

Stefano Chessa

Pisa
13th March 2019

Scientific papers ... & their performances

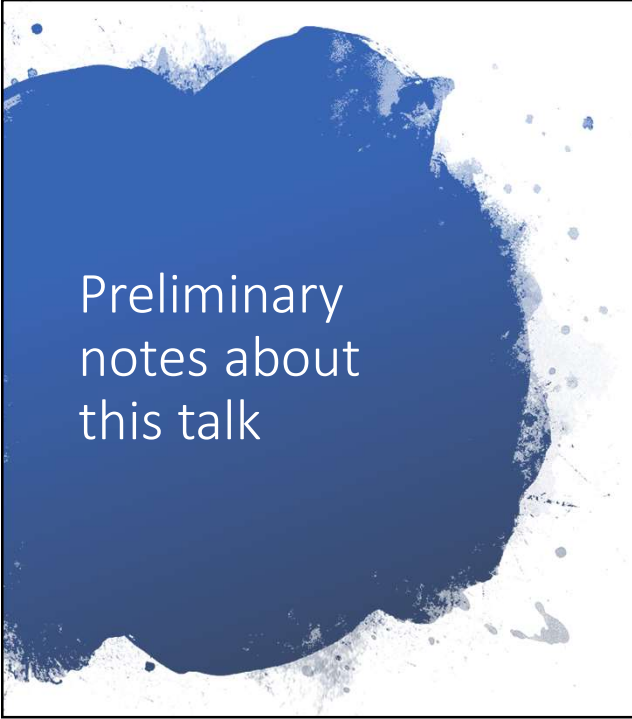
1



About me

- Studies: PhD in Computer Science, 1999
- Past positions: researcher at the University of Pisa 2000-2014
- Current position: associate professor at the University of Pisa
- Since November 2015: Vice-chair of the BSc and MSc curricula in "Computer Science" of the University of Pisa
- Member of the Council of the Doctorate in Computer Science since October 2013
- Supervisor of 7 PhD thesis (2 underway)
- Delegate for the assessment of the quality of research for my department (since 2012)

2



Preliminary notes about this talk

- You are a PhD student, you are learning how to conduct a research
- Your supervisor is your first reference:
 - He/she is experienced, and he/she knows the rules of the game
 - Learn from him as much as you can

This seminar is not intended to replace him/her!

3



Why this talk

In the last years “aggressive” use of bibliometrics to evaluate the research

... and consequent use of “aggressive” strategies by the researchers to improve their bibliometric indexes...

4

Number of papers...

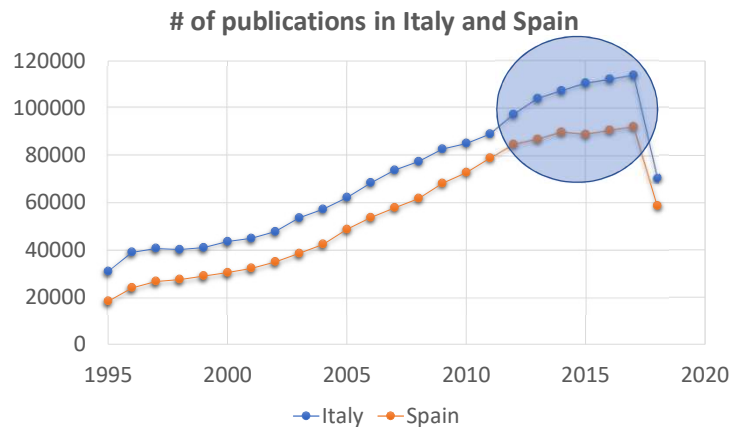


From Scopus, queries:

AFFILCOUNTRY(Italy)

AFFILCOUNTRY(Spain)

Moore's Law for papers:
the number of papers
that are "inexpensively"
produced doubles every
10 years...

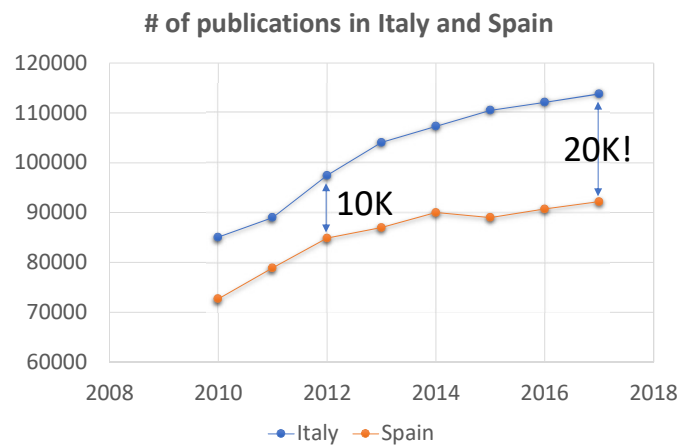


5

Happened in 2012...



- In Italy the rules for recruitment changed drastically
 - Pre-selection based on citations, h-index, #papers
- That's explain the growth in Italy after 2012
- A "**speculative bubble**" ...

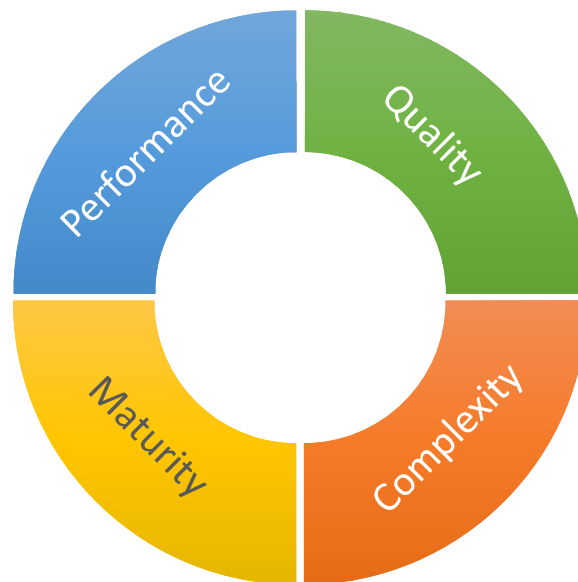


6

Part II: performance indicators & evaluation of research

7

Properties of a
paper



8

Performance			v. s.	Maturity		
venue	year	citations		venue	year	citations
ComCom	2007	893		J. of Algo.	2002	27
INFOCOM	2005	197		IEEE TIT	2012	9
ComCom	2001	121		SP&E	2010	22
SRDS	2001	118		IEEE TC	2001	13

9



Performance indicators

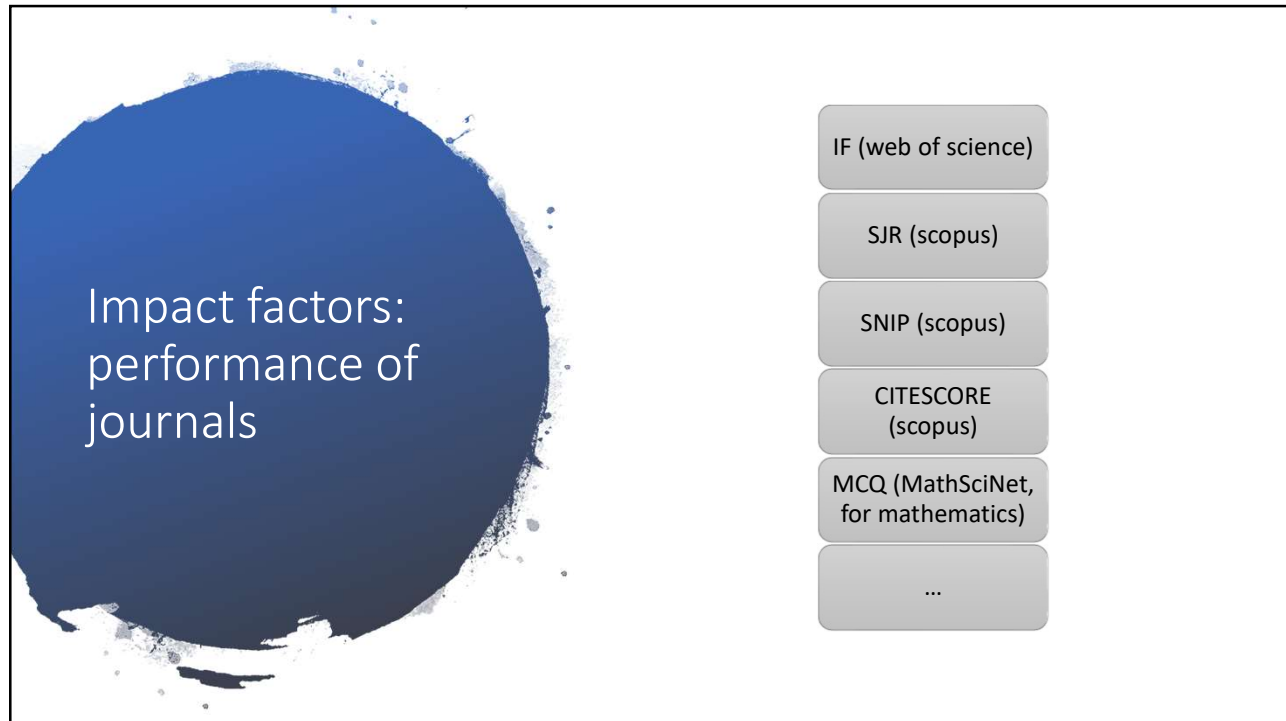
Sometimes they are called “quality” indicators (in Italy for example)

... but they are not. They measure the performance of a paper or of a journal in terms of “diffusion” in the research community

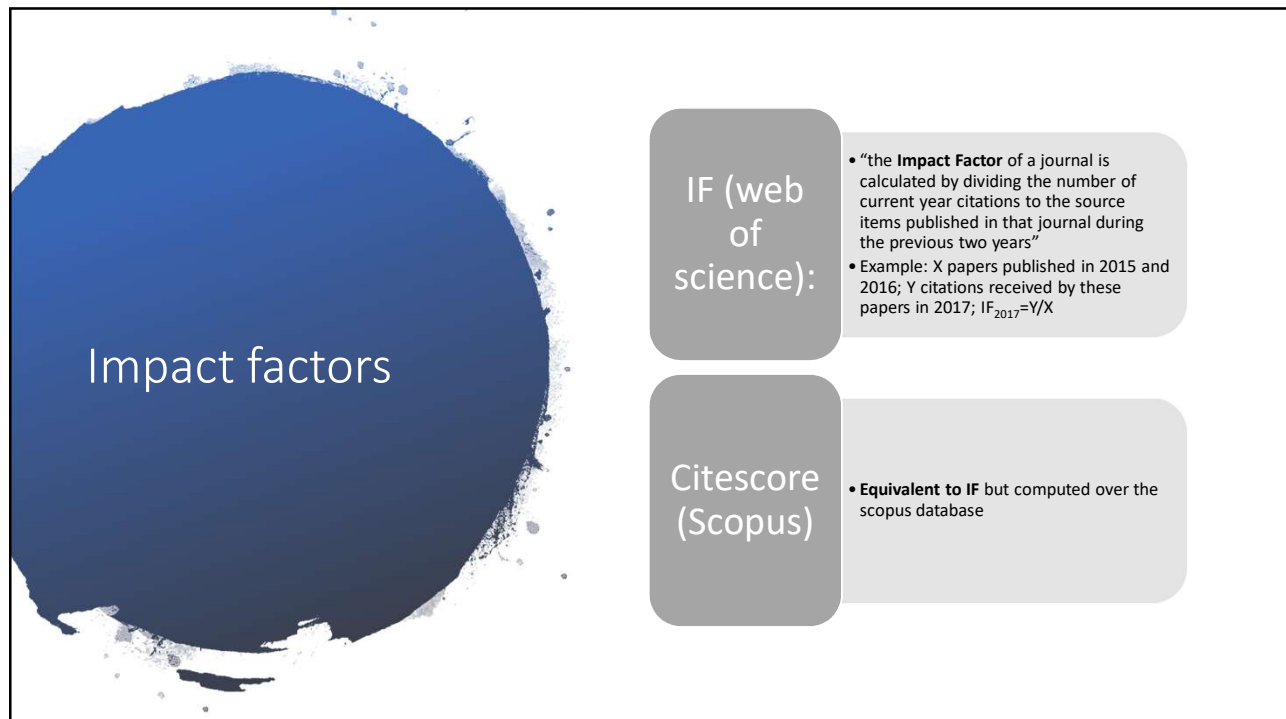
Many different indexes:

- Impact factors
- H-index
- Number of citations
- Number of papers

10



11



12



Impact factors

<p>SJR (scopus):</p>	<p>"SCImago Journal Rank measures weighted citations received by the serial. Citation weighting depends on subject field and prestige (SJR) of the citing serial." Ispirato al PageRank di google</p>
<p>SNIP (scopus):</p>	<p>"Source Normalized Impact per Paper measures actual citations received relative to citations expected for the serial's subject field."</p>

13

Publishing in high impact journals

High Impact →

- high diffusion, many readers
- high chance of being read & cited
- more selective, harder to publish

- In many areas the impact of the journals is taken rather seriously
- ... and recently also for computer science & engineering it is becoming important

14

Publishing in high impact journals (II)

- It's your preliminary choice
- ... but look first at the meaningfulness of the journal for your paper
- and review process may be engaging...

15

The impact
of my favorite
journals
is low!

Ranking of journals of area
«theoretical computer science»
based on CiteScore

CiteScore rank 2016 In category: Theoretical Computer Science

☆ #65	Theoretical Computer Science	0.97	42nd percentile
113			
Rank	Source title	CiteScore 2016	Percentile
#1	Foundations and Trends in Theoretical Computer Science	14.83	99th percentile
#2	IEEE Transactions on Evolutionary Computation	11.83	98th percentile
#3	ACM Computing Surveys	11.16	97th percentile
#4	Computer Science Review	8.20	96th percentile
#5	Information Sciences	5.37	96th percentile
#6	Integrated Computer-Aided Engineering	4.02	95th percentile
#7	ACM Transactions on Intelligent Systems and Technology	3.97	94th percentile
#8	Mathematical Programming Computation	3.86	93rd percentile
#9	IEEE Transactions on Computers	3.42	92nd percentile
#10	SIAM Review	3.26	91st percentile
#11	Parallel Architectures and Compilation Techniques - Conference Proceedings, PACT	3.24	90th percentile
#12	SIAM Journal on Optimization	3.22	89th percentile
#13	International Journal of Bio-Inspired Computation	3.14	88th percentile
#14	IEEE Computational Intelligence Magazine	2.96	88th percentile
#15	International Journal of Intelligent Systems	2.68	87th percentile
#16	Computer Speech and Language	2.67	86th percentile
#17	Journal of Parallel and Distributed Computing	2.60	85th percentile
#18	International Journal of General Systems	2.57	84th percentile
#19	International Journal of Approximate Reasoning	2.55	83rd percentile

16

The impact of my favorite journals is low!

Ranking of journals of area
«hardware & architecture»
based on CiteScore

CiteScore rank 2016

In category: Hardware and Architecture

★ #17	IEEE Transactions on Computers	3.42	88th percentile
141			
Rank	Source title	CiteScore 2016	Percentile
#1	IEEE Internet of Things Journal	9.90	99th percentile
#2	IEEE Network	7.81	98th percentile
#3	International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS	6.45	98th percentile
#4	IEEE Transactions on Cloud Computing	6.24	97th percentile
#5	Information Fusion	5.99	96th percentile
#6	Future Generation Computer Systems	5.60	96th percentile
#7	IEEE Transactions on Parallel and Distributed Systems	5.19	95th percentile
#8	Journal of Network and Computer Applications	4.42	94th percentile
#9	IEEE Transactions on Services Computing	4.15	93rd percentile
#10	Computer Physics Communications	4.04	93rd percentile
#11	Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM	3.80	92nd percentile
#12	Synthesis Lectures on Computer Architecture	3.76	91st percentile
#13	Ad Hoc Networks	3.67	91st percentile
#14	Journal of the ACM	3.56	90th percentile
#15	Mobile Networks and Applications	3.48	89th percentile
#16	Journal of Manufacturing Systems	3.46	89th percentile
★ #17	IEEE Transactions on Computers	3.42	88th percentile
#18	Information Systems	3.33	87th percentile
#19	VLDB Journal	3.27	86th percentile

17

The impact of my favorite journals is low!

Ranking of journals of area
«software»
based on CiteScore


CiteScore rank 2016

In category: Software

★ #52	IEEE Transactions on Computers	3.42	85th percentile
367			
Rank	Source title	CiteScore 2016	Percentile
#1	IEEE Transactions on Pattern Analysis and Machine Intelligence	13.29	99th percentile
#2	Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	12.72	99th percentile
#3	IEEE Transactions on Evolutionary Computation	11.83	99th percentile
#4	International Journal of Computer Vision	11.06	99th percentile
#5	Foundations and Trends in Machine Learning	10.00	98th percentile
#6	Proceedings of the IEEE International Conference on Computer Vision	9.92	98th percentile
#7	IEEE Network	7.81	98th percentile
#8	Journal of Statistical Software	7.71	97th percentile
#9	IEEE Transactions on Cybernetics	7.14	97th percentile
#10	IEEE Transactions on Image Processing	6.73	97th percentile
#11	IEEE Transactions on Neural Networks and Learning Systems	6.48	97th percentile
#12	International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS	6.45	96th percentile
#13	International Journal of Robotics Research	6.30	96th percentile
#14	IEEE Transactions on Cloud Computing	6.24	96th percentile
#15	Information Fusion	5.99	96th percentile
#16	Future Generation Computer Systems	5.60	95th percentile
#17	IEEE Transactions on Software Engineering	5.51	95th percentile
#18	Information Sciences	5.37	95th percentile
#19	Pattern Recognition	5.36	94th percentile

18

Publishing in high impact journals (III)

- However,
high impact  large number of citations
- ... why so?
- The citations received by a paper are an individual value
- The impact of a journal is a collective value
- All high-impact journals have highly-cited and normally/lowly-cited papers

19

Citations and H-Index

- Usually, the number of citations received and the H-index are considered in combination with the journal's impact
- They indicate the “individual” performance of a researcher or of a paper

20

Citations and H-Index

- H-index of a researcher is X if he has exactly X papers each of which received at least X citations
- H-index grows slowly and it is not linear!
 - $1 < 5$ but $11 \ll 15 \lll 19 \dots$
- There are criticisms to H-Index, but it is still widely used

21

Citations and H-Index

are usually a factor of **stress** and **depression**:

- They do not (necessarily) depend on the quality of your work
- They do not (necessarily) depend on your preliminary choice (as impact factors)
- They depend on the future behavior of other researchers, out of your control

22

How to get cited?

there's no guarantee,

depends on many factors

... and may take time...

23

Why do you cite a paper?

- To refer a work strongly related to yours
- To motivate the importance of a research field
- To explain the impact of your work on the society
- To avoid plagiarism
- To defend your approach/methodology
- To defend a statement in your paper
- ...

You cite a paper because it is useful!

24

About usefulness

- Writing papers useful for a research community is not easy
 - Many time you know later whether they are really useful
 - I don't know of anybody who wrote only useful papers
 - In fact, most papers have a limited "usefulness" ...
- Sometimes we write papers just to:
 - to test our ideas,
 - receive opinions from reviewers,
 - document our work
 - ... and sometimes even to witness or to strengthen a cooperation

25

Main factors for citations

1. usefulness
2. venue
3. reputation of the authors
4. size of research community
5. timeliness of the work

26

2. Venue of the publication

- Not only a matter of impact
- The content of the paper should match well the audience of the journal/conference
 - Write the paper for that journal
 - Use terminology, methodology, approach typical of that community
 - i.e. if they expect formal proofs give them formal proofs
 - If they expect simulations give them simulations
 - ... etc...

27

2. Venue of the publication : example

Two papers with a similar idea about routing protocols in ad hoc networks, (almost) same year

- GPSR: Greedy Perimeter Stateless Routing for wireless networks
MOBICOM 2000 – 4940 citations
- Routing with guaranteed delivery in ad hoc wireless networks
Dial-M '99 – 559 citations
 - Later appeared also in Wireless Networks '01 – 781 citations

28

About venue

- You are in the best position to assess your work:
 - if you feel it is very good makes sense to write it for a top journal/conference
 - ... otherwise it may be a good idea to write it anyway and address a minor venue
 - ... but write it for the venue you chose
- A good venue always help good papers...
- ... but it doesn't help poor performing papers

29

2. Venue & 3. Reputation: example

Two papers with a very similar idea about routing protocols in ad hoc networks, same year

- Virtual ring routing: Network routing inspired by DHTs
ACM SIGCOMM '06 – **150 citations**
- Reliable routing in wireless ad hoc networks: The virtual routing protocol
J. of Network and Systems Management '06 – **12 citations**

30

How do you gain reputation?

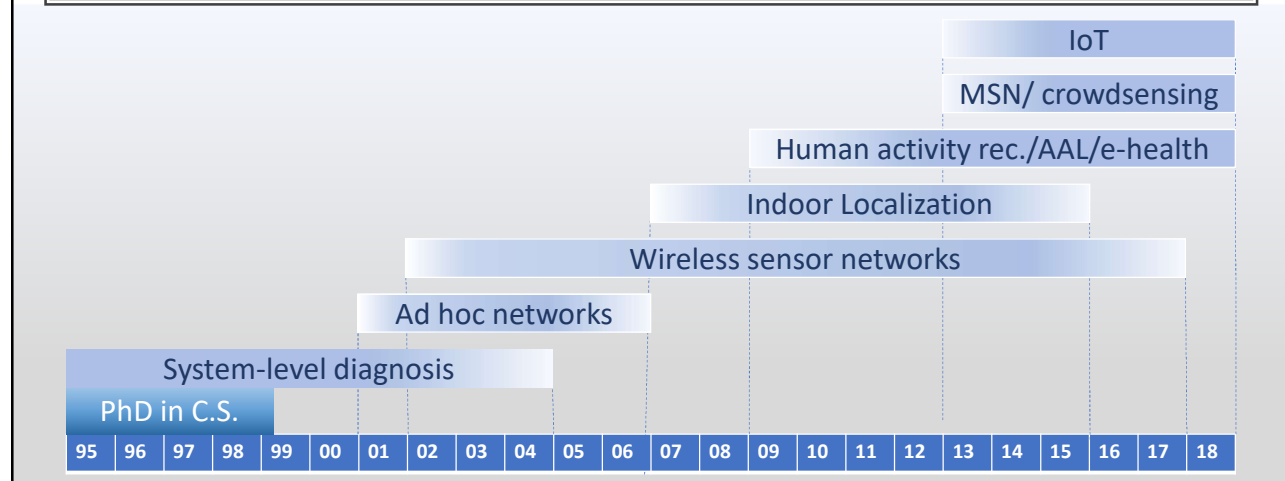
1. Writing high-quality papers
2. Being involved in a research community
 - serve the community
 - take part to the public events
 - ...
3. Being proactive in innovation:
 - proposing new themes of research
 - proposing new workshops/special issues
 - ...
4. Establishing a network of connections

3. Reputation of the authors

31

4. Size of a research community

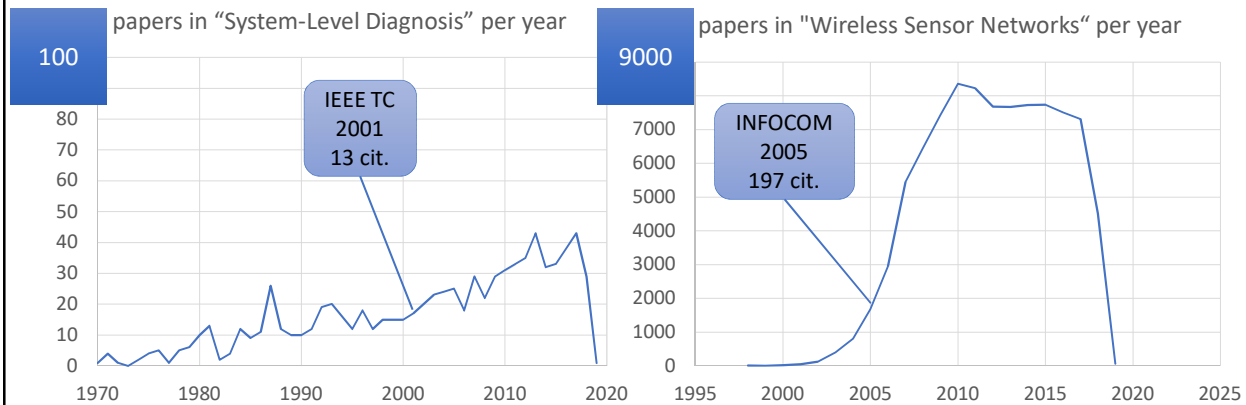
First, my main research areas...



32

4. Size of a research community – II

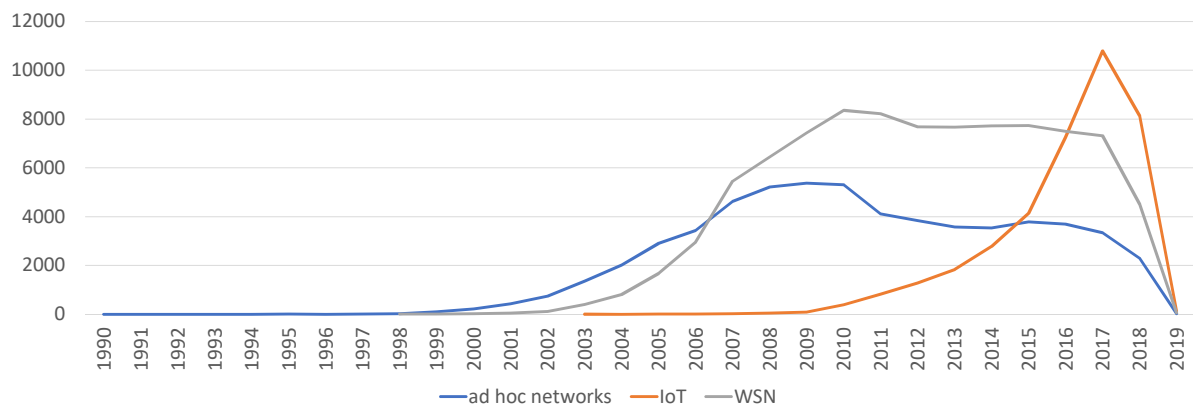
“system-level diagnosis” vs “Wireless sensor networks”



33

4. Size of a research community – III

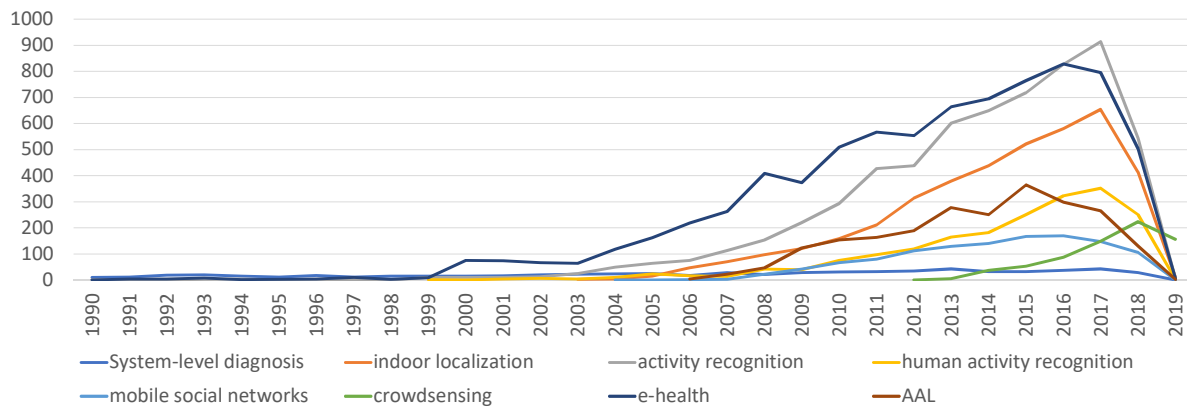
“crowded areas”, number of papers per year



34

4. Size of a research community – IV

“Less-crowded” areas, number of papers per year



35

About size of community

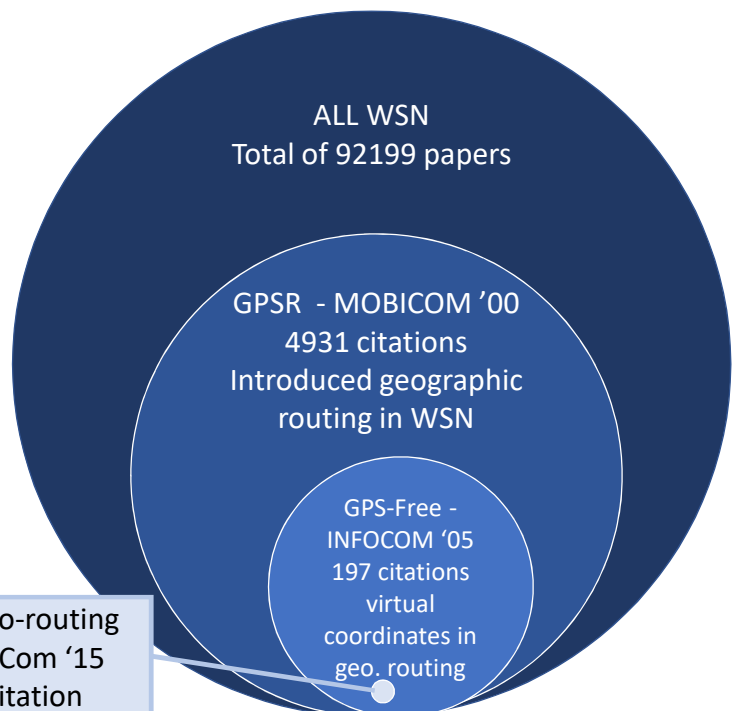
really top papers had been written for communities that did not exist yet...

- don't be obsessed by the size
- ... but don't remain entrapped in a “black hole”
 - If a research field is becoming a desert consider moving ahead

36

4. Size (sub-areas) & 5. Timeliness

3D geo-routing
ComCom '15
1 citation



37

5. Timeliness in geographic routing

Greedy perimeter stateless routing
(GPSR), *MOBICOM 2000*

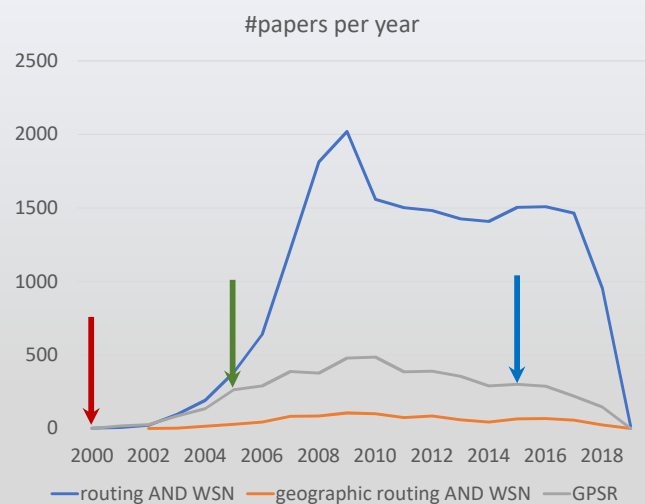
- 4931 citations, a top conference

GPS free coordinate assignment and
routing in wireless sensor networks
(VCAP), *INFOCOM 2005*

- 197 citations, a top conference

Multi-Dimensional Recursive Routing
with Guaranteed Delivery in Wireless
Sensor Networks, *ComCom 2015*

- 1 citation, a good impact journal



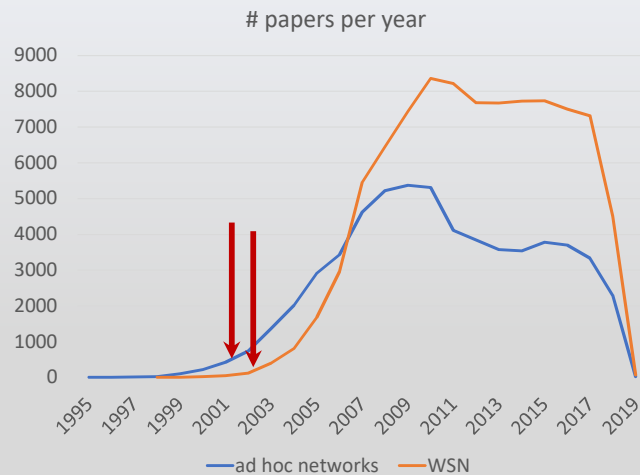
38

5. Timeliness: dependability in WSN & Ad Hoc

Comparison-Based System-Level Fault
Diagnosis in Ad-Hoc Networks, *SRDS*
2001

- 118 citations, conference
- Crash Faults Identification in Wireless
Sensor Networks, *Comp. Comm.* *2002*
- 121 citations, at time a class B journal

Usefulness ↑↑
Venue ↑
Size ↑
Timeliness ↑↑



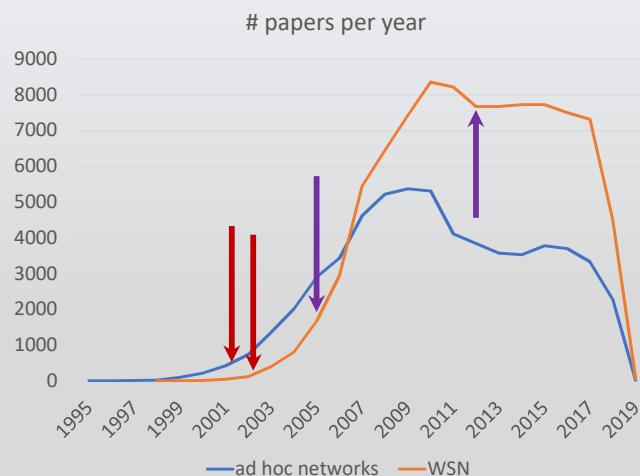
39

5. Timeliness: dependability in WSN & Ad Hoc

Fault Recovery Mechanism in Single-
Hop Sensor Networks, *Computer*
Comm. *2005*

- 13 citations – class B journal
- Energy-Aware Test Connection
Assignment for the Self-Diagnosis of a
WSN, *Journal of the Brazilian Computer*
Society *2012*
- 6 citations, class D journal

Usefulness ↓↓ ↓↓
Venue ↑ ↓
Size ↑ ↑
Timeliness ↑ ↓↓



40

5. Timeliness

Note:

timeliness == right on time

too early may be as bad as too late!

41

About timeliness

Some works deserve to be written anyway:

- If they close definitively a research field (they will probably don't get many citations...)
- If they have other values

Again, don't be obsessed by timeliness, but keep an eye to it

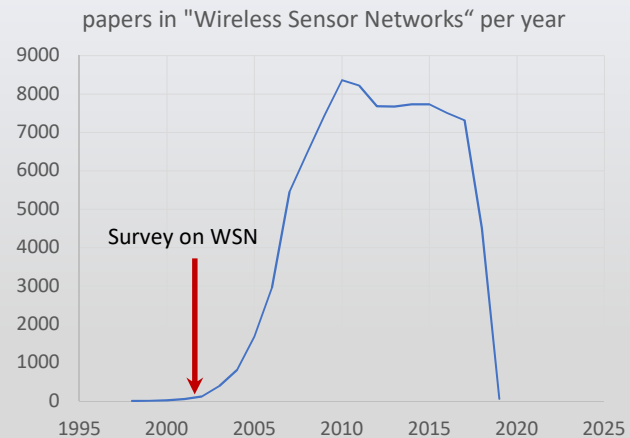
42

A case study... a survey on WSN of 2002!

“Wireless sensor networks: a survey”, Connect 2002

Observe well the dates...

- One expects a survey on a field when it becomes mature enough
- Instead most of these surveys are right at the beginning...
- How is this possible?



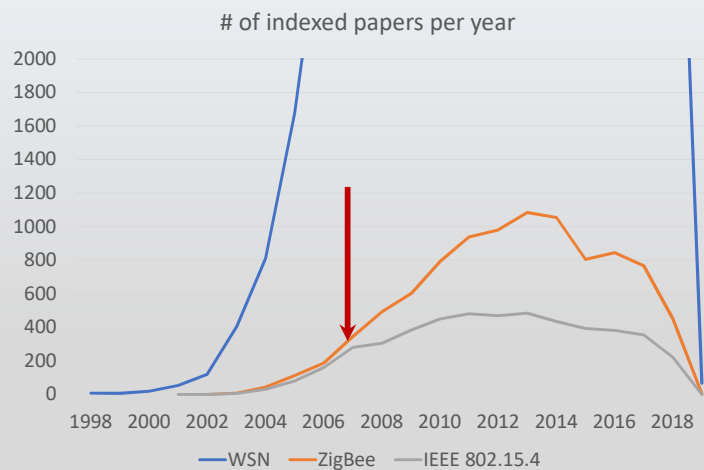
43

What did I do to emulate this survey...

Wireless Sensor Networks: a Survey on the State of the Art and the 802.15.4 and ZigBee Standards

Computer Communications 2007

- 880 citations – at that time a class B journal



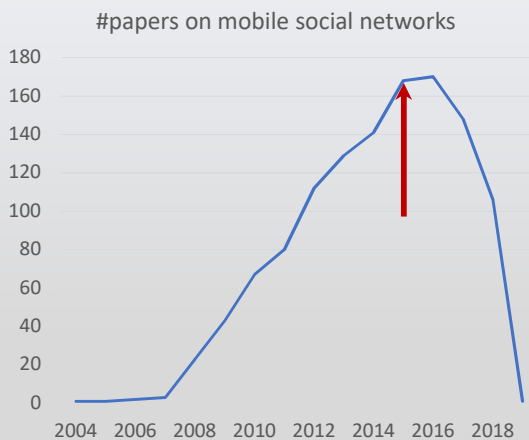
44

What did I do to emulate this survey...

On Service Discovery in Mobile Social Networks: Survey and Perspectives
Computer Networks 2015 – 21 citations

Looks late, but:

- nobody was working on service discovery in MSN
 - most works on routing
- it was a bet
- ... and maybe MSN will keep growing...



45

Part III shortcuts & cheating

Why they are not a good idea

46

Weaknesses of the performance indicators

- The systematic use of performance indicators to assess researchers is producing a “speculative” bubble
 - Number of papers and citations are growing and growing
- Researchers may use strategies to increase their performance surreptitiously:
 - exchange citations
 - request citations of their papers in their reviews
 - unmoderated use of self-citations
- Bad practices of journals to increase their Impact Factors produced new and more complex indexes
 - we already seen a number of impact factors

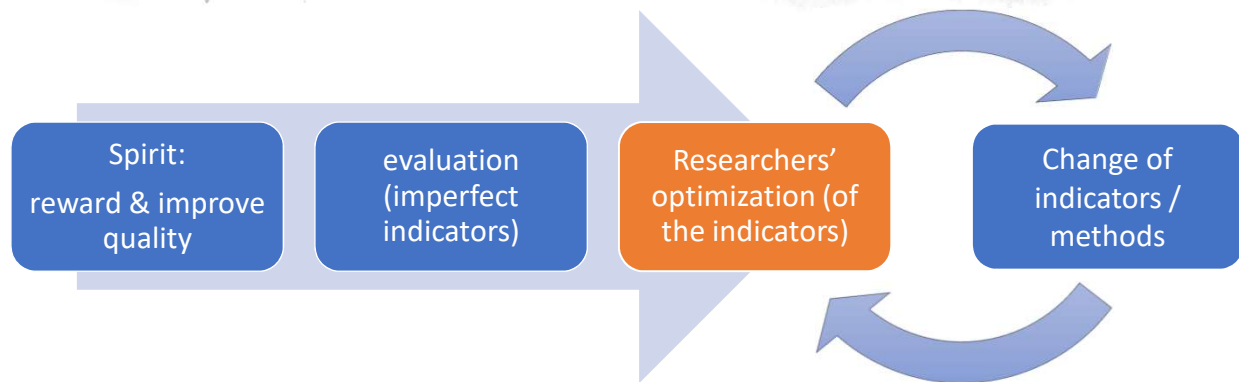
47

Self-citations...

- self-citations are physiological:
 - Your work is related to other previous works of yours
 - You make a bit of advertisement to your past works
- their unreasonable use may become a problem for yourself
 - Easy to locate and filter out
 - They are written on the stone... are visible forever

48

Hyper-specialization: the evaluation loop



49

The risks of bad practices

- Bad practices and cheating may seriously affect your reputation
- Bad practices, cheating and iper-specialization are likely to produce immediate changes in the assessment of research
- The great risk is to follow these changes rather than to be always a step ahead

... but how to be a step ahead?

50

Focus on the quality of your work!

... and, of course, keep an eye to:

1. usefulness
2. venue
3. reputation
4. size of research community
5. timeliness

... and to other factors that may become important in the future:

1. impact on society
2. interdisciplinarity
3. divulgation/teaching
4. ...

51

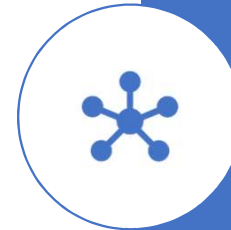
Conclusions

52

Some considerations on really top cited works

In the field of WSN there are some very important works that gave rise to the area:

- Directed diffusion: A scalable and robust communication paradigm for sensor networks, MOBICOM 2000 - 3712 citations
- Greedy perimeter stateless routing, MOBICOM 2000 – 4931 citations
- Maturity and complexity are not really their strengths, so to say...
- ... but from the point of view of reputation, timeliness, venue they are really strong
- they also proved very useful...
they contributed to build a very large research community!



53

About performance

I don't know of anybody that wrote **only** useful, timely papers on top journals for growing communities

I would not give myself such a mission. Consider also writing papers:

- for small communities
- for communities that still do not exist
- useful for you (but write for the other people anyway)
 - to test ideas, receive reviews, document your work etc.
- to witness or to strengthen a cooperation
- even for minor venues, if the idea/work is not so good
- late (even a work that closes a research area is worth of being written)
- ...



54

My best five recommendations

- 1
do a quality job
- 2
write papers
for the others,
not for yourself
- 3
do not be
obsessed by
performance
indicators
- 4
keep an eye on
trends
- 5
understand the
evaluation of
research and its
evolution

55



Thank you!

56