

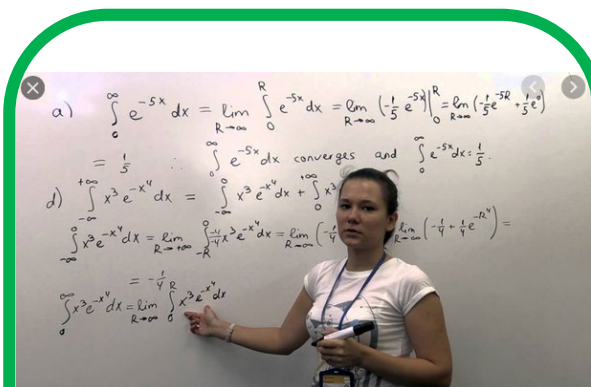
Bibliometric analysis on mathematics

3 snapshots: 2005, 2010, 2015

S. Richard, Q. Sun

Content:

- I) General picture
- II) Individual predictors and the response
- III) Best predictor and predictions
- IV) About countries
- V) Conclusion



Galina Levitina

Cwikel-type estimates on open domains

(Himeji 2022)

I) General picture

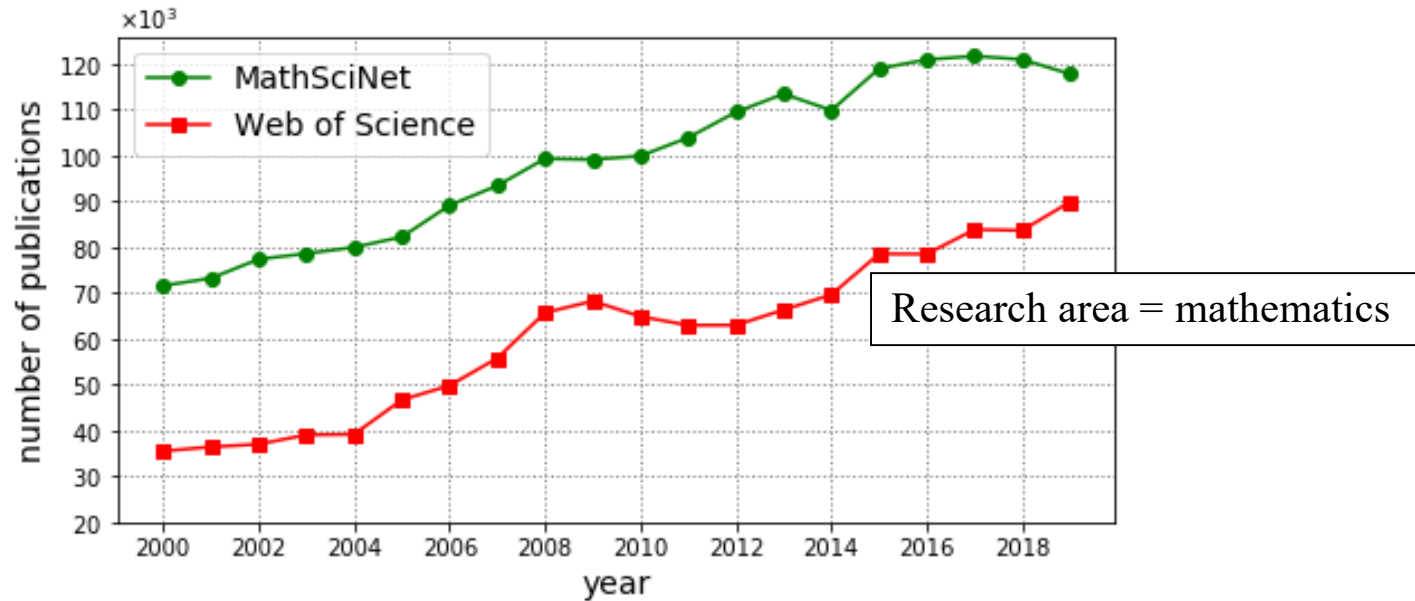
Original question:

What are the important predictors for the publications in mathematics,
if the response is the number of citations ?

Previous works:

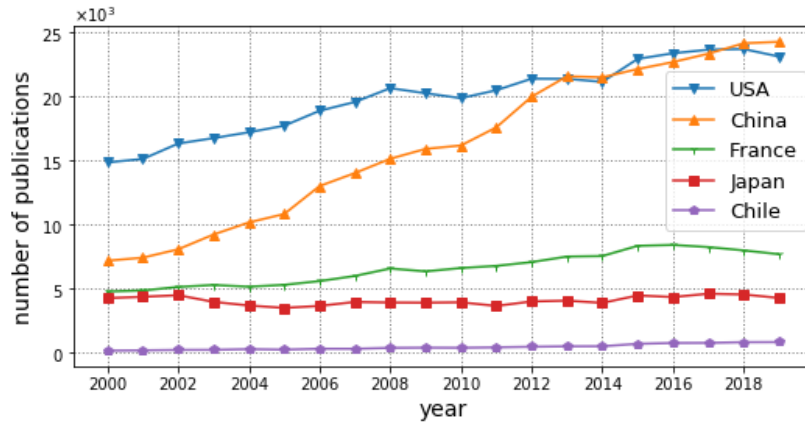
- Several on different fields, but not centred on mathematics
- Grossman (2002, 2005) about research collaboration, with MathSciNet
- Bensman et al. (2010) about citation indices for math journals
- Behrens and Luksch (2011) bibliometric analysis on the period 1868 - 2008
- Özkaya (2018) bibliometric analysis but about math education
- Szomszor et al. (2020) Individual perspective about self-citations
- Verma et al. (2021) bibliometric analysis over 40 years for a single journal
- Dunne (2021) On individual citations, **based on MathSciNet**
- Paik and Rivin (preprint) on US mathematics faculties with some bibliometric tools

Which database ?

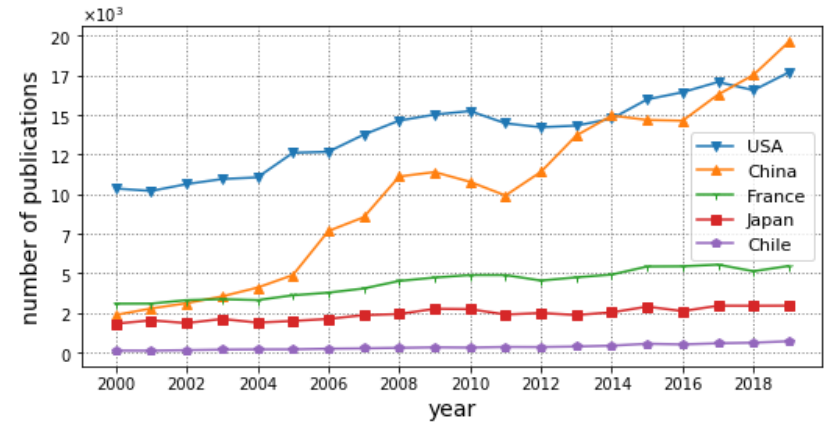


Yearly new indexed math publications

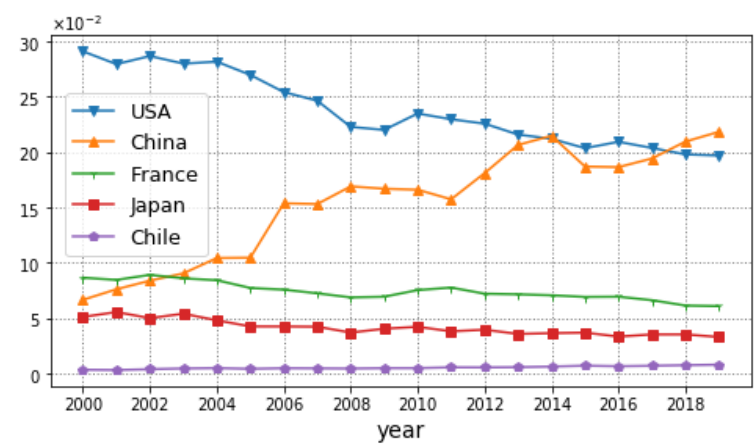
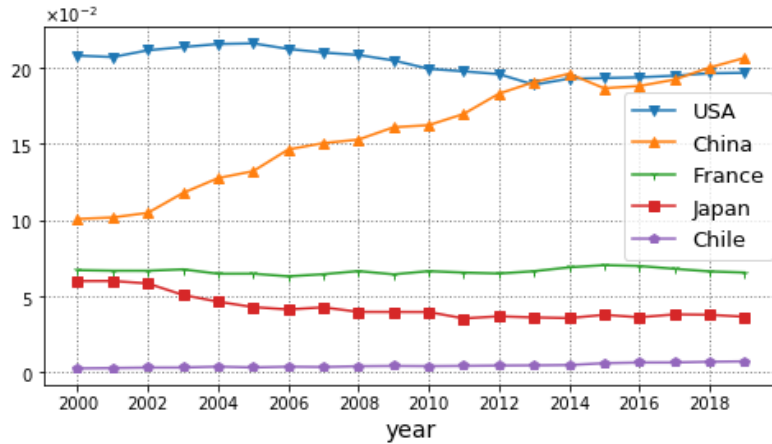
MathSciNet



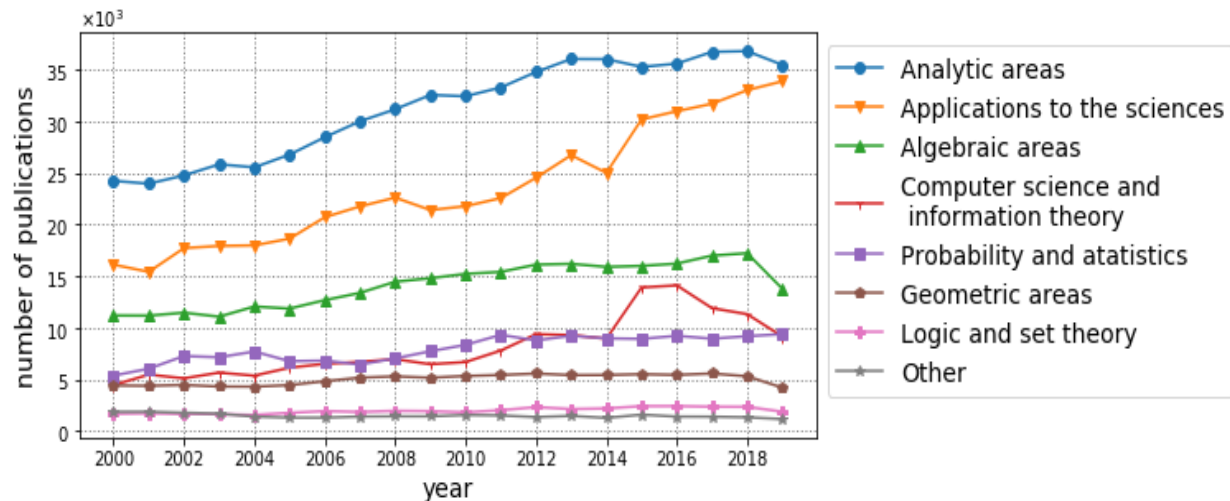
Web of Science



Publications with at least one author from a given country,
Absolute and relative numbers



MathSciNet



More accurate
Centred on mathematics
No automatic downloading
No robot allowed

Web of Science

Nothing
similar

Less accurate
Mathematics in a broad sense
Massive downloading possible



Raw data

Research area = mathematics, clear author(s) and affiliation(s), citation < 64

2005: 42'792 items

2010: 61'084 items

2015: 76'168 items

Predictors:

authors

institutes

countries

references

pages

keywords

open access (Y/N)

JIF

research area

category

Response:

citations

Simple statistics: mean number / publication			
	2005	2010	2015
Authors	2.10	2.23	2.39
Institutes	1.57	1.74	1.90
Countries	1.23	1.28	1.32
References (0 excluded)	19.0	21.7	24.5
Pages	15.5	15.5	16.9
Keywords (0 excluded)	4.38	4.48	4.58
Open Access (by WoS)	16.4%	21.2%	32.3%
Citations	10.0	8.5	5.3

II) Individual predictors and the response

authors:

	2005			2010			2015		
	%	mean	median	%	mean	median	%	mean	median
1 author	35.3	7.8	3	30.4	5.9	2	25.6	3.4	1
2 authors	36.6	10.8	6	36.5	9.0	5	35.9	5.1	2
≥ 3 authors	28.1	11.8	6	33.1	10.3	6	38.5	6.7	3



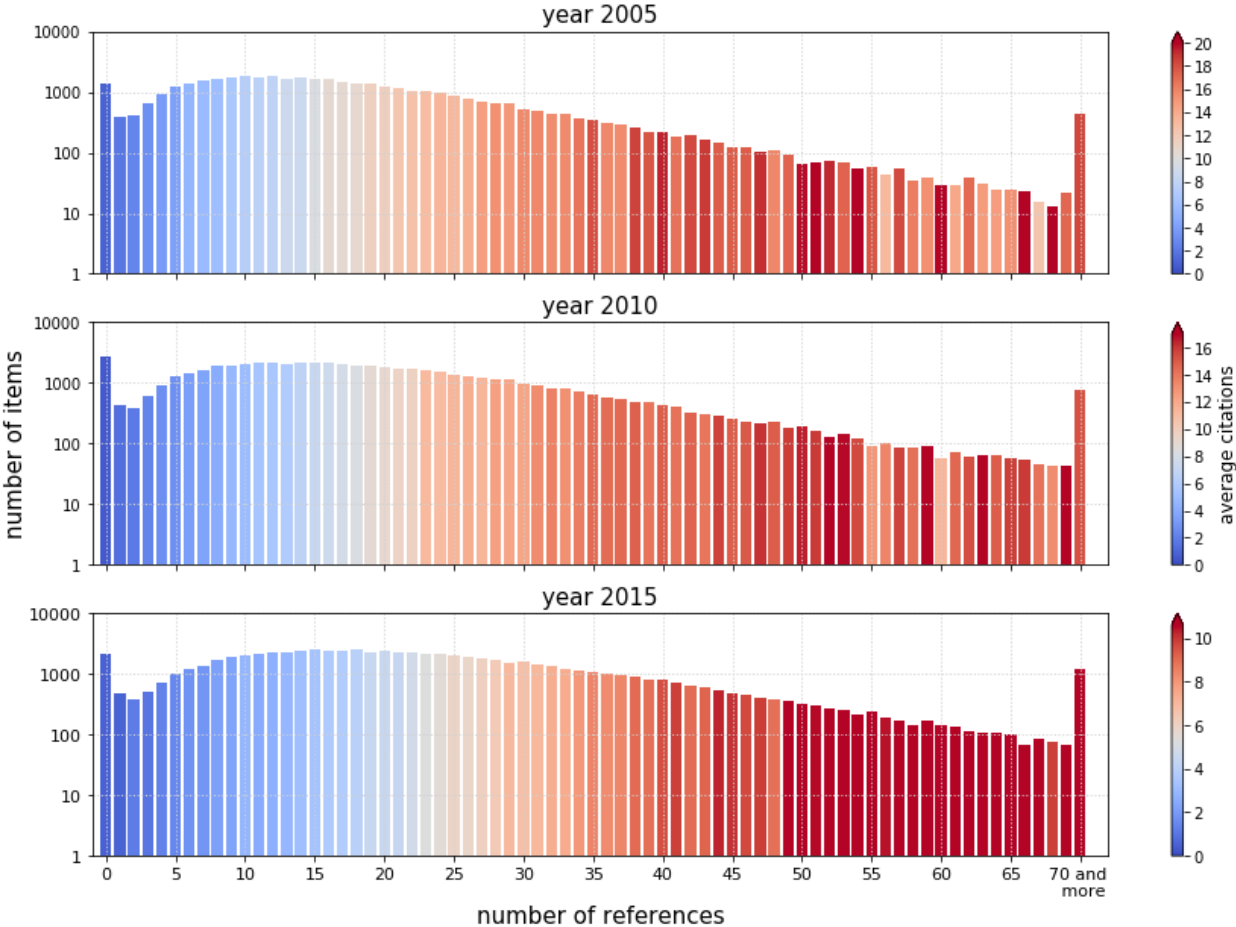
Number of citations

countries:

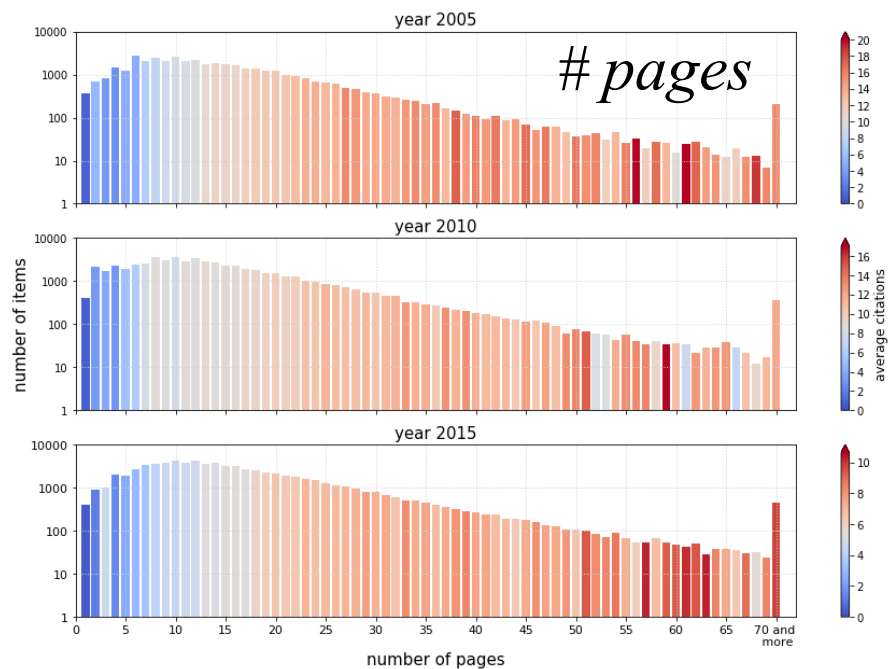
	2005			2010			2015		
	%	mean	median	%	mean	median	%	mean	median
1 country	80.0	9.2	4	76.5	7.7	3	73.7	4.7	2
2 countries	17.4	13.5	8	19.8	10.8	6	21.4	6.6	3
≥ 3 countries	2.6	14.5	10	3.7	12.3	7	4.9	8.3	5

references:

Log
scale



Centred on
the mean:
2005: 10
2010: 8.5
2015: 5.3



keywords

	2005			2010			2015		
nb keywords	%	mean	median	%	mean	median	%	mean	median
0	36.6	8.3	3	28.7	6.7	2	23.7	4.4	1
1	0.3	7.7	3	0.3	4.4	2	0.2	3.1	1
2	4.3	8.1	4	3.9	6.8	3	3.0	3.4	2
3	15.9	9.8	5	16.7	7.6	4	16.2	4.3	2
4	17.6	10.9	6	20.1	8.7	4	21.9	5.4	3
5	13.1	11.5	7	15.8	10.3	6	18.6	6.2	3
6	6.4	13.3	8	7.9	11.2	6	9.5	7.0	4
7	2.7	12.9	7	3.2	10.7	6	3.3	6.3	3
8	1.3	14.8	10	1.7	11.3	7	1.7	6.2	3
9	0.7	15.3	11	0.8	11.5	7	0.9	6.9	4
10	0.4	14.3	9	0.4	12.0	6	0.5	6.0	3
≥ 11	0.5	15.3	11	0.5	11.2	7	0.6	6.6	4

Open access

	2005			2010			2015		
	%	mean	median	%	mean	median	%	mean	median
OA	16.4	12.2	7	21.2	9.7	5	32.3	5.7	3
no OA	83.6	9.6	5	78.8	8.1	4	67.7	5.1	2

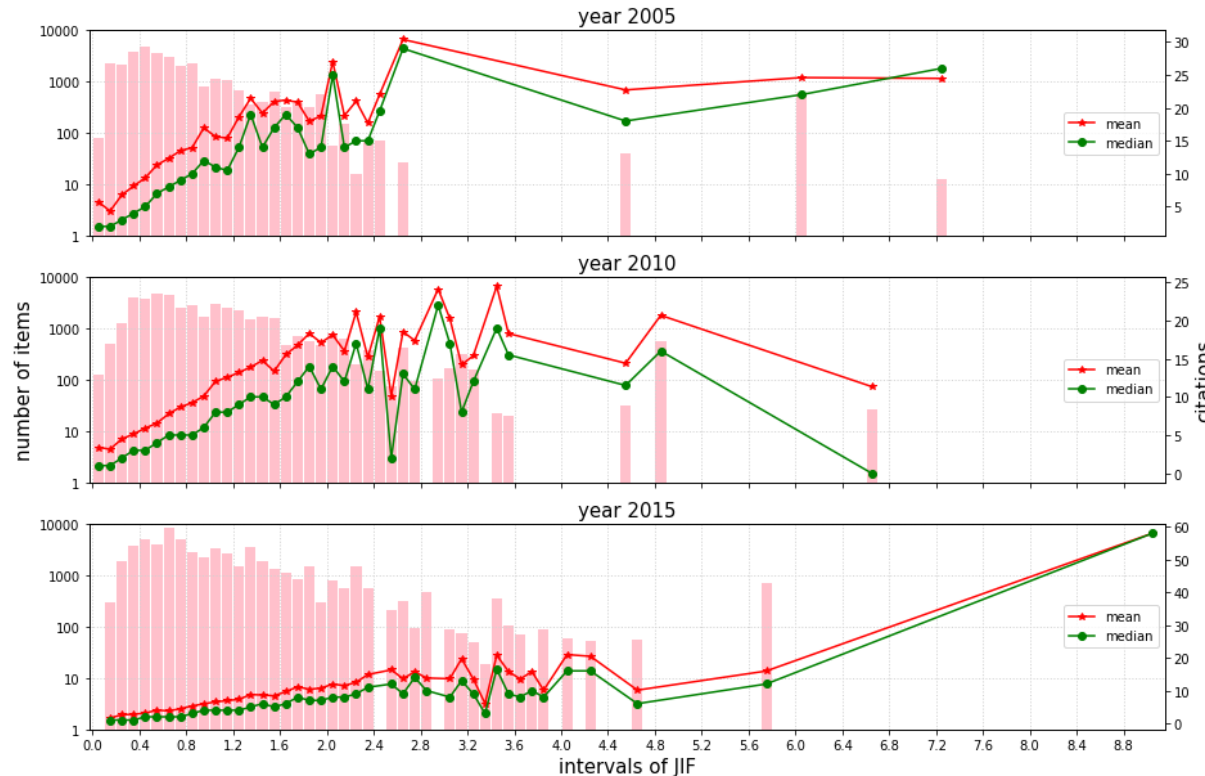
Journal Impact Factor

Research area = mathematics, clear author(s) and affiliation(s), citation < 64, **JIF available**

2005: 35'556 items

2010: 44'639 items

2015: 57'756 items



Citation mean:

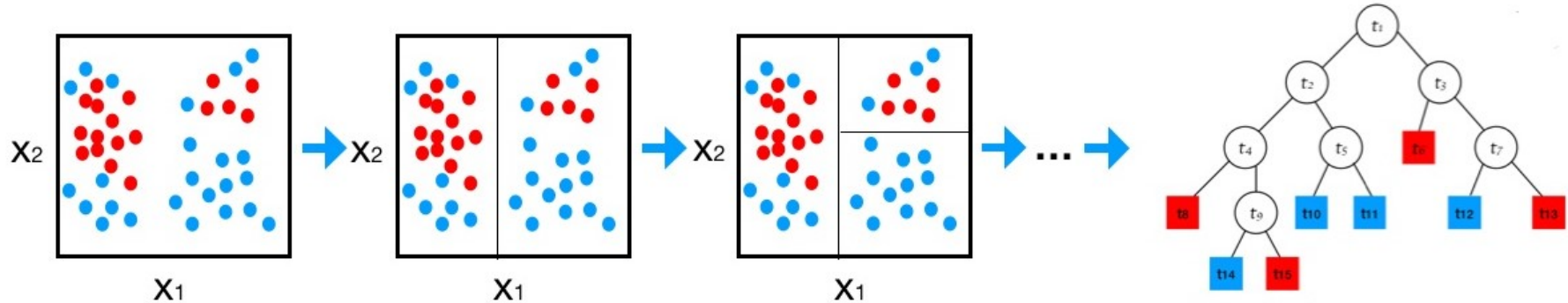
2005: 12.0

2010: 10.2

2015: 6.3

III) Best predictor and predictions

Tree classifier: 10 predictors, response partitioned in J classes



- 1) Generate best splits by minimizing an impurity function
- 2) Leaves labeled by majority of class J items
- 3) Reduce the size of the tree by pruning, best size obtained with training set
- 4) **Best predictor** generates maximal decay of impurity

Predictions: Test new items on the best tree

The outcome:

Free choice

Some predictors

Content matters
more than
bibliometric
data

	2005		2010		2015	
JIF	No	Yes	No	Yes	No	Yes
<u>classes</u>	[0,5] [6,12] [13,20] [21,40] [41,63]	[0,6] [7,14] [15,23] [24,45] [46,63]	[0,4] [5,10] [11,20] [21,63]	[0,5] [6,12] [13,24] [25,63]	[0,3] [4,9] [10,63]	[0,4] [5,11] [12,63]
# leaves	24	15	27	29	49	25
height	7	6	8	6	10	6
av. train acc. (%)	32.9	34.5	39.3	40	51.3	51.5
av. test acc. (%)	31.3	30	38.4	37.8	50.4	50.7
references	1	2	1	2	1	2
authors		4	3	3	3	4
countries						3
institutes	2					
pages	4	3	4	4	2	
jif		1		1		1
research area	3 (ENG.)		2 (ENG.)		4 (Bio. & A. M.)	
Experiment (%)	39	40	44	43	53	53
Random guess:	20 %		25 %		33 %	

IV) About countries

1) Relative importance of each country (at least 1 author from a given country)

	2005					2010					2015				
	%	h.c.	mn	md	r.c.	%	h.c.	mn	md	r.c.	%	h.c.	mn	md	r.c.
US	27.7	40.6	12.0	7	1.2	24.0	37.3	10.1	5	1.2	20.7	32.9	6.5	3	1.2
CN	10.8	12.8	11.0	6	1.1	17.1	17.2	8.6	4	1.0	19.1	29.5	6.5	3	1.2
FR	8.0	9.3	12.0	7	1.2	7.7	7.1	10.2	6	1.2	7.0	7.4	6.0	3	1.1
DE	7.3	8.7	11.2	6	1.1	7.3	7.7	8.9	5	1.0	6.9	10.6	6.1	3	1.2
IT	5.5	4.8	9.9	5	1.0	5.1	4.2	9.4	5	1.1	5.1	7.4	6.6	4	1.3
UK	5.3	7.8	12.2	7	1.2	5.0	8.1	10.7	6	1.3	4.7	11.1	6.6	4	1.2
CA	4.6	5.0	11.2	6	1.1	4.2	5.2	9.8	5	1.2	3.5	3.7	5.7	3	1.1
JP	4.4	2.5	8.5	4	0.8	4.3	2.1	6.4	3	0.8	3.7	2.7	4.2	2	0.8
ES	4.1	3.4	11.0	6	1.1	4.3	4.2	8.9	5	1.1	3.7	3.2	5.8	3	1.1
RU	4.1	1.1	6.2	2	0.6	3.8	0.9	5.6	2	0.7	4.8	1.4	3.8	2	0.7
AU	2.9	3.3	9.7	4	1.0	2.1	3.1	9.7	5	1.1	2.5	4.0	6.1	3	1.2

% = percentage
of publications

mn = citation
mean

md = citation
median

r.c. = relative
citation

h.c. = highly cited. Percentage of publications having
more than 63 citations

2) International collaborations (at least 2 countries involved)

	2005					2010					2015				
	%	mn	md	rcc	rc2	%	mn	md	rcc	rc2	%	mn	md	rcc	rc2
US	30.1	14.4	9	1.2	1.0	34.9	12.3	7	1.2	1.0	42.4	7.5	4	1.2	1.0
CN	23.7	14.1	9	1.3	1.2	21.2	12.4	7	1.4	1.4	25.5	8.6	5	1.3	1.3
FR	38.7	14.4	9	1.2	1.1	48.1	11.5	7	1.1	1.0	54.7	6.6	4	1.1	1.0
DE	40.1	13.6	9	1.2	1.1	47.6	10.8	7	1.2	1.1	51.8	7.1	4	1.2	1.0
IT	33.7	13.6	8	1.4	1.2	42.4	11.5	7	1.2	1.1	49.5	7.8	5	1.2	1.1
UK	46.8	14.3	9	1.2	1.0	54.0	12.3	7	1.1	1.0	61.5	7.3	4	1.1	1.0
CA	51.2	13.6	8	1.2	1.1	53.7	11.4	7	1.2	1.1	60.8	6.4	3	1.1	1.0
JP	23.3	12.3	8	1.4	1.2	30.3	8.7	5	1.4	1.2	33.3	6.3	3	1.5	1.3
ES	36.0	14.2	10	1.3	1.2	46.2	10.5	6	1.2	1.1	54.4	6.4	4	1.1	1.1
RU	29.3	11.5	7	1.8	1.5	30.3	9.0	5	1.6	1.2	25.0	5.9	3	1.5	1.3
AU	40.9	14.2	9	1.5	1.4	56.8	11.2	6	1.2	1.1	57.7	8.2	4	1.3	1.2

% = percentage
of publications

mn = citation
mean

md = citation
median

rcc = relative citation for international collaborations. Ratio of mn by the citation mean over all publications of this country with less than 64 citations

rc2 = relative citations for international collaborations / 2 authors. Ratio of mn by the citation mean over all publications of this country with less than 64 citations and at least 2 authors.

3) Bilateral collaborations

For a country in x-coordinate,
the numbers correspond to the %
of its international collaborations
with a country of the y-coordinate

2005
2010
2015

	US	CN	FR	DE	IT	UK	CA	JP	ES	RU	AU
US		36.7	23.9	24.2	21.6	27.9	37	27.2	25.1	20.7	26.9
		37.5	21.5	22.8	22.2	22.3	34.6	18.5	16.6	19.3	19.7
		40.1	18.8	22.7	19.9	25	32.7	22.3	15.6	18.8	19.5
CN	11.3		3.7	5.3	3.3	6.4	13.5	13.1	2.2	0.4	15.2
	16.1		5.6	5.3	3.5	7.7	13	18	2.9	3.1	22
	22.2		6.7	6.8	3.1	10.3	18.7	16.7	4.2	4.1	26.8
FR	8.8	4.5		8.2	16.8	7.2	8	6.1	12.5	13.8	3.2
	9.5	5.7		10	17.9	8.8	9.3	6.7	12.4	13.8	7.7
	8.2	5.3		9.9	15.2	9.6	10	9.5	11.7	11.4	4.8
DE	8.4	6.1	7.8		11.5	10.1	5.4	13.1	5.8	14.6	6.1
	9.4	5.1	9.3		11.7	11.7	5.5	8.5	8.2	12.6	8.8
	9.3	5	9.2		12.1	11.6	6.2	8.8	7.5	12.2	6.1
IT	4.8	2.4	10.2	7.3		5.4	2.6	4.8	6.4	10.1	2.7
	5.7	2.1	10.4	7.3		5.8	3.7	4.4	10.1	6.7	4.8
	5.7	1.6	10	8.5		8.8	6.7	5	8.9	8.5	3.5
UK	8.2	6.1	5.8	8.5	7.2		5.5	3.7	5.5	10.7	15
	7.1	5.7	6.4	9.1	7.3		6.6	7.3	6.2	8.4	11.8
	8.2	6.1	7.3	9.3	10.1		7.6	6.1	6.7	8.3	12.8
CA	10.4	12.4	6.1	4.3	3.2	5.2		6.1	4.6	5.7	7.8
	9.3	8.1	5.7	3.6	3.8	5.5		5.5	4.3	3.8	8.2
	8	8.3	5.6	3.7	5.7	5.6		4.4	3.2	2.6	5.8
JP	3.3	5.2	2	4.6	2.6	1.5	2.6		1.6	2	2.5
	2.9	6.5	2.4	3.2	2.7	3.5	3.2		1.9	4.1	2
	3.2	4.3	3.1	3.1	2.5	2.6	2.5		1.9	2.6	2.7
ES	4.5	1.3	6.1	3	5.1	3.3	2.9	2.4		4.4	3.4
	3.9	1.6	6.6	4.7	9.2	4.5	3.7	2.8		3.2	3.5
	3.6	1.8	6.2	4.3	7.2	4.7	3	3		4	3.4
RU	3	0.2	5.4	6	6.6	5.2	2.9	2.4	3.6		2.8
	2.6	1	4.2	4.1	3.5	3.5	1.9	3.5	1.9		2
	2.6	1	3.6	4.1	4.1	3.5	1.5	2.5	2.4		2.4
AU	3.8	6.9	1.2	2.4	1.7	7.1	3.9	2.8	2.7	2.8	
	2.7	7.1	2.4	3	2.6	5.1	4.2	1.8	2.1	2.1	
	3.1	7.8	1.8	2.4	2	6.3	3.8	3	2.4	2.8	

V) Conclusion

1) Everything increase

	2005	2010	2015
# publications	45'035	62'945	76'788
# authors	2.10	2.23	2.39
# institutes	1.57	1.74	1.90
# countries	1.23	1.28	1.32
# references	19	21.7	24.5
# pages	15.5	15.5	16.9
# keywords	4.38	4.48	4.58
open access (%)	16.4	21.2	32.3

2) Predictors can not predict so much

3) Good practices are useful

4) International collaborations are beneficial

5) MathSciNet should allow such investigations



is useful and should be praised

Thank you