

Introduction to my previous research and background

A quantitative sociology of academic work in an era of hypercompetition and rankings

Ali (Aliakbar Akbaritabar)

University of Milan - University of Brescia

25/04/2018

Outline of presentation

- ▶ Introducing myself
- ▶ Introduction of my latest studies
- ▶ My plans during these three years!

Outline of research

- ▶ Introduction to a hyper-competitive academic landscape
(Chapter 1)
 - ▶ Quantity versus quality (Quantitative research evaluation)
 - ▶ Funding, promotion and reputation circulation in 21st century
 - ▶ Responsible metrics and credit allocation
 - ▶ Sociological theories
 - ▶ Matteo effect, winner takes all?
 - ▶ Attaching preferably to a few star scientists/leaders?
 - ▶ Fragmentation of ideas, sociological small world of disconnected islands?
 - ▶ Embeddedness and organizational ambiguity at work?
 - ▶ Core of leaders and periphery of followers?
 - ▶ Summing up and introduction to individual, societal, meso, macro, and multi level analyses

Outline of research

- ▶ Individual level research productivity ([Chapter 2](#))
 - ▶ Internationalization
 - ▶ Co-authors similarity
 - ▶ Individuals embedded in organizational settings
- ▶ Structural & Societal effects (Diversity in research productivity) ([Chapter 3](#))
 - ▶ Gender & Ethnic diversity in research productivity
 - ▶ Ivy-League effect in research productivity
- ▶ Meso level (Networks effect in research productivity) ([Chapter 4](#))
 - ▶ Groups collaboration behavior; subjective similarity and research communities?
 - ▶ Community membership effect
 - ▶ Community evolution

Outline of research

- ▶ Macro level (Top-down process of inspiring research productivity) ([Chapter 5](#))
 - ▶ Policy effect evaluation
 - ▶ ANVUR effect in national scale
- ▶ Multi-level approach ([Chapter 6](#))
 - ▶ Individuals embedded simultaneously in communities and organizations
 - ▶ Multi-level ERGM of research productivity
- ▶ Summary, conclusions and discussion ([Chapter 7](#))

A screenshot of a web browser window showing a GitHub page for "akbaritabar". The URL in the address bar is "akbaritabar.github.io". The page has a dark header with navigation links: "About Me", "CV", and "Research Projects & Ideas".

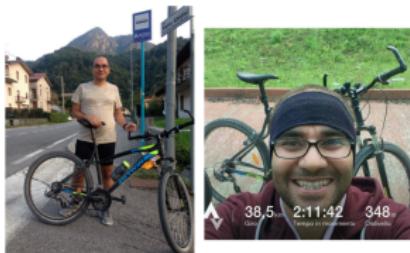
Aliakbar Akbaritabar (Ali)

Hello there!

I am Ali (the simplest version of Aliakbar Akbaritabar, believe me, I have tried other versions!). Below you find pictures of me in my happiest moments. Those moments normally involve a bike ([see my Strava](#)) and my good friends!

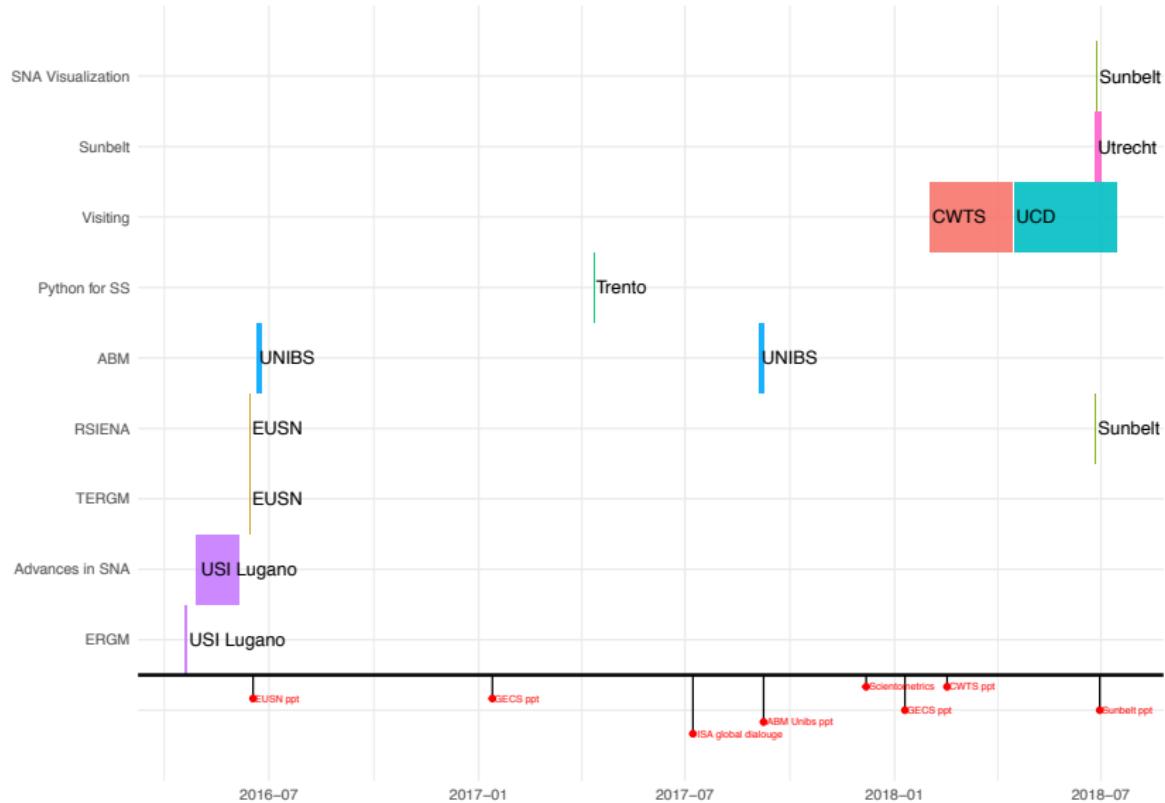
I like to call myself a **Social Data Scientist** (I don't know when it is going to officially become a thing, till then, it is mainly what data scientists are doing plus a sociological approach which is what I have been studying for most of my academic life).

You can find a detailed **CV** of me [here](#) and introductions to *Research Projects & Ideas* I have been working on [here](#)



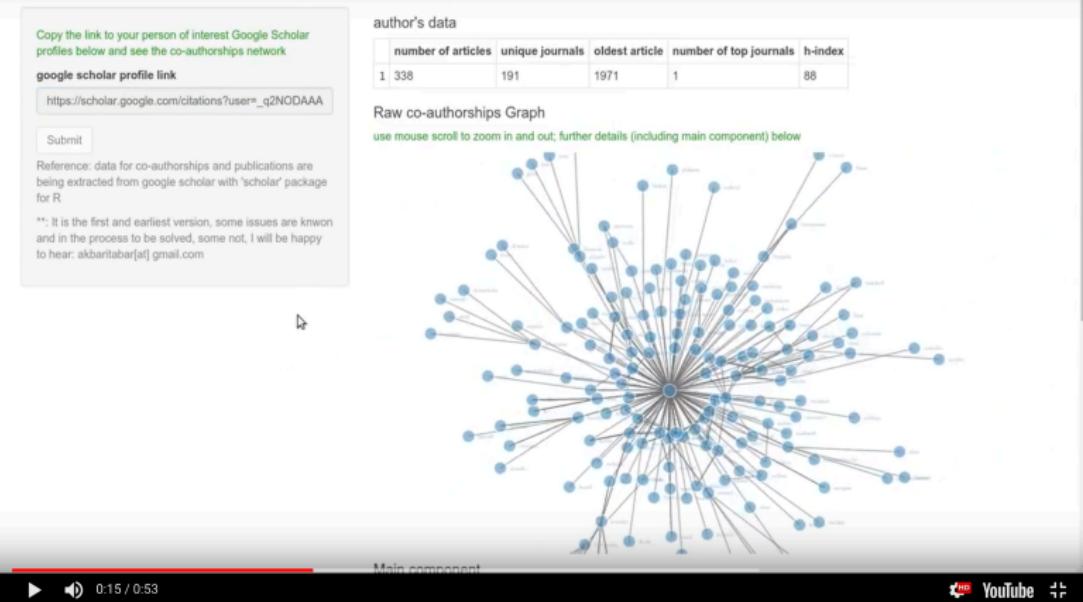
Find me here: akbaritabar.github.io

Activities 2016-2018



R & shiny application for google scholar data crawl and netowrk analysis

Who are researchers mostly co-authoring with?



Brief Shiny application to see google scholar coauthorships network
+ R package to extract paper/author network from Scopus and add numID

1st paper: “The conundrum of research productivity. A study on sociologists in Italy”

- ▶ **How is the situation of our community ?**
- ▶ **Is it a matter of embeddedness?**

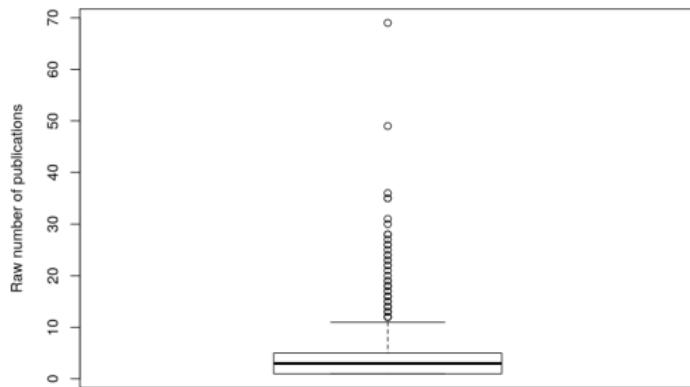
Find it here on Scientometrics:
[https://link.springer.com/article/10.1007/
s11192-017-2606-5](https://link.springer.com/article/10.1007/s11192-017-2606-5)

Hierarchical linear modeling

- ▶ Examining the importance of institutional embeddedness
- ▶ Each scientist is nested in different clusters
- ▶ Three clustering levels:
 - ▶ department; promotion and careers are eventually decided locally
 - ▶ scientific disciplinary sector
 - ▶ university, important in Italy after the last reform of the national university system in 2010; governance for incentivising and measuring scientist productivity, e.g., by establishing rewards for research excellence and allocating internal resources on productivity indicators.
- ▶ Following (Snijders and Bosker 1999; Faraway 2005; Zuur et al. 2009), we used hierarchical linear models
 - ▶ for example, an assistant professor of political sociology (i.e., this is the scientific disciplinary sector) who was enrolled in a department of social sciences in a given university could have different intercept (starting point) and slope (growth rate) in a regression model compared to a full professor of economic sociology in the same university and department.

Results

- Highly skewed distribution of publications; in line with (Nygaard 2015; Ramsden 1994; Coile 1977; Ellwein, Khachab, and Waldman 1989).



Box-plot distribution (black line is the median)

Results

Table 1: Comparative table of Multi level regression models

	Research productivity as dependent variable	
	FSS (1)	Total Publications (2)
internationalisation	0.062*** (0.023)	0.099*** (0.025)
Coauthors Similarity	0.042 (0.048)	0.460*** (0.050)
Gender (male)	0.032*** (0.010)	0.033*** (0.010)
Associate professor	-0.011 (0.011)	0.015 (0.012)
Full professor	-0.025** (0.013)	0.035*** (0.014)
Postdoc	0.081*** (0.016)	0.102*** (0.018)
Constant	0.036*** (0.011)	0.015 (0.012)
Observations	456	473
Log Likelihood	389.517	362.026
Akaike Inf. Crit.	-757.034	-702.052
Bayesian Inf. Crit.	-711.686	-656.302

Note:

* p<0.1; ** p<0.05; *** p<0.01

2nd paper: “Evaluating the effects of ANVUR policies on research productivity of Italian sociologists”

- ▶ **Can policies reorient publication patterns against endogeneous forces?**

Table 3: Models comparing trend of publications in fascia A and non fascia A journals after ANVUR

	Contrasting general trends of publications	
	Number of papers in Fascia or non-Fascia journals as dependent variable	
	(1)	(2)
Publications in fascia A journals	0.007 (0.018)	
Publications in non fascia A journals		0.017*** (0.006)
Constant	0.080*** (0.016)	0.033*** (0.006)
Observations	362	645
Log Likelihood	152.887	663.171
Akaike Inf. Crit.	-295.773	-1,316.343
Bayesian Inf. Crit.	-276.315	-1,293.997

Note:

*p<0.1; **p<0.05; ***p<0.01

Gender and ethnic patterns of publication in top sociology journals

(Success stories only!)

Aliakbar
Akbaritabar (Ali),
Flaminio
Squazzoni

Introduction

Data & Method

Results

Conclusions

Aliakbar Akbaritabar (Ali) Flaminio Squazzoni

GECS - University of Brescia

January 10th, 2018

Dataset

Gender and ethnic
patterns of
publication in top
sociology journals

Aliakbar
Akbaritabar (Ali),
Flaminio
Squazzoni

Introduction

Data & Method

Results

Conclusions

- ▶ All AJS and ASR publications (extracted from **Scopus** on 20th January 2017)
- ▶ **Fields:** article title, authors' names and affiliation, and number of citations received.

Journal name	# papers	Sample Starts	Sample Ends
American Journal of Sociology	1153	1946	2016
American Sociological Review	1440	1965	2016
Total number of papers	2593	-	-

Gender imputation!

- ▶ Used authors' **first names**
- ▶ Send automatic requests with R scripts to a database of numerous names extracted from social media profiles (Wais 2016)
- ▶ Simultaneously a research assistant (RA) **hand-coded author gender**
- ▶ Any conflicting attribution case, the RA researched the online profile of authors, whenever available.
- ▶ Cases of differences (*41 out of 2,897 authors*), used the hand-coded gender
- ▶ Missing data in the hand-coded procedure (*22 out of 2,897 authors*), used the automatic gender extracted from API
- ▶ Only 17 out of 2,897 missing genders (0.58%)

Ethnicity imputation!

- ▶ Used another API, based on **US census data** (Khanna, Imai, and Jin 2017)
- ▶ **Last names** of authors, provides a probability distribution of names in: *white, black, Hispanic, Asian* and *other*
- ▶ In parallel, our RA checked each of the 2897 authors' online profiles
- ▶ A photograph of the author exist, **hand-coded the ethnicity**
- ▶ Cases of differences between the two procedures (*132 out of 2,897 authors*), used the hand-coded category
- ▶ Cases of missing data in our hand-coded procedure (*426 out of 2,897 authors*), used the automatic ethnicity extracted from API
- ▶ We ended up without any missing ethnicity

Women: AJS & ASR authors vs. ASA members

Gender and ethnic
patterns of
publication in top
sociology journals

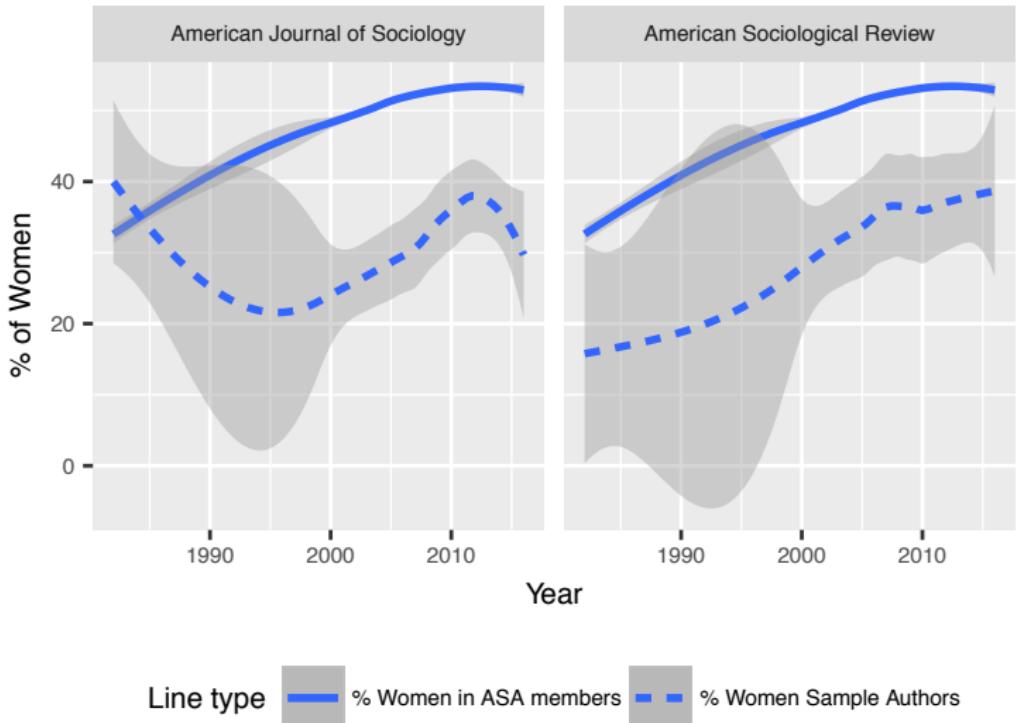
Aliakbar
Akbaratabar (Ali),
Flaminio
Squazzoni

Introduction

Data & Method

Results

Conclusions



Co-authorship patterns

- ▶ **84%** of articles in AJS and ASR had **at least one** (or more) *male* author(s)
- ▶ Only **40%** had at least one (or more) *female* author(s)
- ▶ Picture approximates a 70/30 ratio, which is slightly better than what suggested by Young (1995)'s study in political sciences
- ▶ But similar to what found by Teele and Thelen (2017)
- ▶ It seems that fields such as sociology and political sciences do not dramatically differ in terms of gender patterns.

Journal Name	# All Papers	# All Authors	# Men	% Men	# Women	% Women
AJS	1153	2023	1469	72.61	547	27.04
ASR	1440	2686	1860	69.25	813	30.27
Total number	2593	4709	3329	-	1360	-

Cross-gender co-authorship increased!

Gender and ethnic
patterns of
publication in top
sociology journals

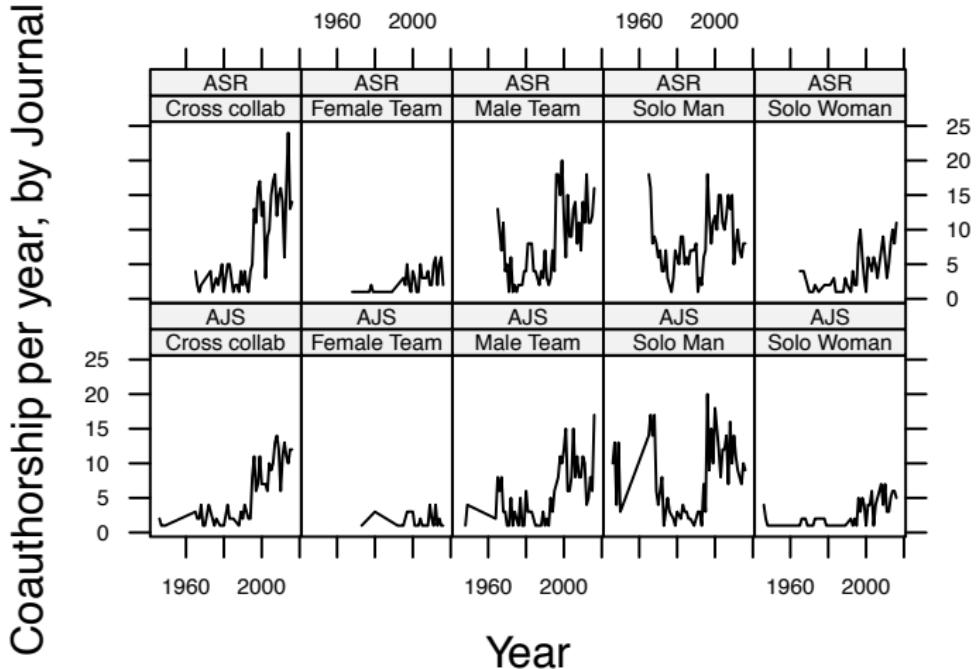
Aliakbar
Akbaratabar (Ali),
Flaminio
Squazzoni

Introduction

Data & Method

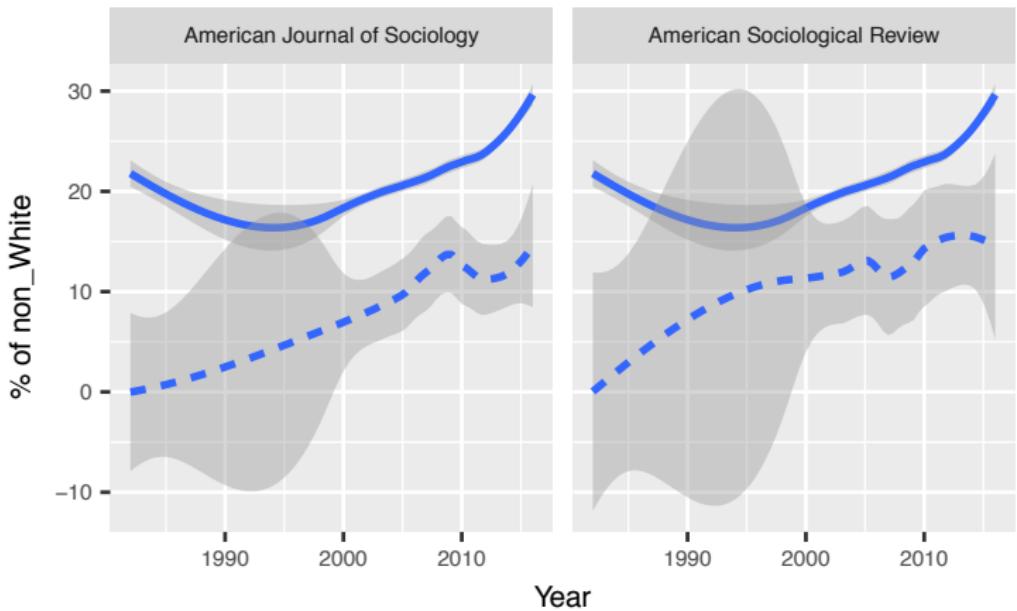
Results

Conclusions



Non-white: AJS & ASR authors vs. ASA members

- ▶ Number of non-white authors **has increased**, but less pronounced than non-white members in ASA (a maximum 30% of members)
- ▶ The ratio was below 20% in AJS and below 25% in ASR.



Gender and ethnic
patterns of
publication in top
sociology journals

Aliakbar
Akbaratabar (Ali),
Flaminio
Squazzoni

Introduction

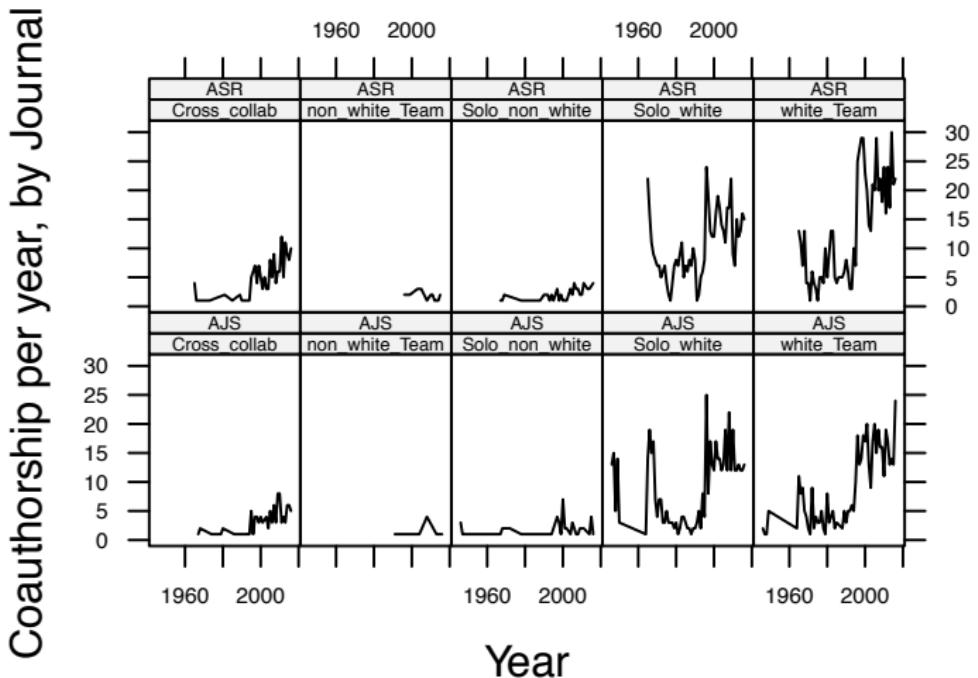
Data & Method

Results

Conclusions

Ethnic co-authorship dynamics in AJS and ASR

- Ethnic gap in the academic élite seems to persist for a long time



Introduction

Data & Method

Results

Conclusions

Comparative models of number of publications

Gender and ethnic
patterns of
publication in top
sociology journals

Aliakbar
Akbaritabar (Ali),
Flaminio
Squazzoni

Introduction

Data & Method

Results

Conclusions

	Research products as dependent variables		
	Total Publications (1)	Publications before 2000 (2)	Publications after 2000 (3)
Gender Male	0.021*** (0.004)	0.192*** (0.064)	0.260*** (0.059)
Ethnicity Black	-0.006 (0.017)	0.108 (0.285)	-0.209 (0.222)
Ethnicity Hispanic	0.009 (0.021)	0.563* (0.325)	-0.086 (0.276)
Ethnicity White	0.020*** (0.007)	0.245** (0.124)	0.174* (0.091)
Constant	0.014 (0.012)	1.051*** (0.151)	1.250*** (0.120)
Observations	2,716	1,293	1,708
Log Likelihood	2,212.361	-1,842.347	-2,706.775
Akaike Inf. Crit.	-4,408.723	3,700.694	5,429.549
Bayesian Inf. Crit.	-4,361.468	3,742.012	5,473.094

Note:

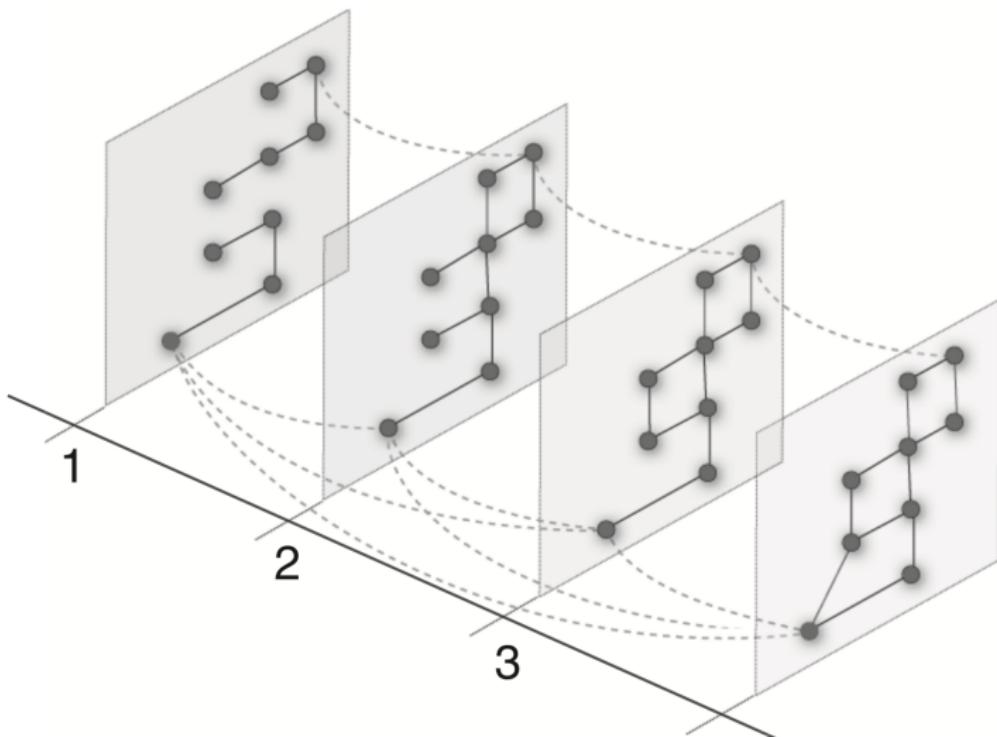
* p<0.1; ** p<0.05; *** p<0.01

Detecting invisible colleges

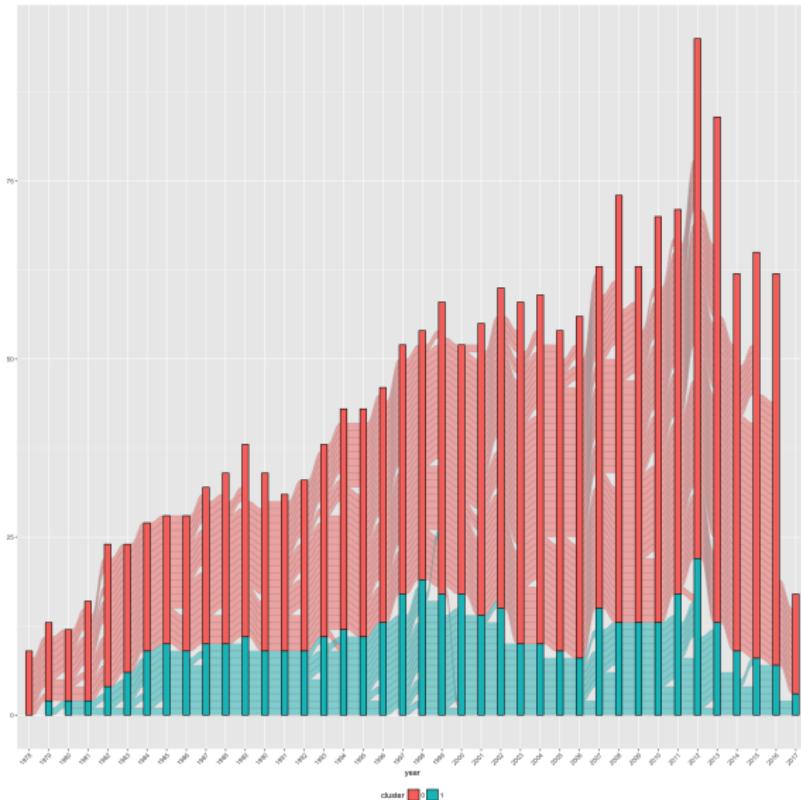
- ▶ One of the long standing dreams of sociologists have been to find a way to *detect, explore, explain* and *track schools of thoughts* or in other words **cohesive sub-groups of scientific community** which are more tied **within themselves** than **between** their group and others.
- ▶ To see:
 - ▶ Why and How are they formed?
 - ▶ How are these cohesive groups collaborating together?
 - ▶ How do they evolve through the time?
 - ▶ What can be the **glue** which attaches them together in terms of **social solidarity and cohesion** which detaches them from other parts of the community (or simply makes them more connected to each other and preserve weak ties with other parts of community).

- ▶ Present Research communities (invisible colleges) detected!
 - ▶ External files (zoom possibility)
- ▶ How temporal community detection works?

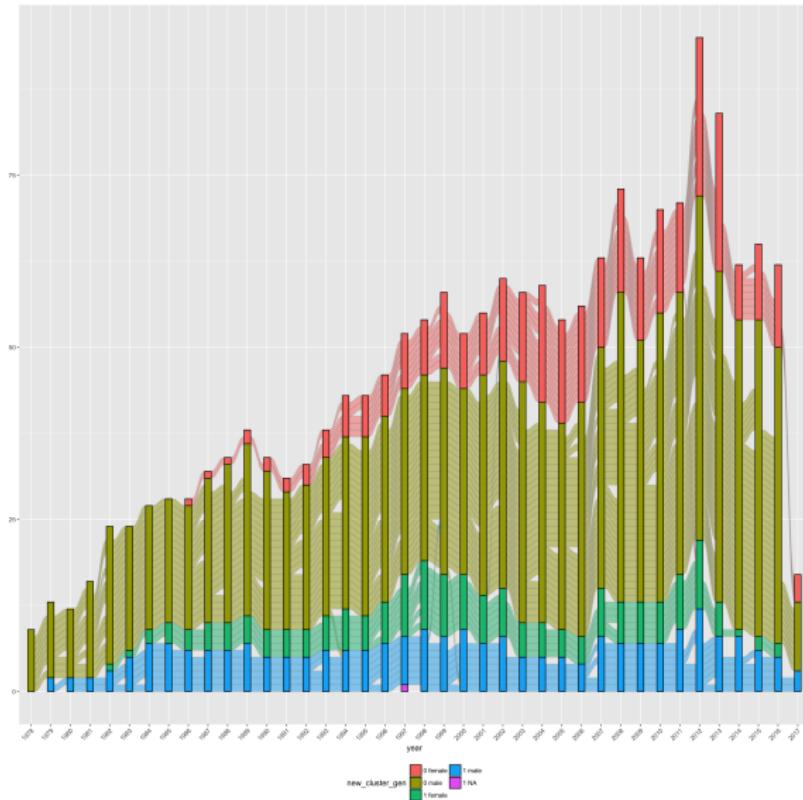
- Photo credit: Mucha, et.al (2010). Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980), 876-878.



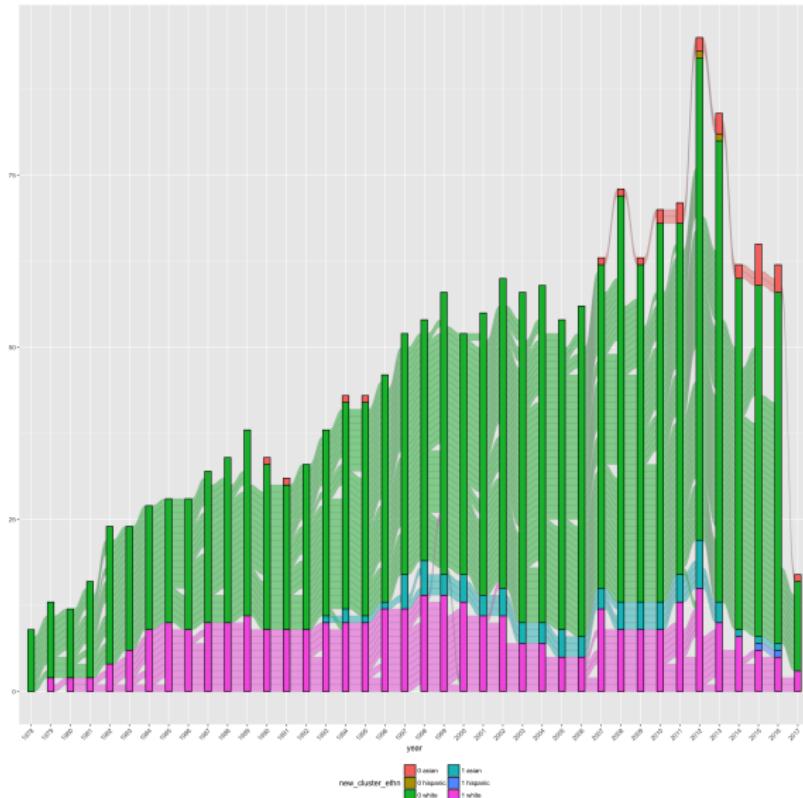
TCD Results (SN papers)



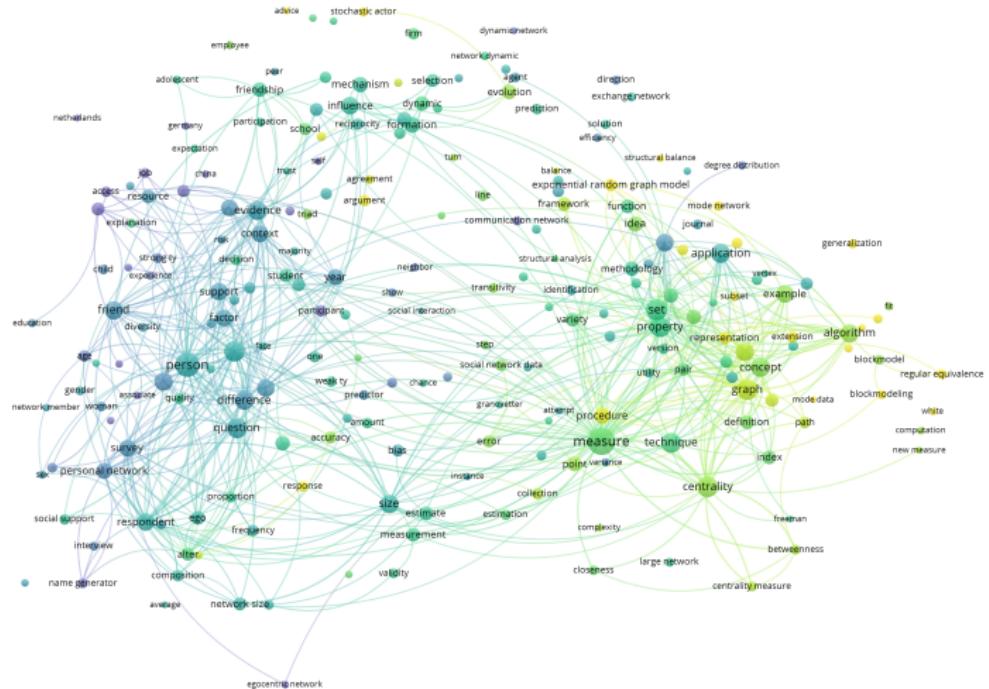
TCD Results (SN papers vs gender)



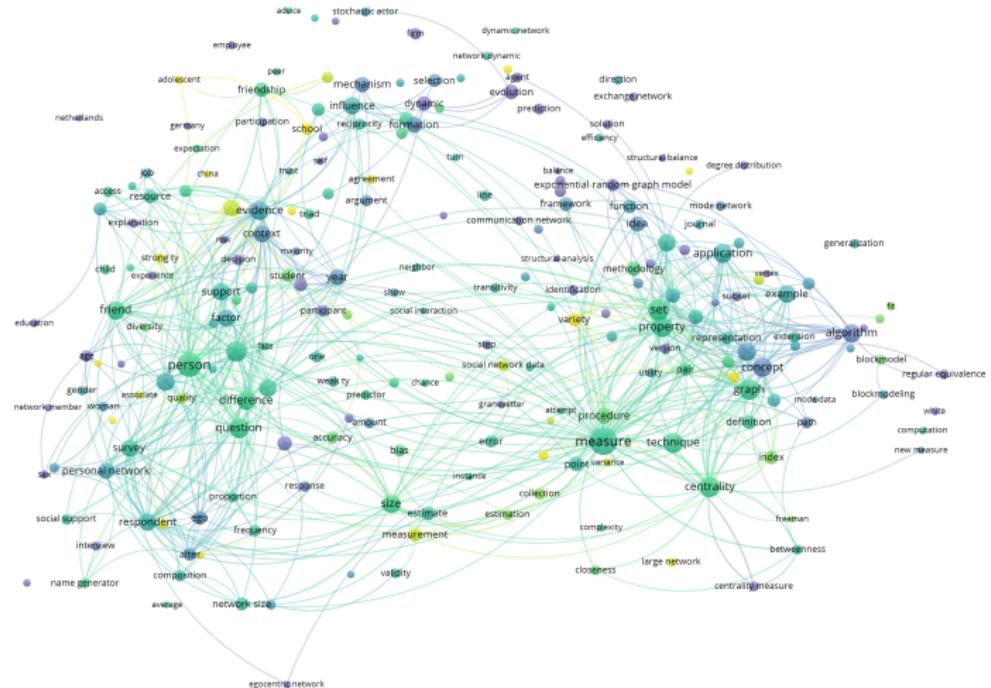
TCD Results (SN papers vs ethnicity)



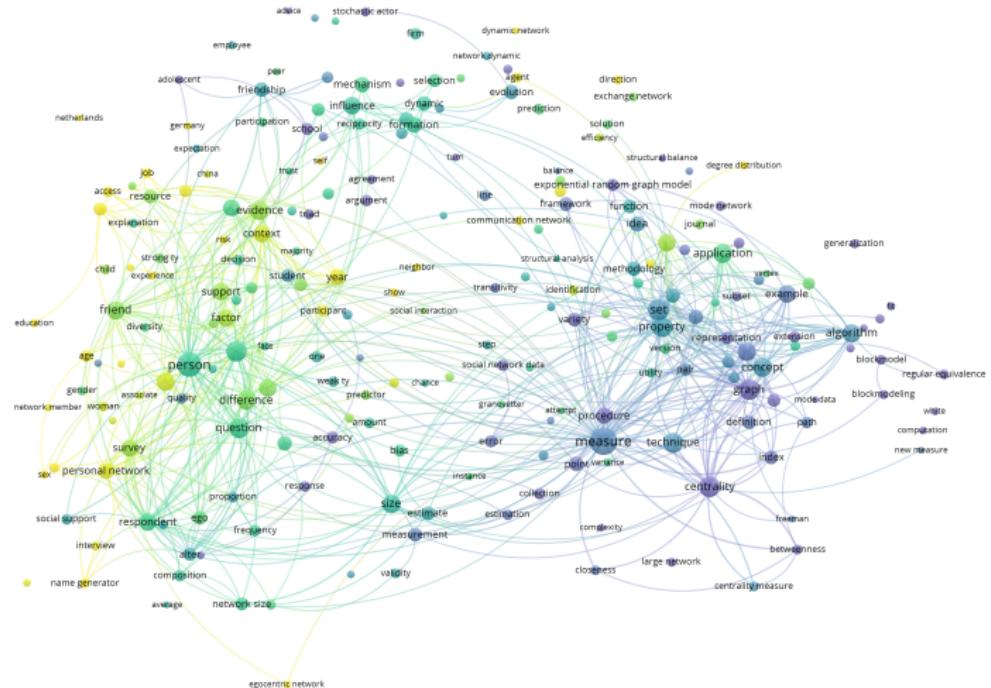
Term map overlays (SN papers cluster 0)



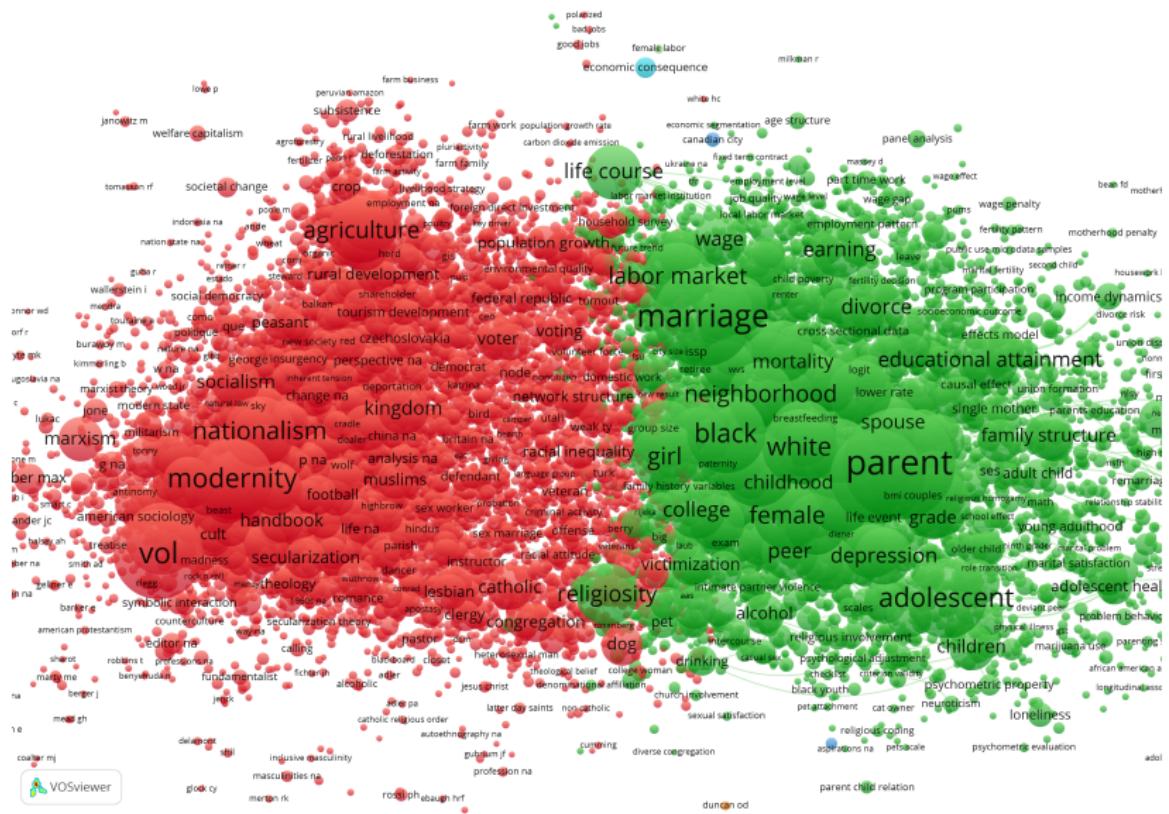
Term map overlays (SN papers cluster 1)



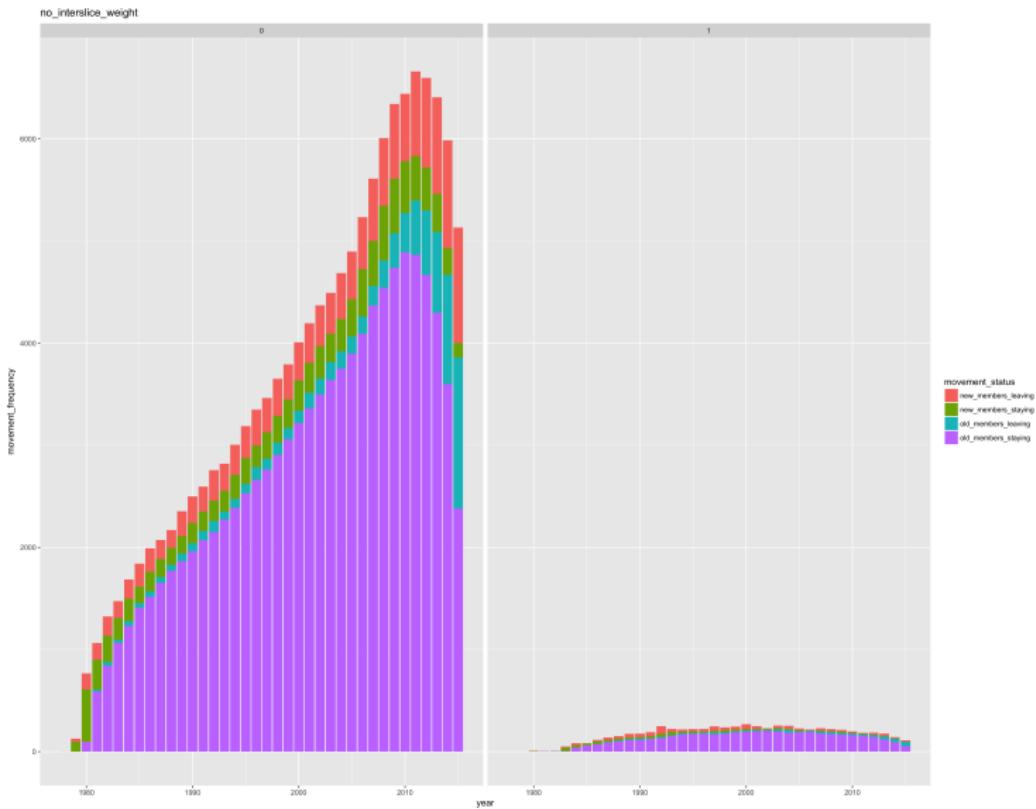
Term map overlays (SN papers cluster NA)



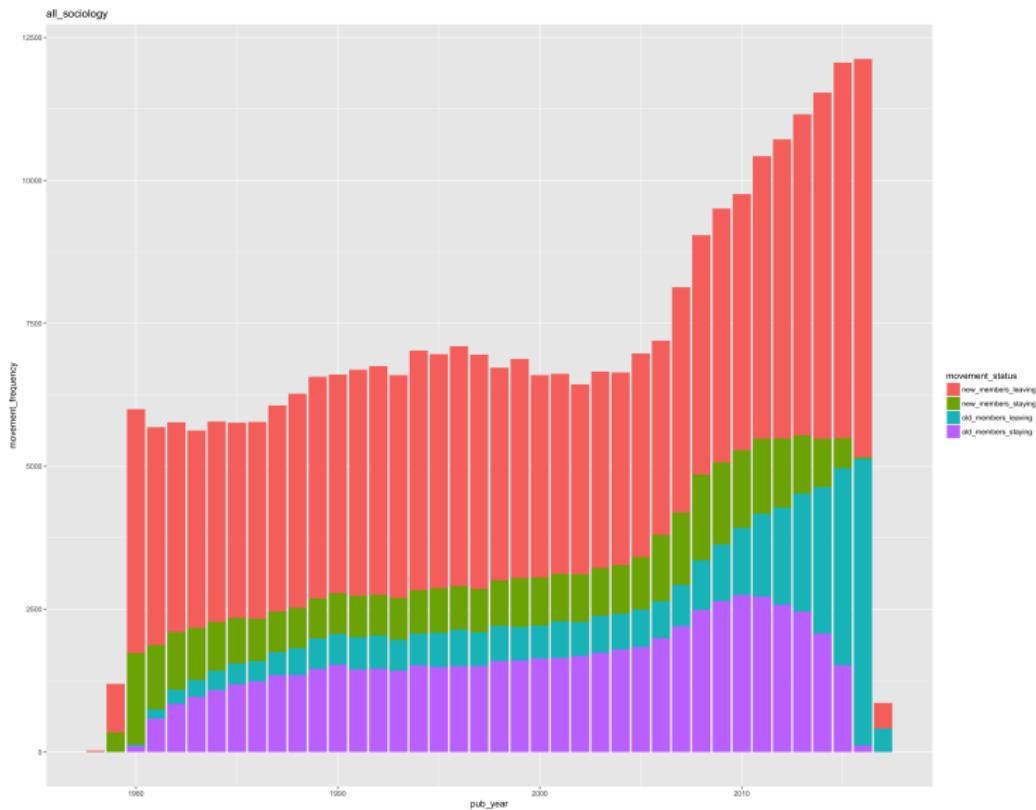
All Sociology term clusters



Giant Component Evolution



All Sociology Evolution



What I would like to do in next 3 years?!

- ▶ Do research that matters and has empirical implication on the life of those being studied
- ▶ Do solution oriented social data science!
- ▶ Empower my CV
- ▶ Learn new things (constant improvement)

What I have done recently!

- ▶ Need in society for social scientists doing high quality science, keeping the applications and outcomes of this research work in the core of their perspective.
- ▶ Solution oriented social science
- ▶ 1st PhD thesis, studying *people's life and their social capital on social media* and how they were using it to provide and receive social support
- ▶ Designed and developed an online tool, [Social Village](#)
- ▶ 2nd PhD thesis on *scientific collaborations*, my ambition was to develop an online tool to provide real-time insights to scientists about how they have been doing science throughout their career. (*see an early version of the tool here*).
- ▶ In other words, it can give insights to researchers, while giving them voice to take part in their research evaluation by submitting their feedback to the reports they see while using this research toolbox.

Related Experiences to scientometrics & academia

- ▶ Research productivity of Italian sociologists and gender and ethnic patters of publications in top sociology journals can be helpful for *contextualized scientometrics*
- ▶ My background in social media studies can serve me well in *altmetrics* area
- ▶ Experience in systematically gathering data with R scripts from Scopus API and building an integrated database and write it on a web hosted My-SQL DB or local SQLite DB.
- ▶ My efforts on quantitative content analysis on title, keywords and abstract of the papers published in samples mentioned above with structural topic modelling and theme extraction can help me relate to *full-text analysis*
- ▶ The tool I have developed based on Google Scholar data described above and an R package I have developed for **co-authorship network based literature review (CNBLR)** with a shiny web interface to make literature review easier and smoother
- ▶ Organized, highly motivated and dedicated researcher.
- ▶ Worked in organizational contexts which required meeting deadlines
- ▶ Providing targeted reports for both academic publications and less specialist audience (e.g., blog posts and organizational magazines).
- ▶ Experience in organizing **applied workshops** on how to formulate research questions, gather the suitable data, steps in using the software to analyze
- ▶ **Write reproducible report** with a social network analysis approach
- ▶ Previous university teaching experience

**Thanks a lot for your attention
I will be happy to hear your comments/suggestions**
Email: akbaritabar@gmail.com