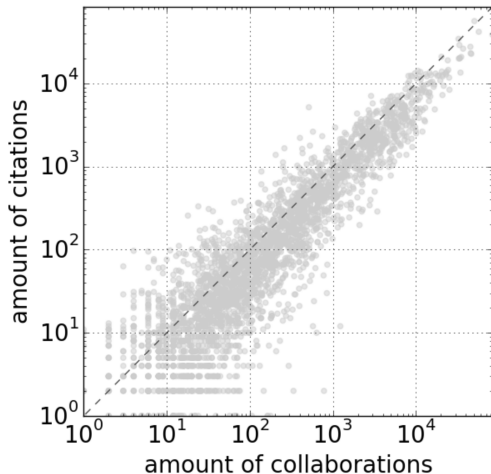
- Focus on genuine contribution
- *Surprise*: Occurrence of success

Two different levels:

- Individual scientists
- Institutions
 - aggregate over collaborations



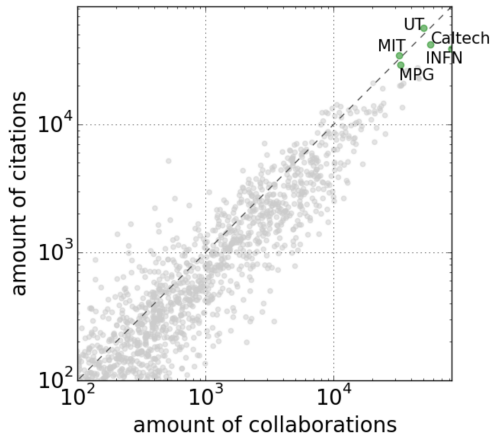
Citations increase with collaborations



- *dots*: universities, research institutions
- almost linear relationship between collaborations and citations (log-log)
- **1st order strategy**: increase collaboration

Data: APS 2005-2010, co-authors, citations aggregated to institutional level

Citations increase with collaborations



And the winner is ...
 ... the usual suspect
 ... some exceptions (APS)

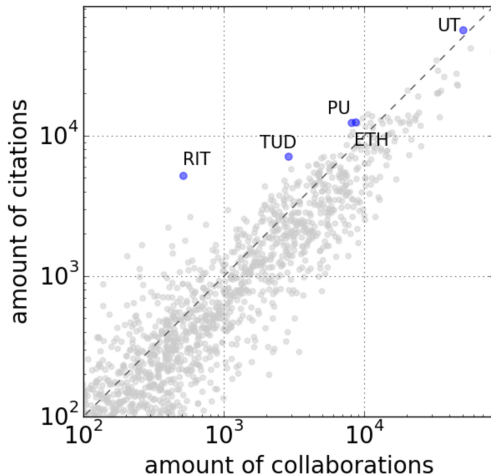
Highest ranked institutions:

University of Tokyo
 California Institute of Technology
 Istituto Nazionale di Fisica Nucleare
 Massachusetts Institute of Technology
 Max Planck Gesellschaft
 Stanford University
 University of California Berkeley
 Atomic Energy Commission
 Fermi National Accelerator Laboratory
 University of Maryland

Is this a good measure of success?

Data: APS 2005-2010, co-authors, citations aggregated to institutional level

Citations increase with collaborations



Hidden Champions:
above the diagonal

Highest ranked institutions:

University of Tokyo
Rochester Institute of Technology
Princeton University
Delft University of Technology
ETH Zurich
Harvard University
National Institute of Standards and Technology
University of Washington
University of Innsbruck
National Aeronautics and Space Administration

Data: APS 2005-2010, co-authors, citations aggregated to institutional level

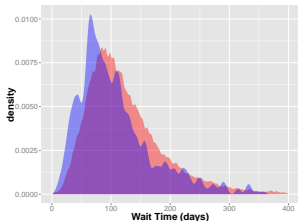
Do social relations pay off in science?

- **Data set:** PLOS ONE (2007-2014) (1 out of 7 PLOS journals)
 - 120.000 publications (i.e. 15.000 p/a), current charge: 1.500 USD
 - 350.000 authors, 4.238 (out of 7.836) editors/authors

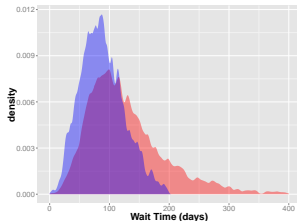
Do social relations pay off in science?

- **Data set:** PLOS ONE (2007-2014) (1 out of 7 PLOS journals)
 - 120.000 publications (i.e. 15.000 p/a), current charge: 1.500 USD
 - 350.000 authors, 4.238 (out of 7.836) editors/authors
- **Analysis:**
 - measure distance of editors-authors on co-authorship network ($d_e = 1$)
 - calculate handling time (submission-accept)

Payoff:

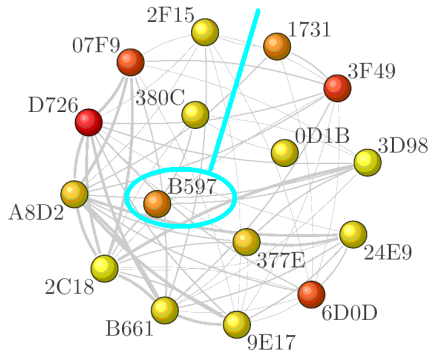


Previous co-authors gain **19 days**



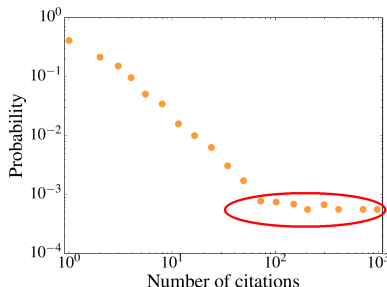
central "repeat customers" gain **38 days**

The Strategic Perspective



A small elite

- **Example:** PACS 89 community (2000-2011)
 - total size of the community: $N = 4100$
 - top 10% most cited authors: 230, top 5%: 120

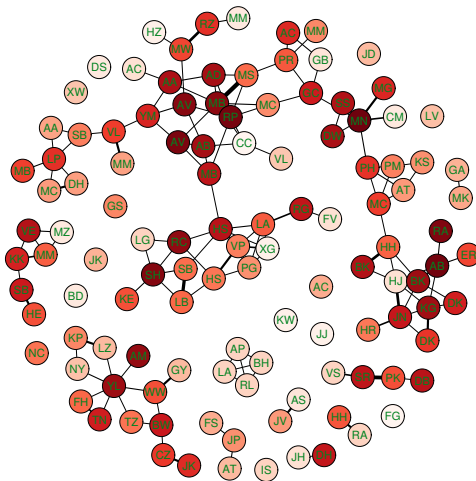


- **biggest problem:** top range authors review/cite top range authors

PACS: Physics and Astronomy Classification Scheme

PACS 89: Other areas of applied and interdisciplinary physics

No 'rich club' effect



- distinctive groups, same reputation \Rightarrow strong competition

Better than the rest?

■ Observations

- very small elite, clustered, sparsely connected
- comparable reputation (# citations)

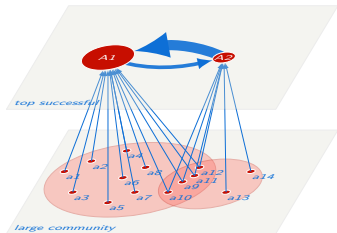


■ How to improve own strategic position?

- 1 *increase own* reputation by other means \Rightarrow *popularity*
 - public appearance, mass media, social media, ...
- 2 *decrease others* reputation \Rightarrow *malicious behavior*
 - anonymous reviews, discredit, omission

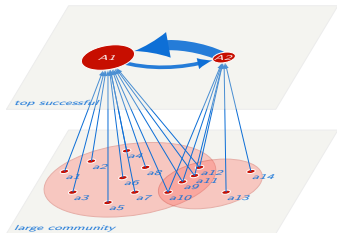
\Rightarrow **systemic feedback on strategic behavior**

Strategic citing behavior



- known: $\hat{k}_{i \rightarrow j}$: number of citations
- *How many citations top author A_i deserves from top author A_j ?*
 - depends on topical overlap
 - proxied by community similarity
 - results in asymmetric weight $w_{i \rightarrow j}$

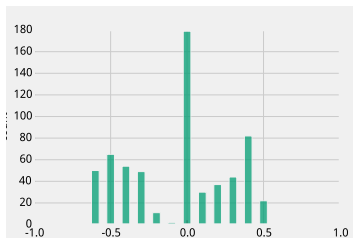
Strategic citing behavior



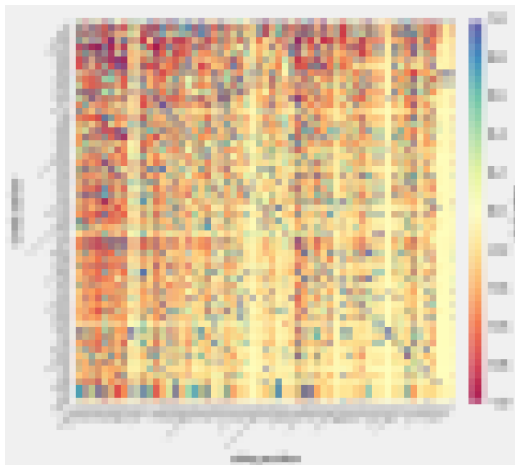
- known: $\hat{k}_{i \rightarrow j}$: number of citations
- *How many citations top author A_i deserves from top author A_j ?*
 - depends on topical overlap
 - proxied by community similarity
 - results in asymmetric weight $w_{i \rightarrow j}$

■ Friend-or-Foe index (FoF)

- $w_{i \rightarrow j} \Rightarrow$ network ensemble $\Rightarrow P(k_{i \rightarrow j})$
- $P(k_{i \rightarrow j}) \Rightarrow$ CDF
- $\hat{p}_{i \rightarrow j}$: percentile, in which $\hat{k}_{i \rightarrow j}$ falls

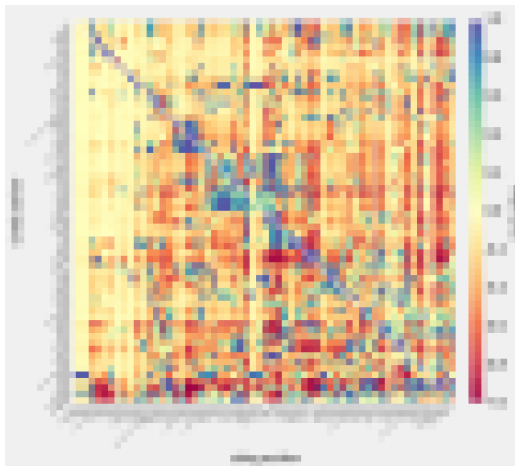


FoF index: Top 100/ PACS 89



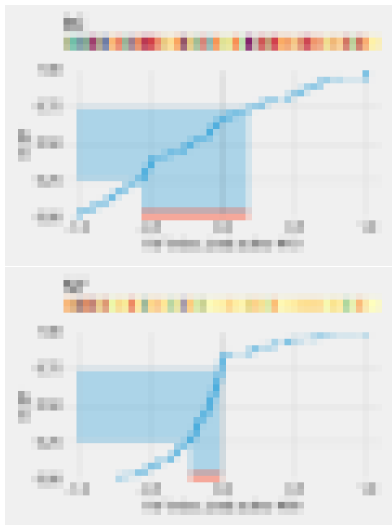
- Bands of asymmetric relations
- Cooperating clusters
- Top authors undercite everyone
- Top authors are overcited by everyone

FoF index: Top 100/ PACS 89



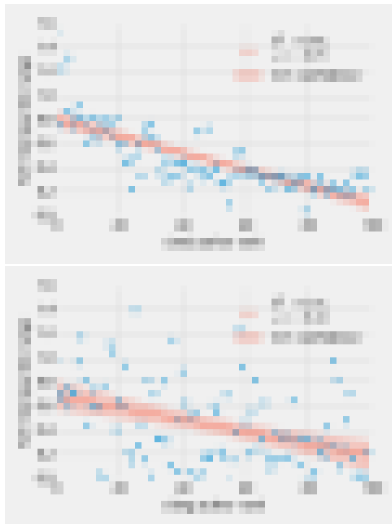
- Bands of asymmetric relations
- Cooperating clusters
- Top authors undercite everyone
- Top authors are overcited by everyone

FoF index: Individual researchers/ PACS 89



- **RC**: ranked 10th
NY: ranked 90th
- Empirical CDF of FoF values
 - broad: large interquartile range
polarizing
 - narrow: small interquartile range
rather neutral
- Different attitudes: 10. vs 90.
 - negative bias

Friends and Foes: PACS 89



- *the larger the interquantile range, the more polarizing*
- **cited author:**
 - the better the rank, the stronger the FoF
 - reason: strong competition
- **citing author:**
 - the better the rank, the stronger the FoF
 - but less pronounced
- **good news:** *from rank 20 on, reasonable behavior*

Conclusions

Role of social network in having success

- 1 **Top 10% paper:** partly predictable by social position
- 2 cannot make a bad paper a good one, but a good one a top one
- 3 **Unknown champions:** success without network

E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, F. Schweitzer: Predicting Scientific Success Based on Coauthorship Networks, EPJ Data Science (2014)

Conclusions

Role of social network in having success

- 1 **Top 10% paper:** partly predictable by social position
- 2 cannot make a bad paper a good one, but a good one a top one
- 3 **Unknown champions:** success without network

Quantifying strategic interaction

- 1 **Friend-or-Foe index:** the better ranked, the more strategic citations
- 2 **competition:** malicious strategies invade communities

E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, F. Schweitzer: Predicting Scientific Success Based on Coauthorship Networks, EPJ Data Science (2014)