

# The conundrum of scientific collaboration in a competitive science: individual motivations and social contexts

**Supervisor:** Prof. Flaminio Squazzoni

**Student:** Ali (Aliakbar Akbaritabar)

[Aliakbar.akbaritabar@unimi.it](mailto:Aliakbar.akbaritabar@unimi.it)

[www.scholar.socialvillage.me](http://www.scholar.socialvillage.me)

# research question(s)

## Big questions:

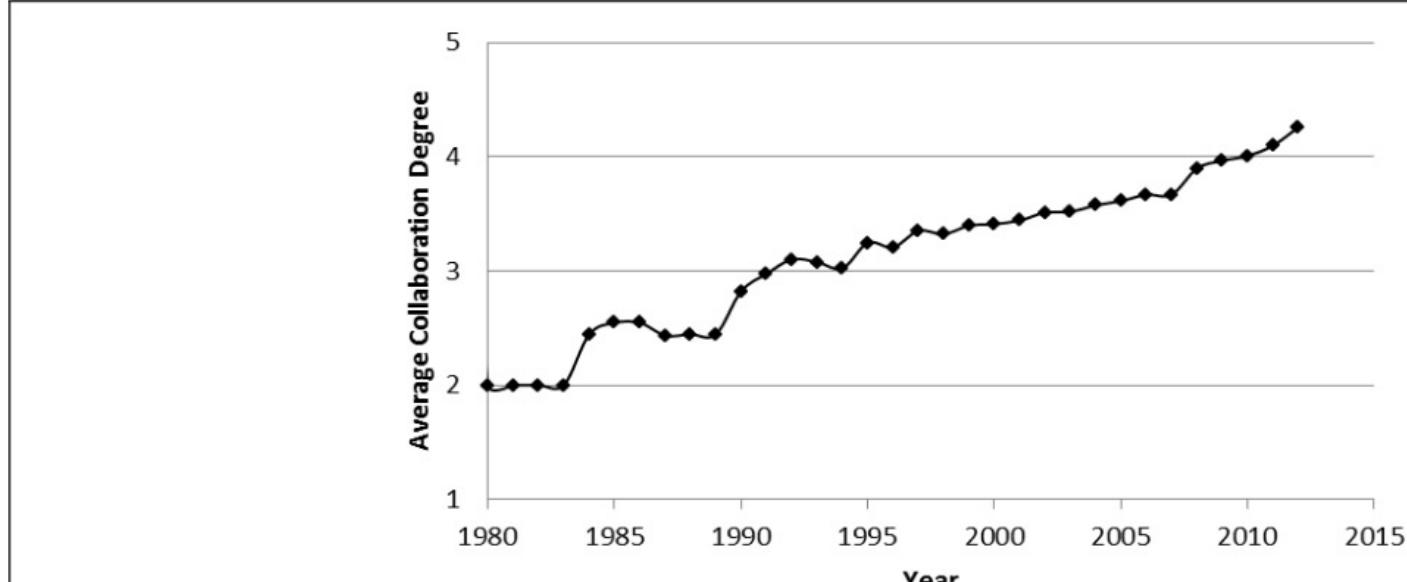
- How scientific community balances competition and collaboration?
- Is Mertonian science old wine in an old bottle?
- How scholars manage **collaboration** and **career competition**, i.e., the scientific community and the academic organization?
- Is there anything specific to **sociology** and **sociologists** in this?

## Smaller questions:

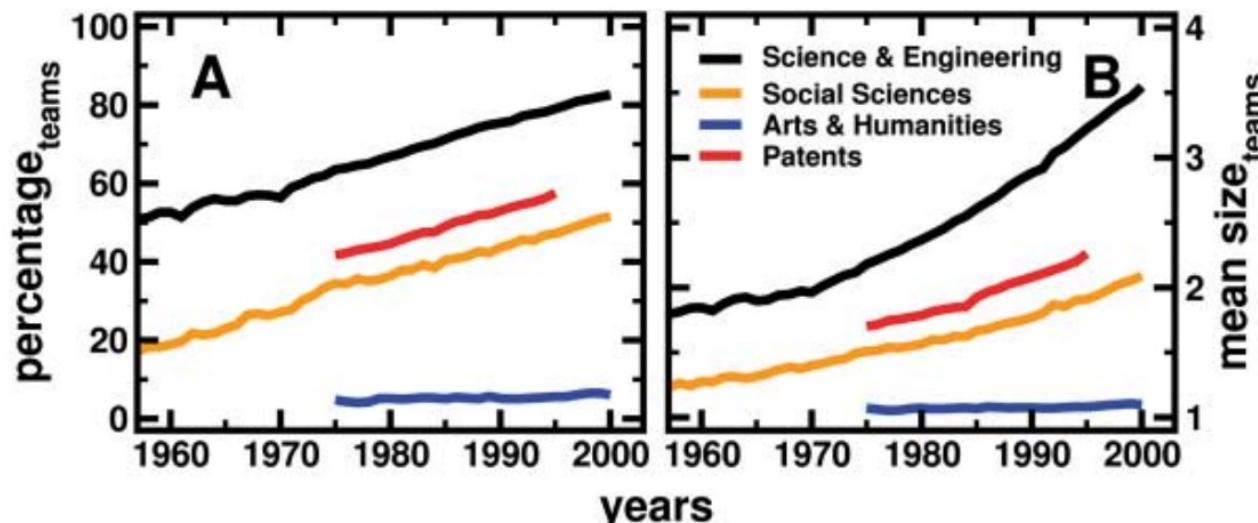
- **Descriptive** questions ([so far this report](#)):
  - Does scientific collaboration change over time? How?
  - Is collaboration context-specific? Fields? Universities? Ranks?
  - Does it only reflect the increasing complexity of research or is it more a social capital channel?
- **Explanative and more theory oriented** questions ([future steps](#)):
  - Does scientist social capital influence scientific collaboration patterns?
  - Are scientists strategically managing social capital in a competitive manner?

# Part of Literature reviewed

Xu, et al.  
Journal of the Association  
for Information Systems  
Vol. 15, Issue 12,  
pp. 835-859,  
December 2014



**Figure 2. Changes in Average Degree Centrality Over Time (in Yearly Networks)**



Wuchty, et al.  
Science 316 ,  
1036 (2007)

**Fig. 1.** The growth of teams. These plots present changes over time in the fraction of papers and patents written in teams (**A**) and in mean team size (**B**). Each line represents the arithmetic average

# Part of Literature reviewed

- Becher and Trowler (2006; 124); **laboratory and facilities needed in different fields** (like **physics**) and their effect on collaborations and necessity of team work (in contrast to situation of **mathematics** and **sociology**; who are known to usually be solo-writers)
- In social sciences, fields closer to natural and hard sciences (like **psychology**) have features similar to them in collaboration patterns
- New policies like **Horizon 2020** and **COST projects** to foster collaborations rather than single researches;
  - Possible question: how much it has been successful (not just based on names, in action, based on the papers and presentations ... published based on those cost projects ...)

## Part of literature reviewed

Examples of affecting variables (based on Tie formation versus tie persistence in scientific collaborations Dahlander and McFarland (2013))

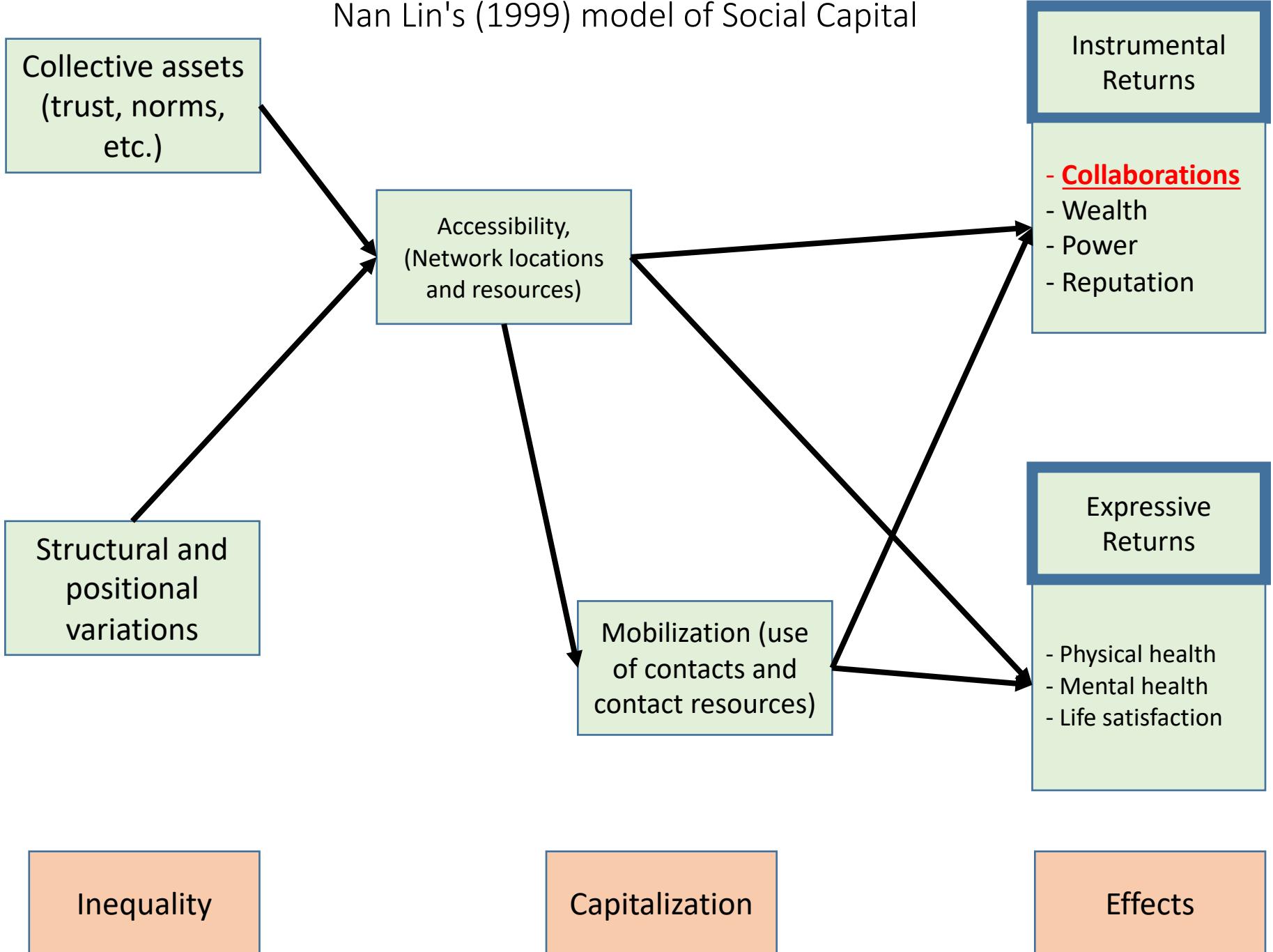
### Independent variables

- Organizational foci
- Status and value homophily
- Cumulative advantage
- Triadic closure
- Tie inertia
- Means-ends rationalization

### Dependent variables

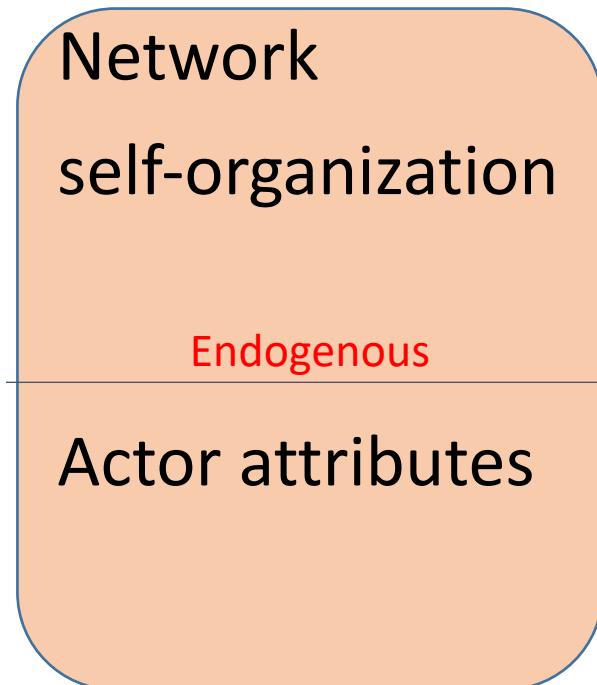
- Scientific collaborations
  - Tie formation
  - Tie persistence
  - (they have seen both papers that has been published and grants they have applied for, either successful or not)
- Control variables (tenure status, gender, and ethnicity, Exclusion restriction Appointment year difference)

# Nan Lin's (1999) model of Social Capital



# Theories of social influence and social selection in tie formation

How and  
why  
collaboration  
ties form?



Activity / popularity

Reciprocity

Closure

Brokerage

Sender effect

Receiver effect

Mix of sender & receiver

Other networks (multiplexity)

Social factors & context

# Design : case selection, methods, ...

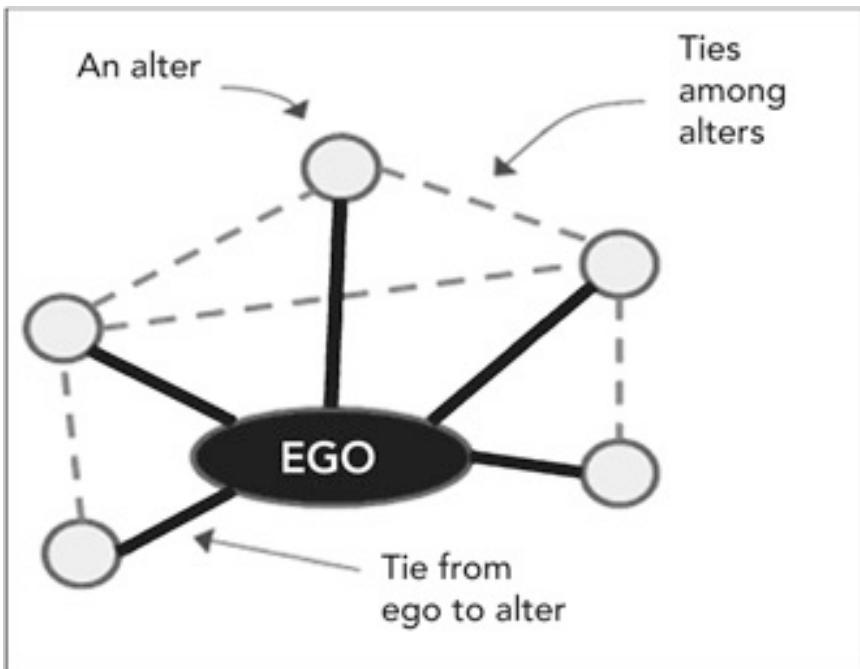
**Cases:** all papers of each individual scholar (**as ego**); (unit of analysis is individual scholars not journals or papers as usual in the literature)

**Methodology:** Social network analysis (mostly with **ego-centric approach**) to see the embeddedness of scholars in their context and how they have mobilized their potentially accessible resources to get returns

**Data:** So far google scholar data; next, web of science, PEERE project data (probably).

**Operationalization:** co-authorship networks as a proxy of how scholars interact with each other

**Results:** Checked different structural properties of ego-networks and here some are presented



# A shiny web application of R function developed for data crawl

Who are researchers mostly co-authoring with?

see how it works



Copy the link to your person of interest Google Scholar profiles below and see the co-authorships network; go here to see a video of how it works <https://goo.gl/053N2o>

google scholar profile link

<https://scholar.google.com/citations?user=BEzh-pcAAA>

Submit

Reference: data for co-authorships and publications are being extracted from google scholar with 'scholar' package for R

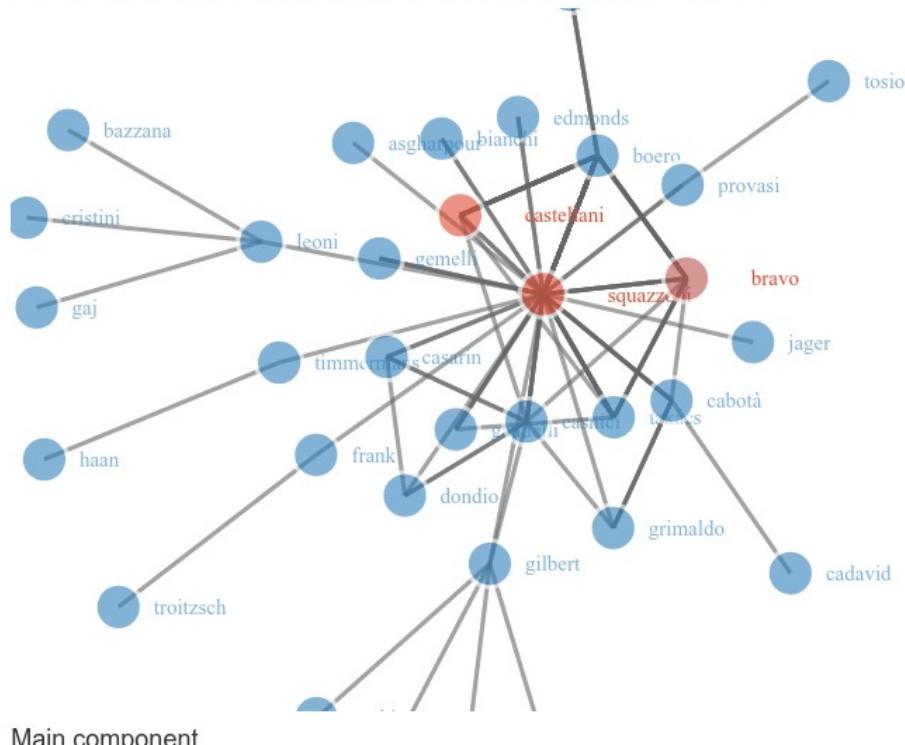
\*\*: It is the first and earliest version, some issues are known and in the process to be solved, some not, I will be happy to hear: akbaritabar[at] gmail.com

author's data

	number of articles	unique journals	oldest article	number of top journals	h-index
1	92	57	2000	0	15

Raw co-authorships Graph

use mouse scroll to zoom in and out; further details (including main component) below



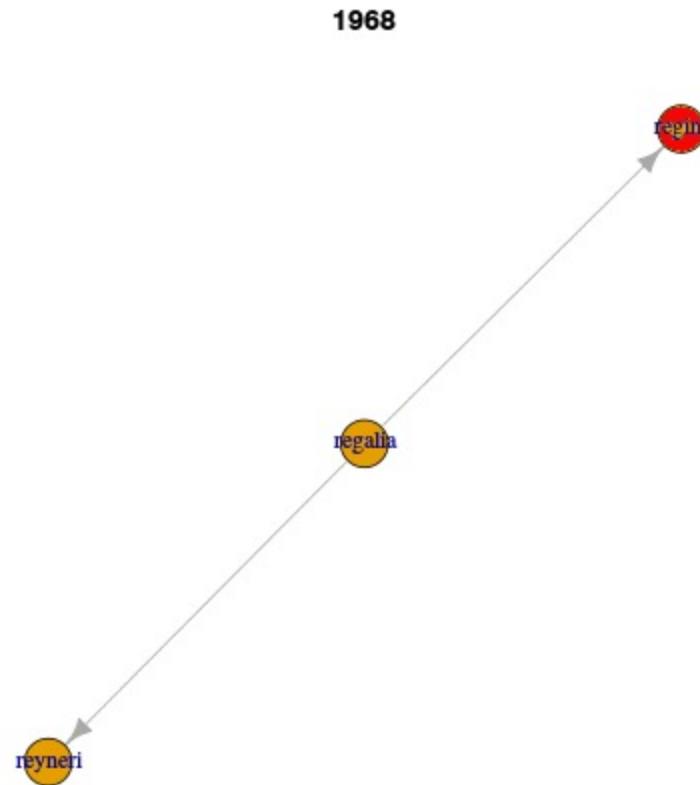
Scholar.socialvillage.me

# 2 points of view to co-authorships

- Choosing **some journals** and following all papers published in them to see co-authorships (can be called sociocentric and complete network view)
- Choosing **some scholars**, and following all their papers and publications in different journals to see co-authorships (can be called ego-centric view)
  - Consider scholar like “**network manager**” type in **Burt’s** terms, someone who decides and manages his scientific collaborations through time (**like a stochastic actor oriented model assumption**)
- How you define your nodes and ties, changes the way you are capturing the network and model the social phenomenon under study ...
  - Revisiting the Foundations of Network Analysis, Carter T. Butts, Science 325 , 414 (2009)

# Trend of changes in co-authorships

- Aged, full professor, 1968 – 2015
- Ego colored in Red



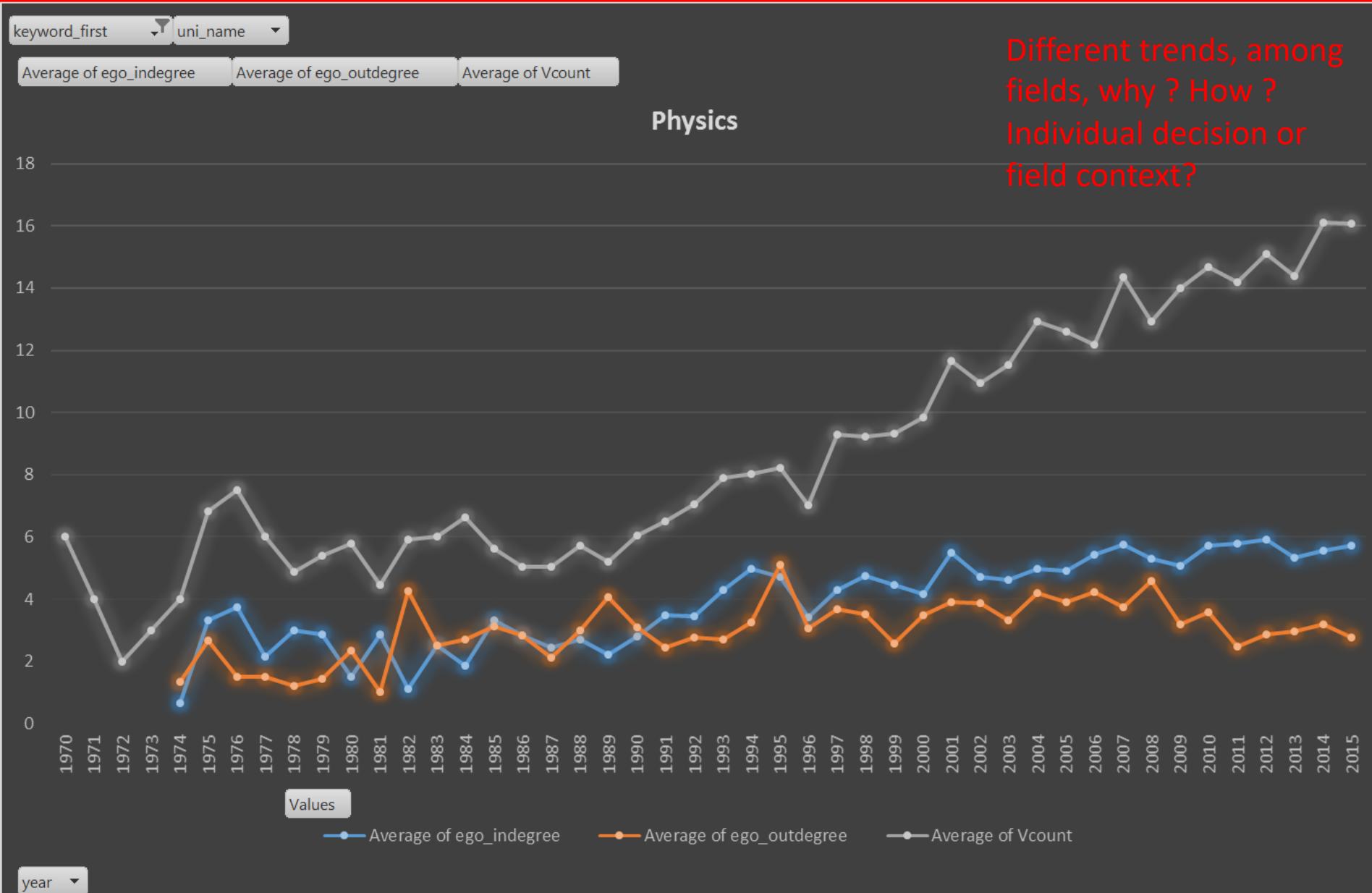
# Study sample (so far)

Sample of study	
Total sample of scholars	8373
Number of scholars <b>without name inconsistencies</b> (reported here)	<b>2578</b>
Next step	Clearing remaining names

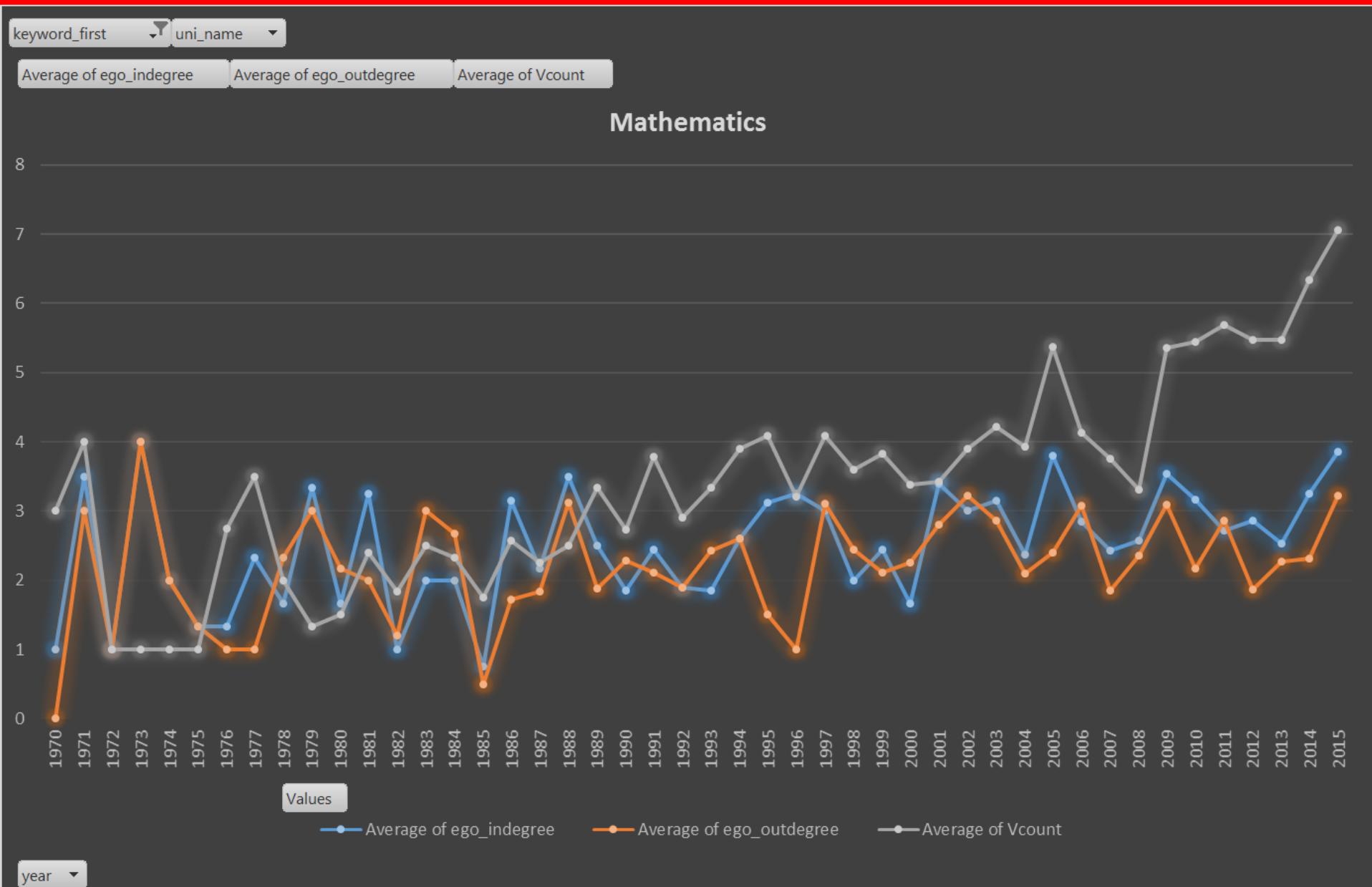
- **gathering:** Feb 20-21; **cleaning:** Feb 22 – May 8
- **Author names inconsistencies** that took long to solve (took same approach similar to research like Dahlander and McFarland (2013) to solve)
- Limiting time span from 1970 to 2015 (45 years)
- Randomly chosen scholars

University name	Initial sample before name correction	Number of scholars in report
Higher_School_of_Economics	530	311
Lomonosov_Moscow	433	251
UAB	529	164
ETH_Zurich	530	161
Milan_(unimi)	500	161
McGill	530	158
ANU	529	158
MIT	530	148
Tehran	496	145
Cambridge	530	138
Ecole_superieure	229	126
UCL	530	120
Oxford	498	95
Brescia	223	87
Stanford	530	86
NU_Singapore	530	84
Harvard	504	77
SciencesPo	106	60
Telecom_Bretagne	86	48
<b>Total</b>	<b>8373</b>	<b>2578</b>

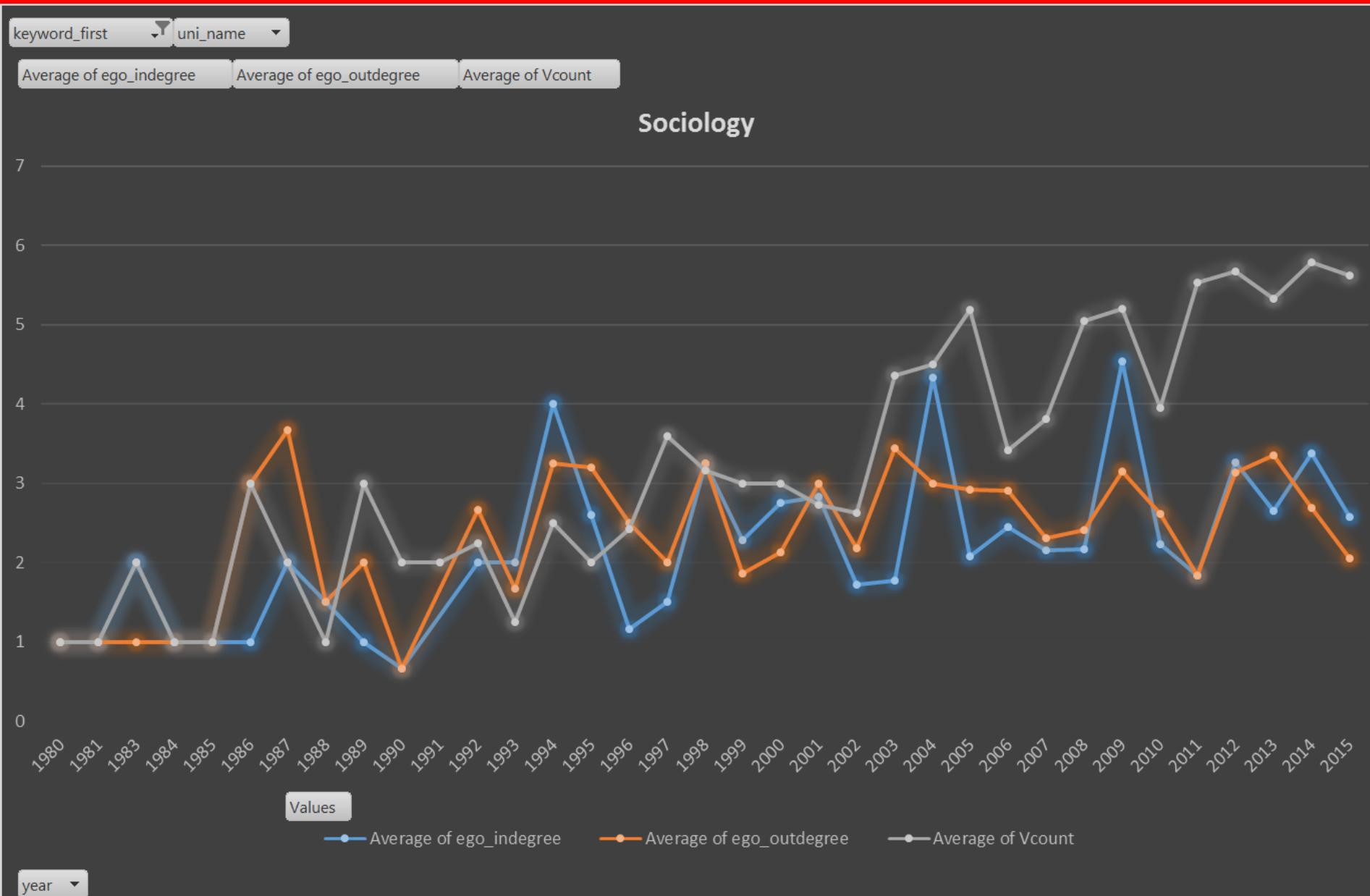
# Trend of co-authorships based on fields? (1/3)



# Trend of co-authorships based on fields? (2/3)

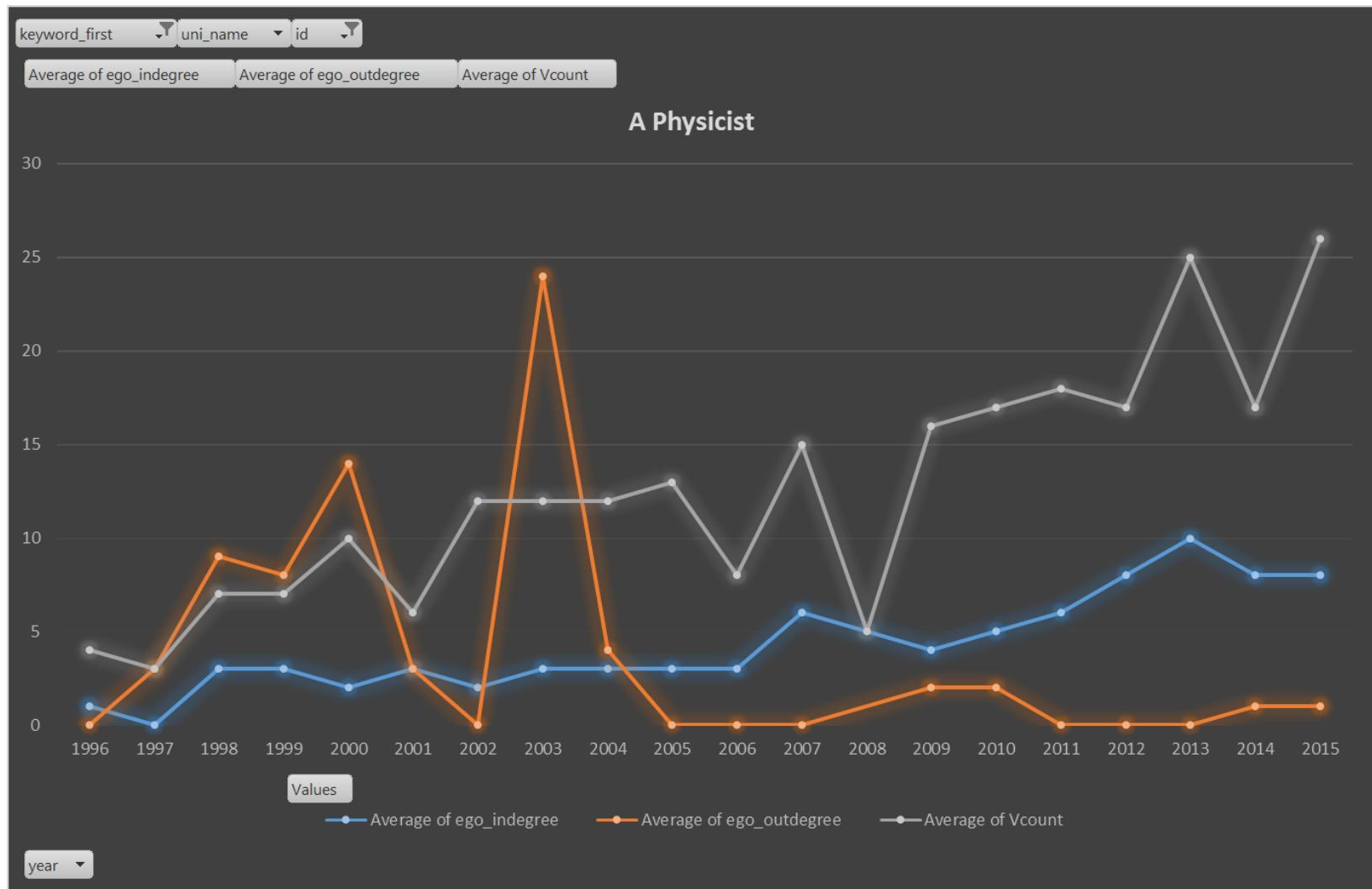


# Trend of co-authorships based on fields? (3/3)



# A physicist

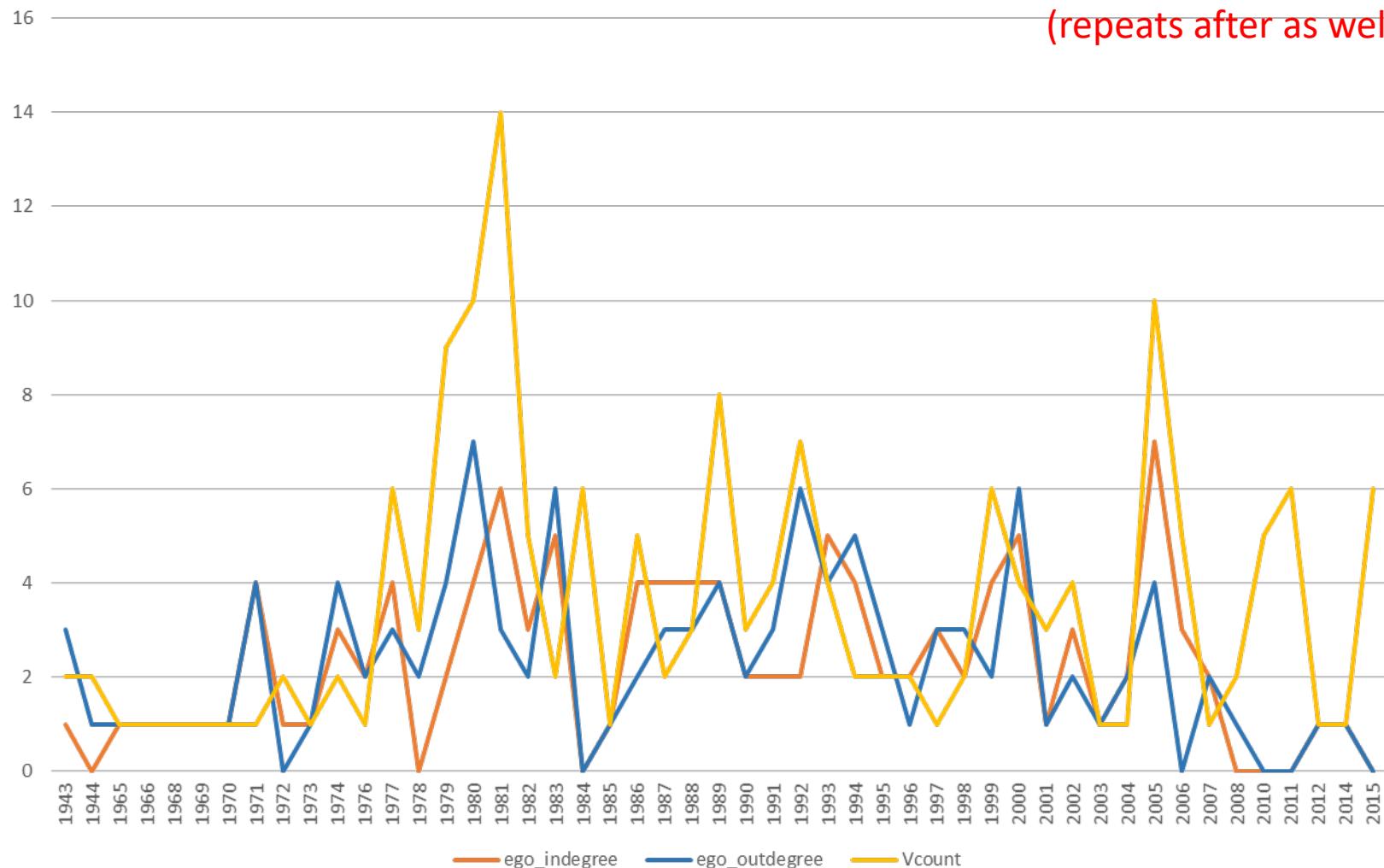
From 1995 to 2004 he is mostly lead author, from 2004 to now, 2<sup>nd</sup> to last author, why ?



Obvious difference in in and out degree; usual increase in vcount

# An anthropologist

Not so clear trend as we saw in case of physicist !  
Why ?  
1944 – 1970 solo-writer  
(repeats after as well)



# ERGM to answer why observed network is happened

Effects	Lambda	Parameter	Stderr	t-ratio	SACF	
ArcA	2	-2.4129	0.265	-0.044	0.422	*
ReciprocityA	2	-3.6059	12.966	-0.078	-0.006	
Effects	Lambda	Parameter	Stderr	t-ratio	SACF	
ArcA	2	-2.0859	0.676	0.025	-0.036	*
ReciprocityA	2	-4.8138	12.976	-0.078	-0.006	
In2StarA	2	-1.236	0.803	0.016	0.009	
Out2StarA	2	0.4081	0.052	-0.01	-0.03	*

DIRECTED ONE-MODE NETWORKS (A & B)

Label	Configuration	Label	Configuration
ArcA		ReciprocityA	
ArcB		ReciprocityB	
In2StarA		Out2StarA	
In2StarB		Out2StarB	

Statistics	Observed	Mean	StdDev	t-ratio			GOF		
ArcA	16	16.101	3.187	-0.032					
ReciprocityA	0	0.006	0.077	-0.078	ATA-TD	4	3.671	4.221	0.078
In2StarA	4	4.13	2.402	-0.054	ATA-TU	4	3.5238	3.956	0.12
Out2StarA	25	26.573	25.394	-0.062	ATA-DU	4	3.5538	4.02	0.111
In3StarA	1	0.247	0.539	1.398	ATA-TDU	6	5.3743	6.094	0.103
Out3StarA	25	62.055	99.749	-0.371	A2PA-T	23	13.4798	6.424	1.482
TwoPathA	23	13.681	6.564	1.42	A2PA-D	25	26.348	25.121	-0.054
Transitive-TriadA	2	1.856	2.158	0.067	A2PA-U	4	3.9236	2.174	0.035
Cyclic-TriadA	0	0.323	0.559	-0.578	A2PA-TD	48	39.8278	25.472	0.321
T1A	0	0	0NaN		A2PA-TU	27	17.4033	7.745	1.239
T2A	0	0	0NaN		A2PA-DU	29	30.2716	26.152	-0.049
T3A	0	0	0NaN		A2PA-TDU	52	43.7513	26.751	0.308
T4A	0	0	0NaN		stddev_indegreeA	1.3093	1.3027	0.208	0.032
T5A	0	0.002	0.045	-0.045	skew_indegreeA	-1.1069	-1.236	0.078	1.663
T6A	0	0	0NaN		stddev_outdegreeA	2.1712	2.0845	0.775	0.112
T7A	0	0.002	0.045	-0.045	skew_outdegreeA	-0.2343	-0.1882	0.967	-0.048
T8A	0	0.044	0.654	-0.067	clustering_A_tm	0.087	0.161	0.221	-0.335
SinkA	10	6.268	1.938	1.926	clustering_A_cm	0	0.0544	0.096	-0.566
SourceA	2	1.671	1.089	0.302	clustering_A_ti	0.25	0.1847	0.174	0.376
IsolateA	0	1.113	1.143	-0.974	clustering_A_to	0.04	0.0341	0.038	0.156
AinSA	3.5	4.007	2.24	-0.226					
AoutSA	15.8125	12.3808	6.634	0.517					
AinSA2	3.5	4.007	2.24	-0.226					
AoutSA2	15.8125	12.3808	6.634	0.517					
AinAoutSA	6.3203	8.5746	3.153	-0.715					
ATA-T	2	1.8205	2.079	0.086					
ATA-C	0	0.9578	1.641	-0.584					
ATA-D	2	1.8505	2.143	0.07					
ATA-U	2	1.7033	1.888	0.157					

## Summary, Empirical and policy implications (1/2)

- I hope to understand the individual strategies in scientific collaborations; **efficient strategies to suggest to others**
- I hope to find ways to **decrease the expenses in economy of science** like university and research expenses through suggesting more fruitful strategies of collaboration
- tracing the **strategic use of social capital by scholars for career** in a competitive manner
- Better understanding of the interplay of **collaboration and competitive pressures in the scientific community** and their evolution

## Summary, Empirical and policy implications (2/2)

- An **economic sociology** approach to **science** that looks at strategic, economic motivations and social norms and consider context-specific factors (e.g., disciplinary fields)
- How **institutes can improve research initiatives** and motivate scholars
- **Where to publish**, to get more noticed and build better collaborations in future
- **Students**, where to apply, where to affiliate, who to work with (personality and strategies); (**online laboratory, give your past, know your future**)
- Is it the university and department that matters, or individual initiative  
...
- Maybe better **evaluation indexes for universities, departments, fields ...**

# Future steps

- my main question is to understand the social mechanisms behind scientific collaboration How scholars connect to each other to build up their scientific career, the first way I tried to answer this question was **through co-authorship networks**, the next ways I am thinking to add to this (to build up the parts of the puzzle) is **citation networks**, and **conceptual map of papers keywords** (that can bee seen as a kind of indirect relationship between authors/scholars); if possible **scholars personal networks**; PEERE data (**peer review as indirect interaction between scholars**) probably ... **what are other ways you think would help me explore and answer this question?**
- I have been reading about **stochastic actor oriented models (SAOM)**, **TERGM**, and I have heard about longitudinal data analysis with **Rsiena**; I am thinking to apply them
- **ABM (agent-based modeling)** to explore social and network mechanisms more detailed than what **ERGM** allows me to do

# Many thanks for your attention

## Acknowledgements:

- I am thankful for comments & suggestions of:

- My supervisor and co-supervisor
- Alessandro Lomi
- Paola Zappa
- Tom Snijders
- Viviana Amati
- Christoph Stadtfeld
- Juergen Lerner
- Federico Bianchi
- Marian-Gabriel Hancean
- Ruhollah Jalili
- Nastaran Keshavarz
- my colleagues in ESLS 31<sup>st</sup> cohort

## To find me:

Ali (Aliakbar Akbaritabar)

NASP & Economic Sociology,  
University of Milan / Brescia

[Aliakbar.akbaritabar@unimi.it](mailto:Aliakbar.akbaritabar@unimi.it)

@akbaritabar

[www.scholar.socialvillage.me](http://www.scholar.socialvillage.me)

Presentation is finished;

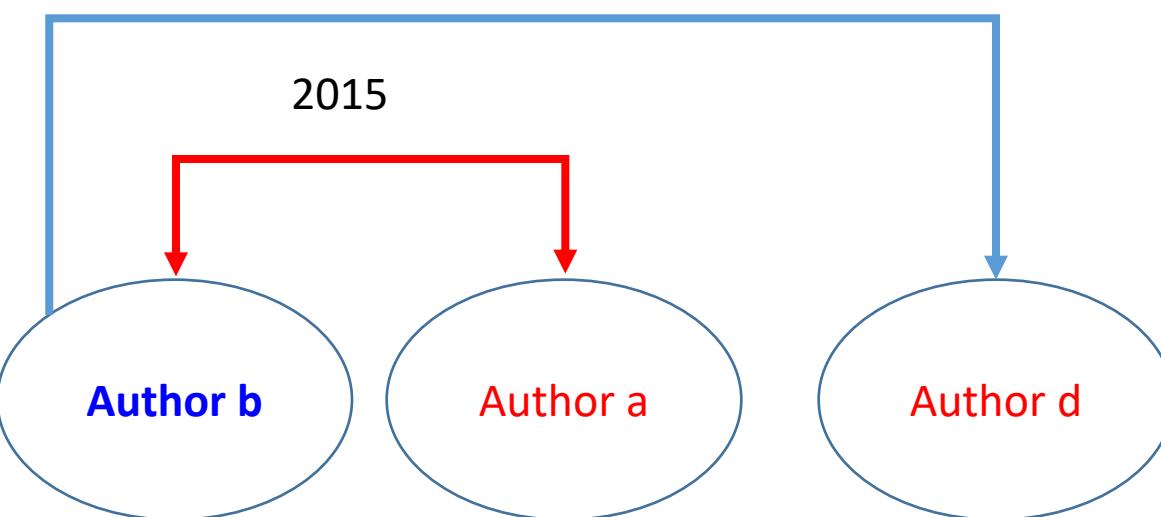
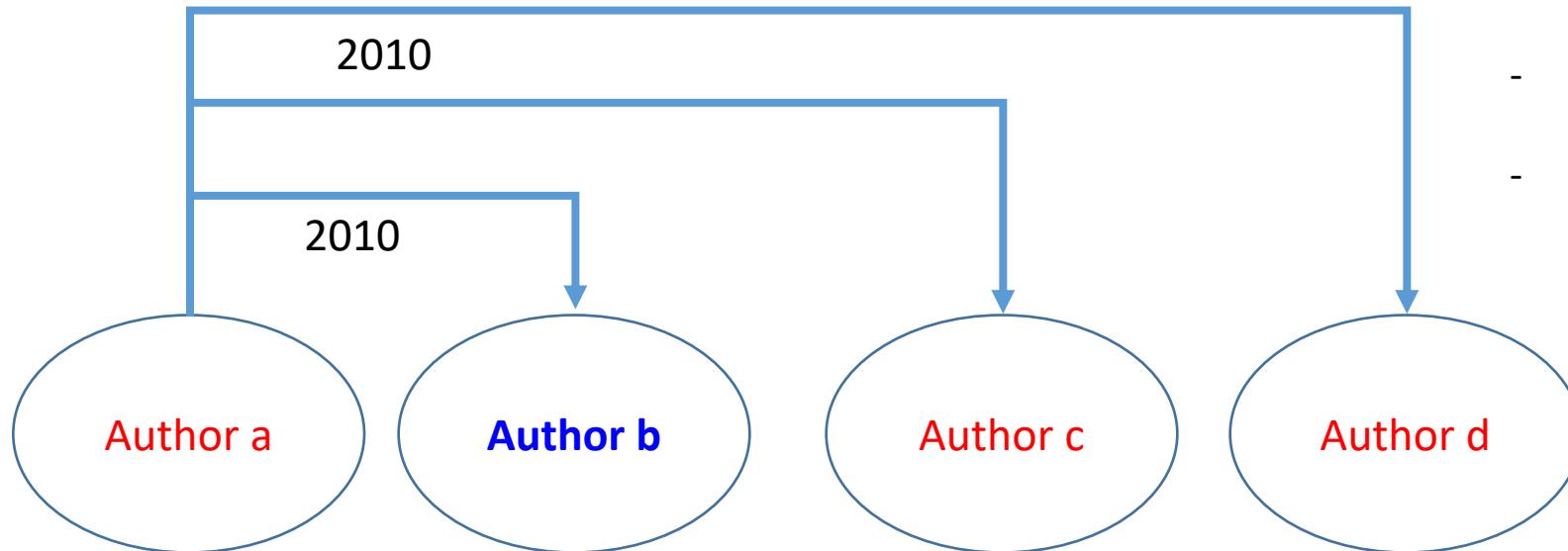
Following slides are here to be discussed in case of questions ...

# Adjacency lists based on papers' author names

Paper 1: author a, **author b**, author c, author d (2010) ..., ...

Paper 2: **author b**, author a, author d (2015) ..., ...

Paper N: ...



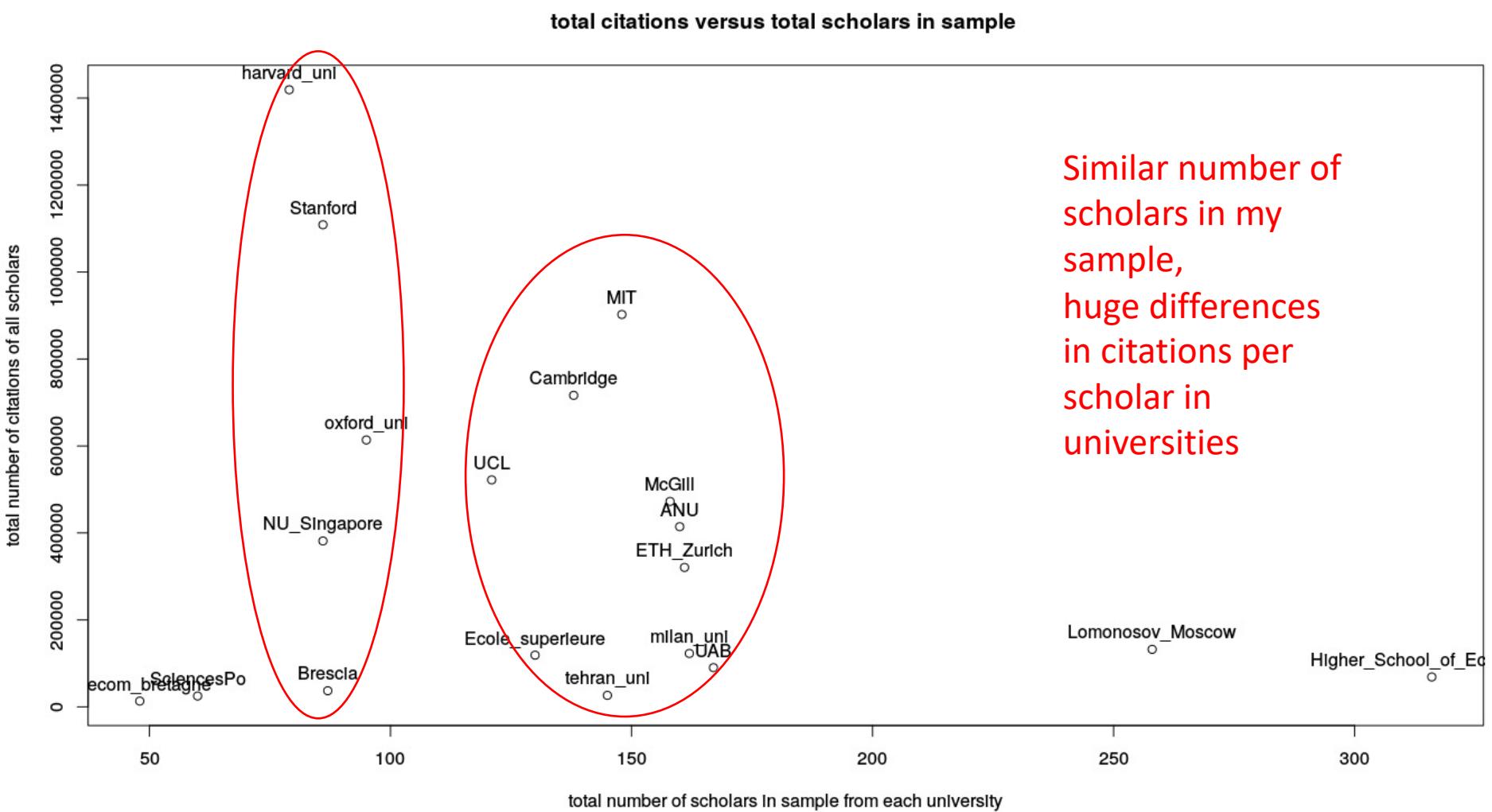
## Innovations:

- ego-centered vs. complete
- One scholar vs. one journal
- Directed vs. undirected

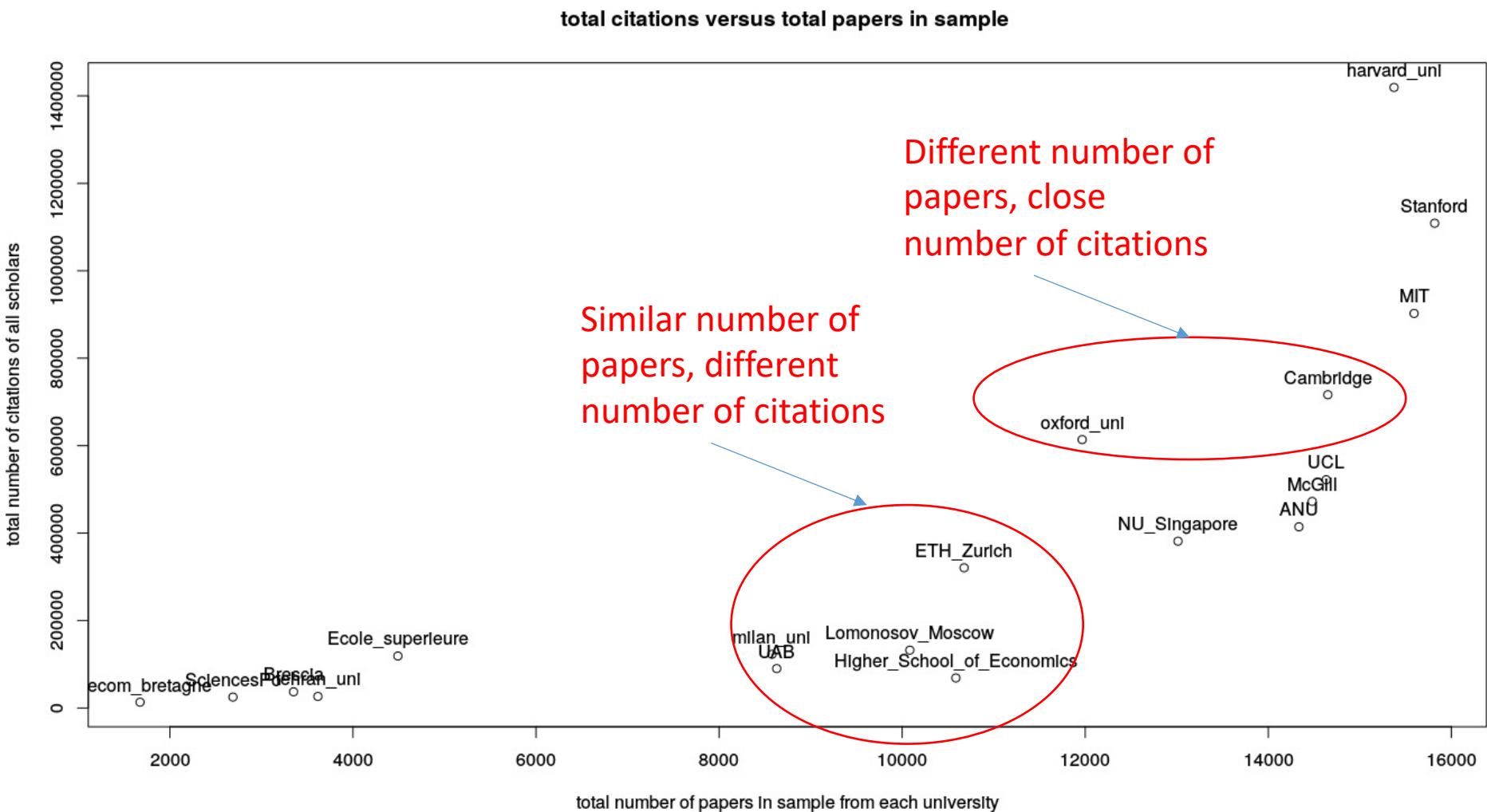
**Rationale:** the agreements between scholars about name order; takes all career long with different agreements into account, focuses on Ego

**Ego = Author b**

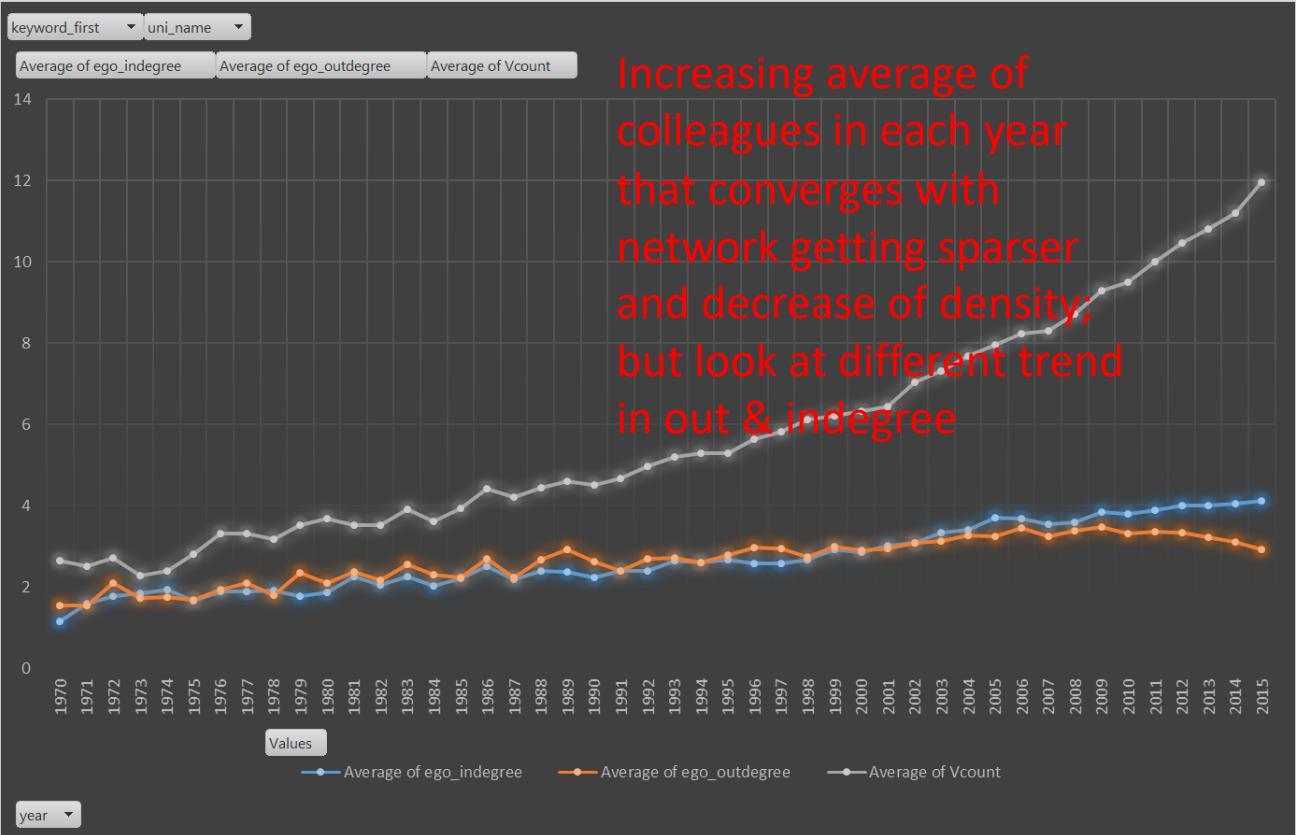
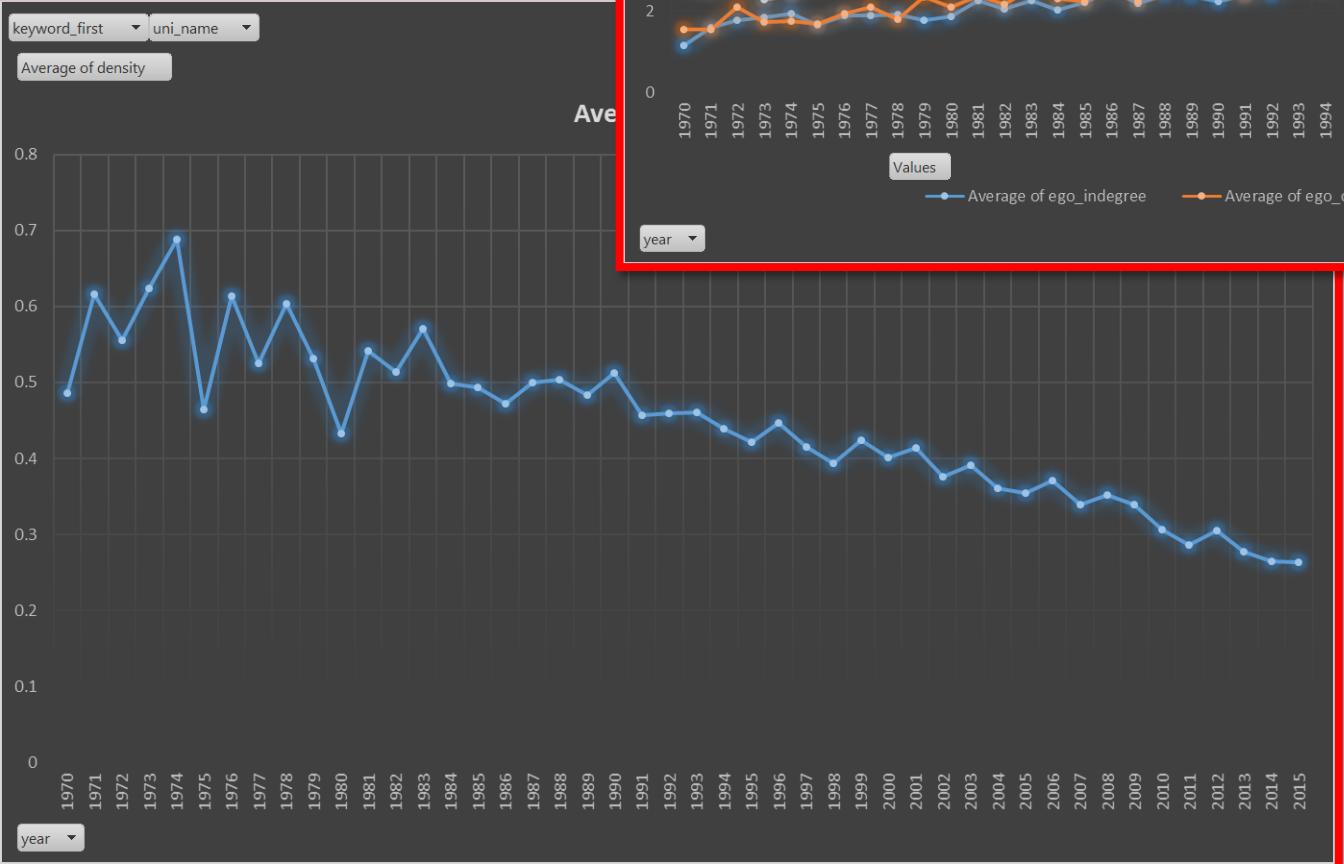
# Preliminary results



As expected, affiliation matters; Number of papers in sample vs. citations

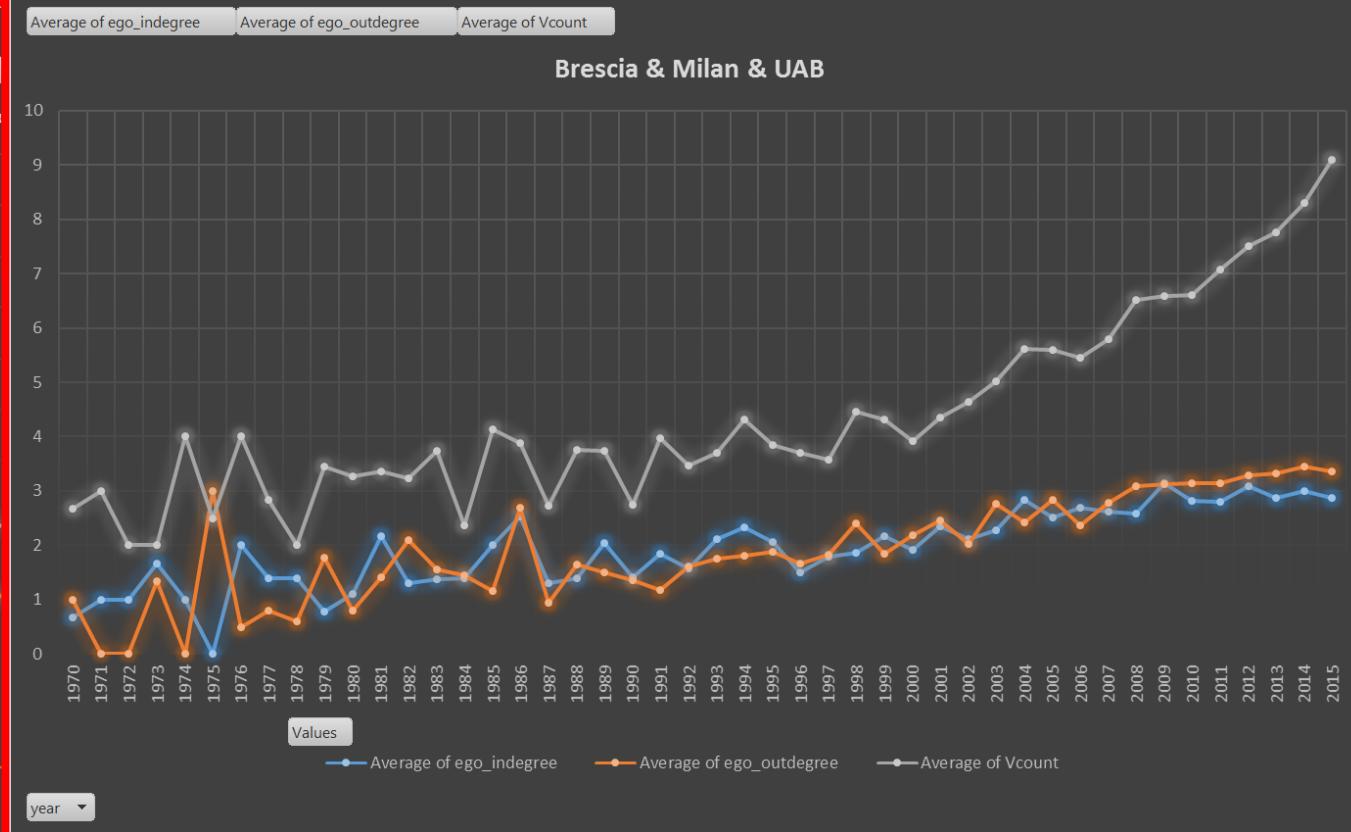
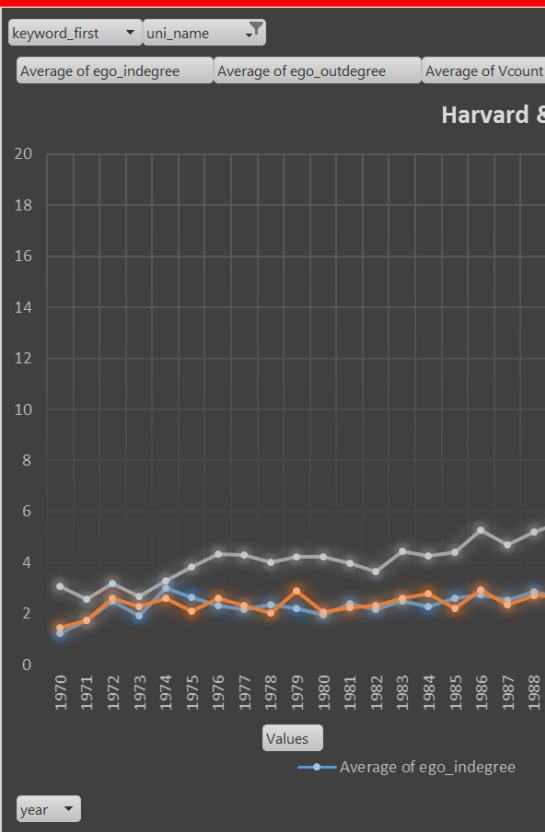
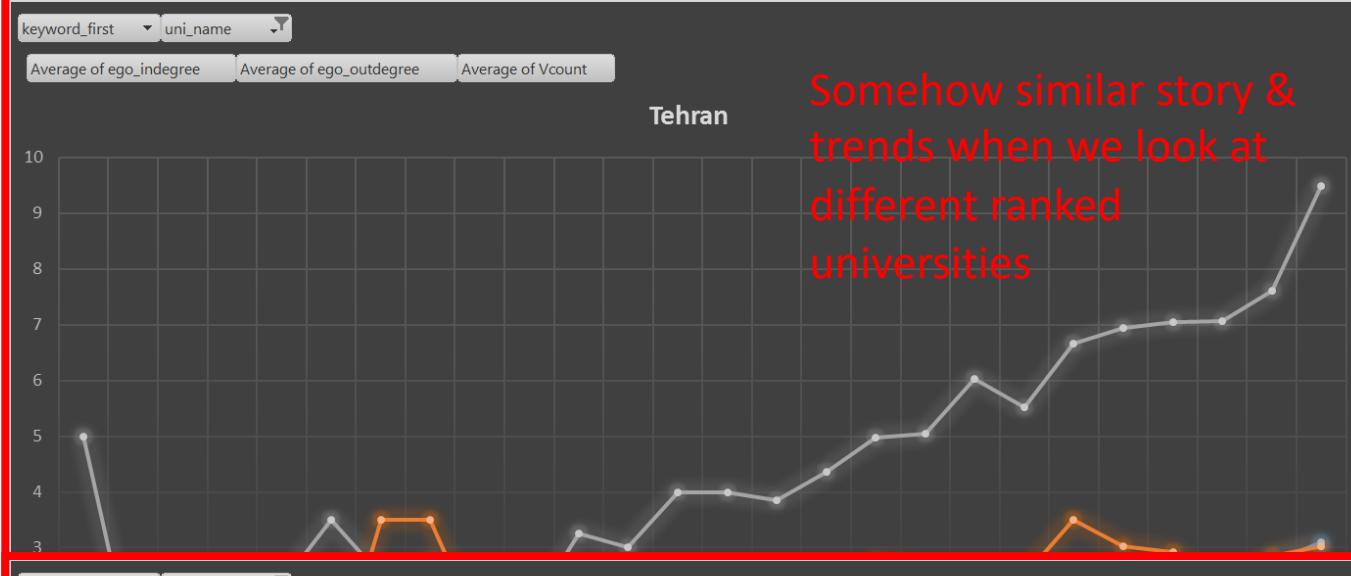


# All fields all universities

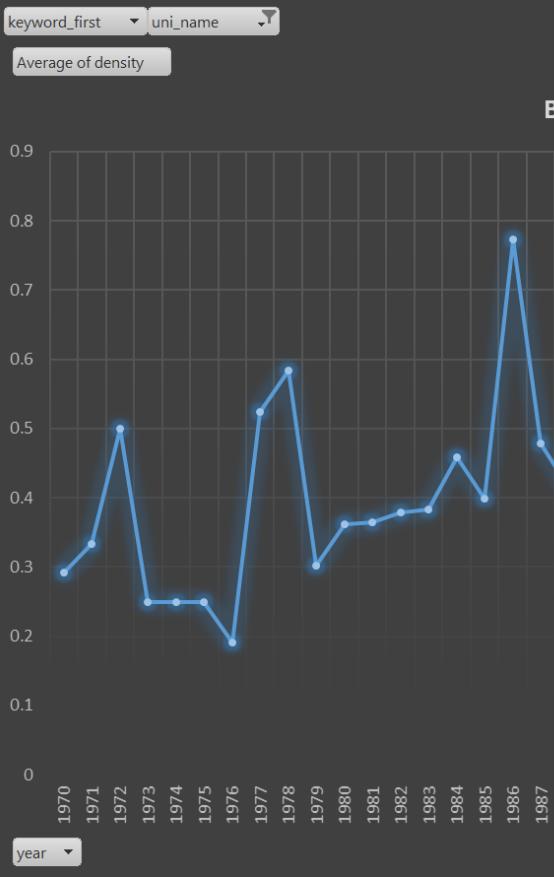
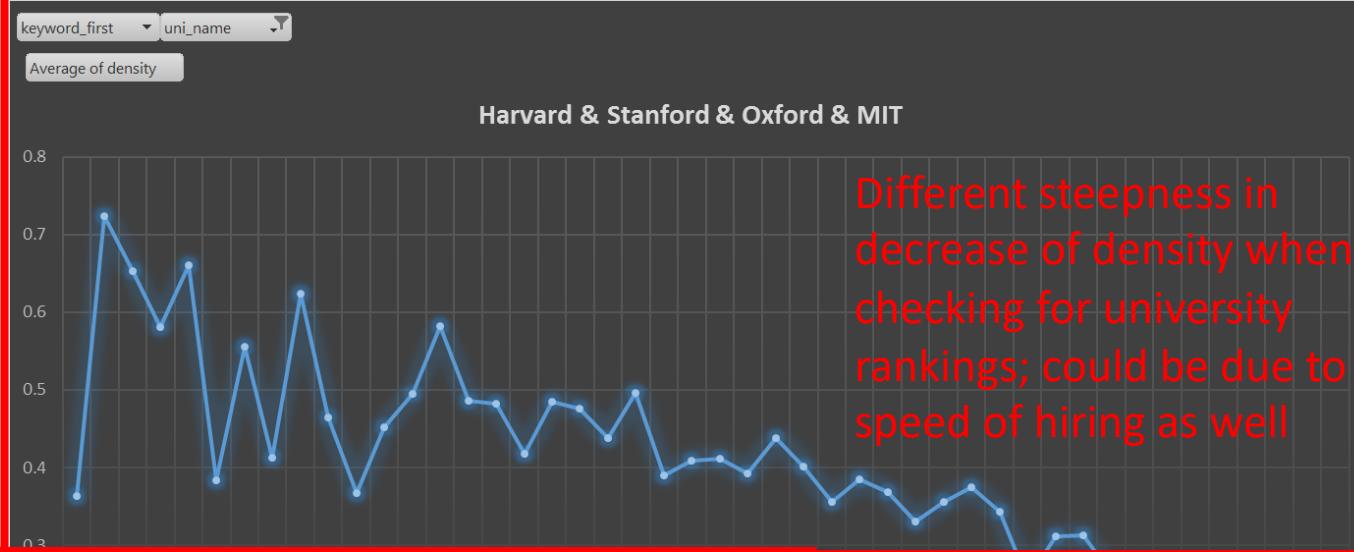


Increasing average of colleagues in each year that converges with network getting sparser and decrease of density; but look at different trend in out & indegree

# All fields universities ranked



# Average density universities ranked



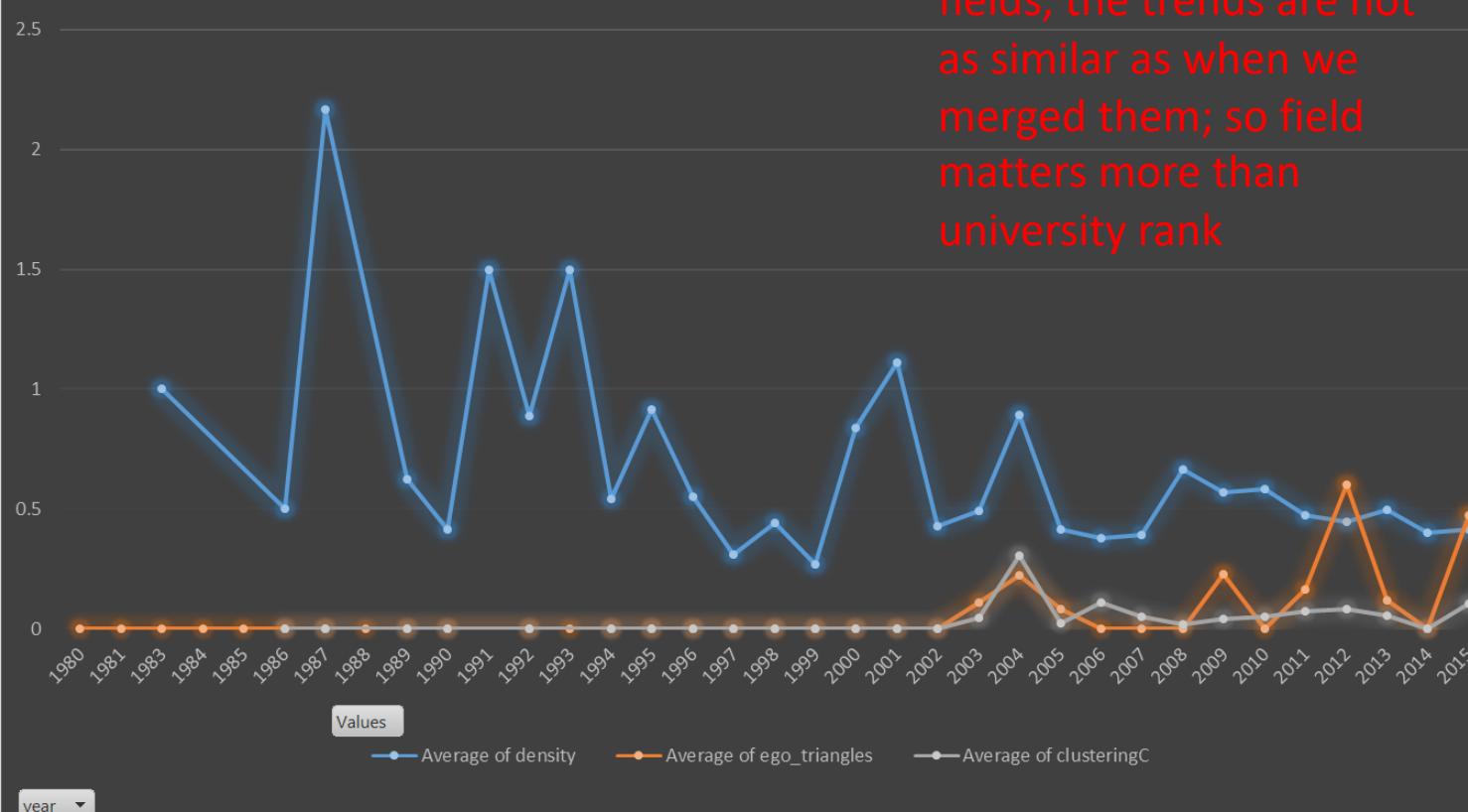
keyword\_first ▾ uni\_name ▾

Average of density Average of ego\_triangles Average of clusteringC

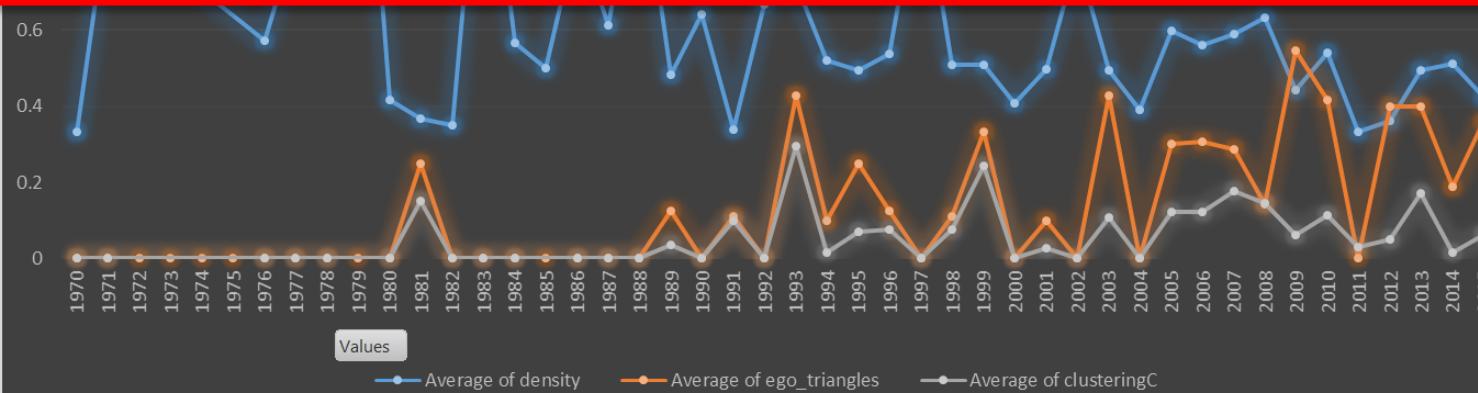
### Sociology

But when we control for fields, the trends are not as similar as when we merged them; so field matters more than university rank

How about fields?



year ▾



year ▾



gC