

Reflections on 40+ years of software engineering research and beyond *an insider's view*



Carlo Ghezzi

DeepSE Group

Dipartimento di Elettronica e Informazione

Politecnico di Milano, Italy

carlo@polimi.it

Outline of the talk

- A journey through 40+ years of software engineering research, mostly observed through ICSE
- *What did we produce?*
- *How can their success (impact) be "defined"?*
- *How can it be measured?*
- *What can we learn from past?*
- *Can we do better?*
- *Will the paradigms followed in the past survive in the future?*

Part 1: Looking backwards



- Where does ICSE come from?
- What can we learn by mining ICSE data?

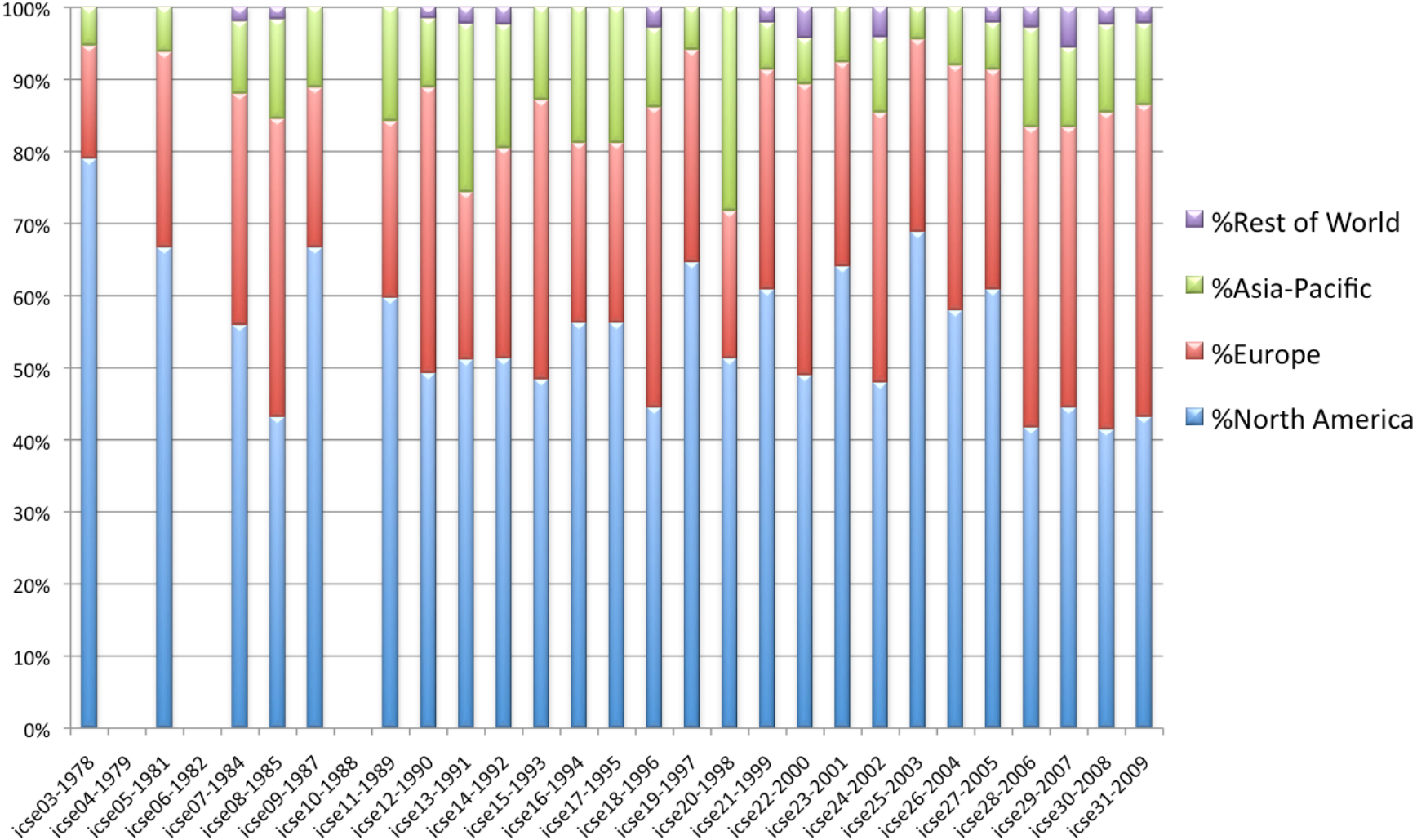
A bit of history

- The roots are in the NATO meetings in Garmisch-Partenkirchen (1968) and Roma (1969)
- **ICSE** started in 1975, this is the 31-st
 - first conference called ICSE is actually ICSE2
 - ICSE1 was NCSE---1st National Conf on Software Engineering, Sept 1975
 - became annual in 1987
- **TSE** started in March 1975
- **ESEC** started in 1987, **FSE** started in 1993, they joined in 1997
- **TOSEM** started in January 1992

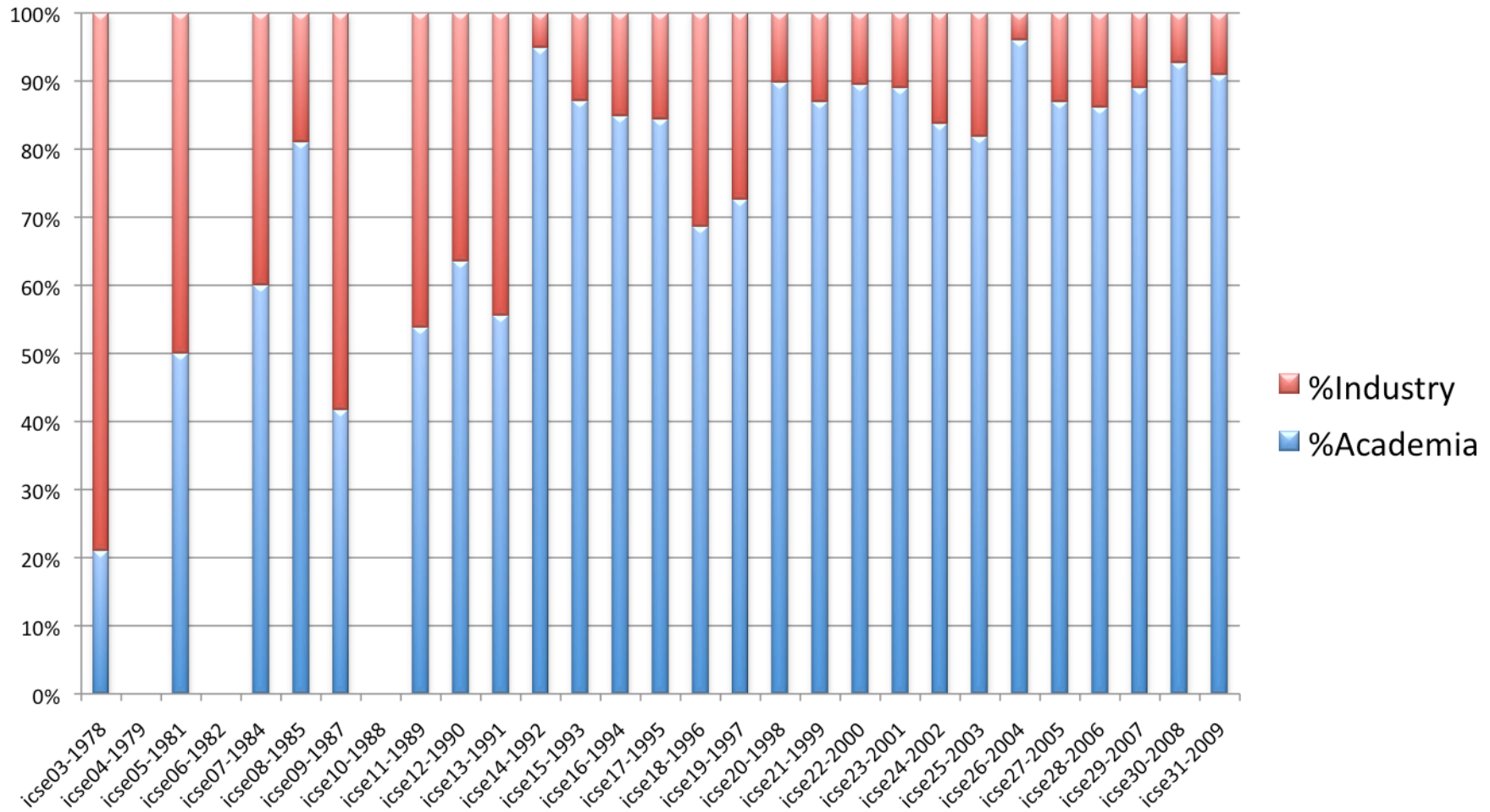
Mining ICSE data

- International Conference on Software Engineering
 - How much **international**?
 - How are the different **engineering** research stakeholders (academia, industry) represented?
 - scientific leadership, contributors
- Paper demographics
 - Is ICSE an attractive venue?
- Research areas
 - How did they evolve?

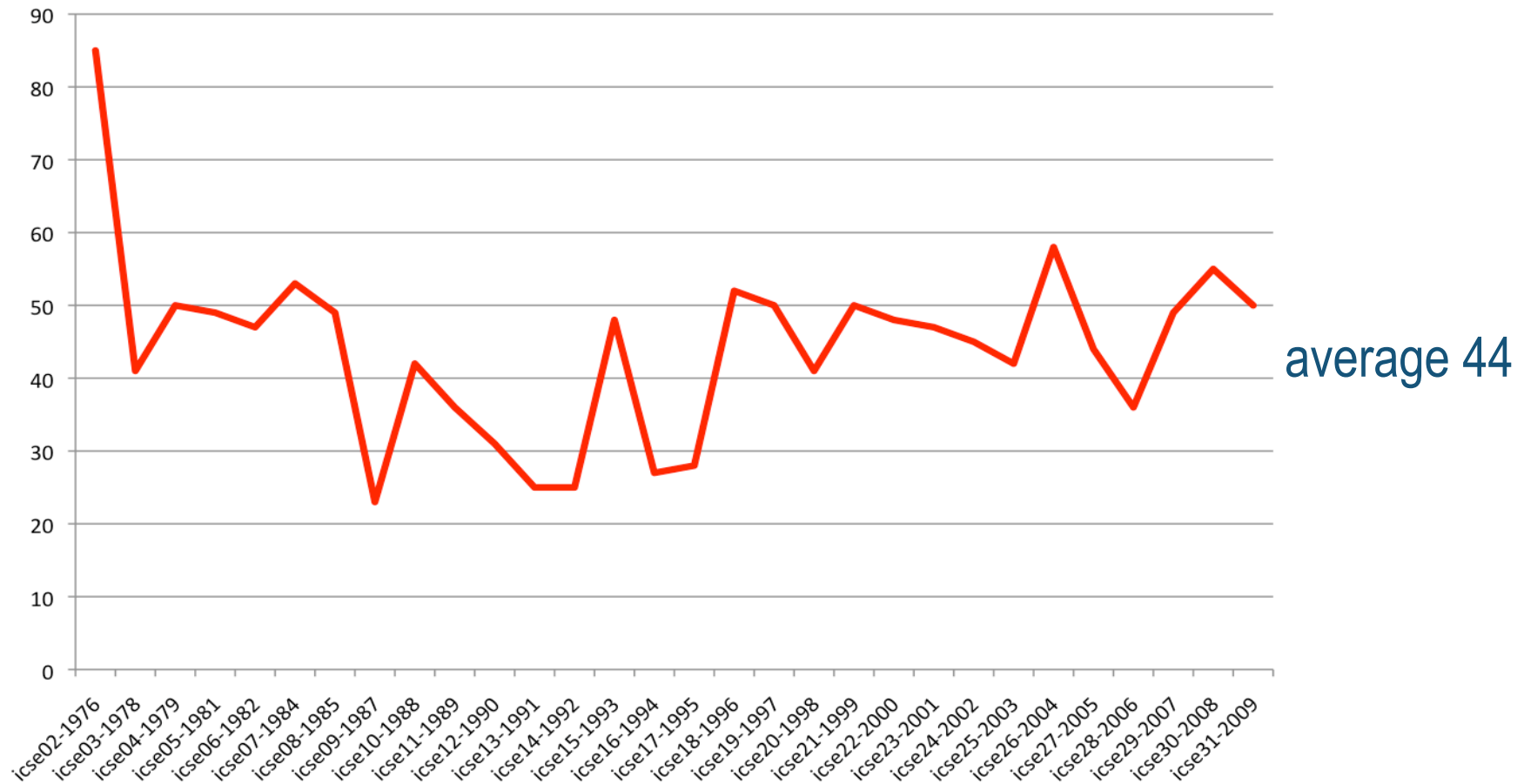
ICSE PCs



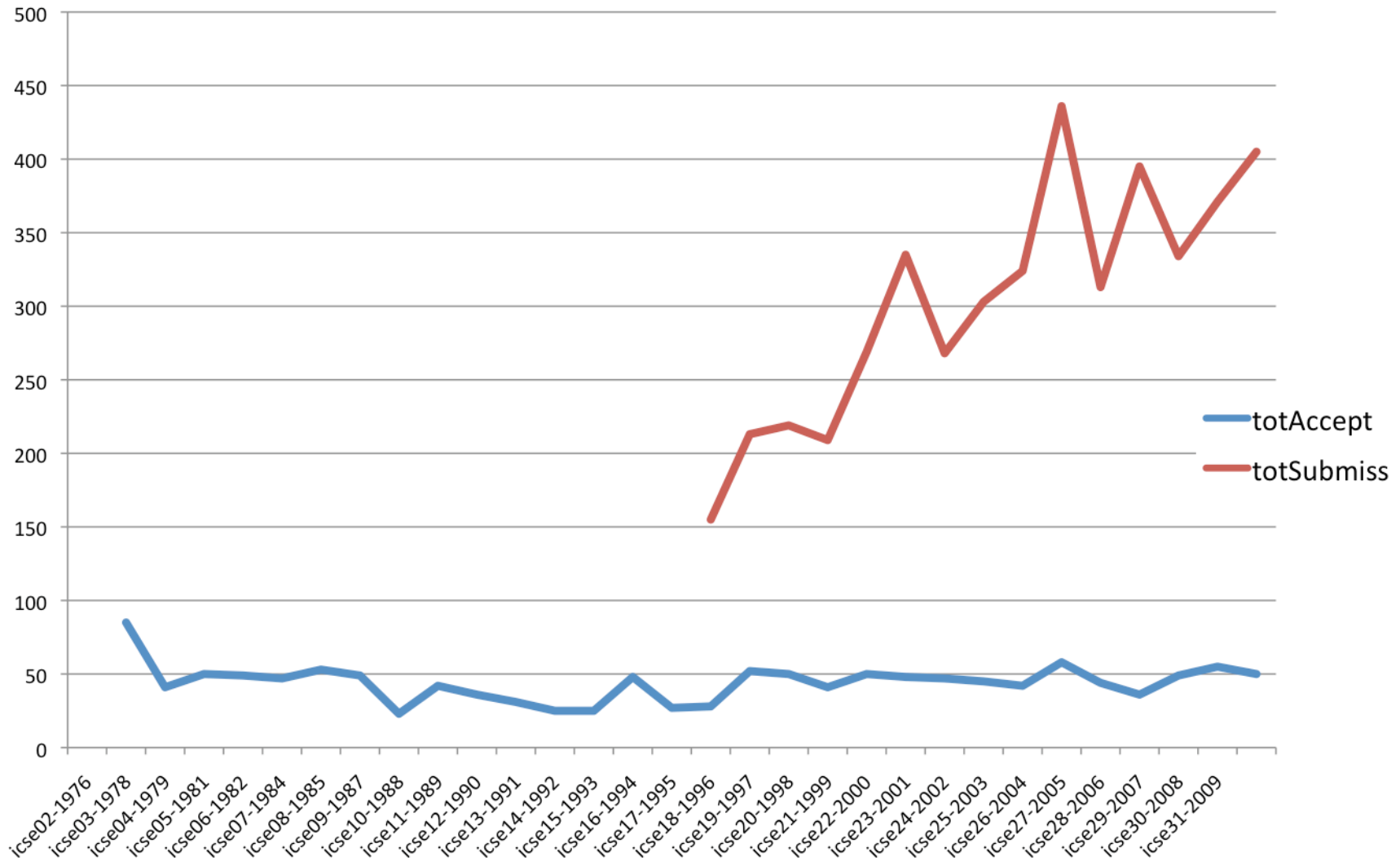
ICSE PCs: industry/academia



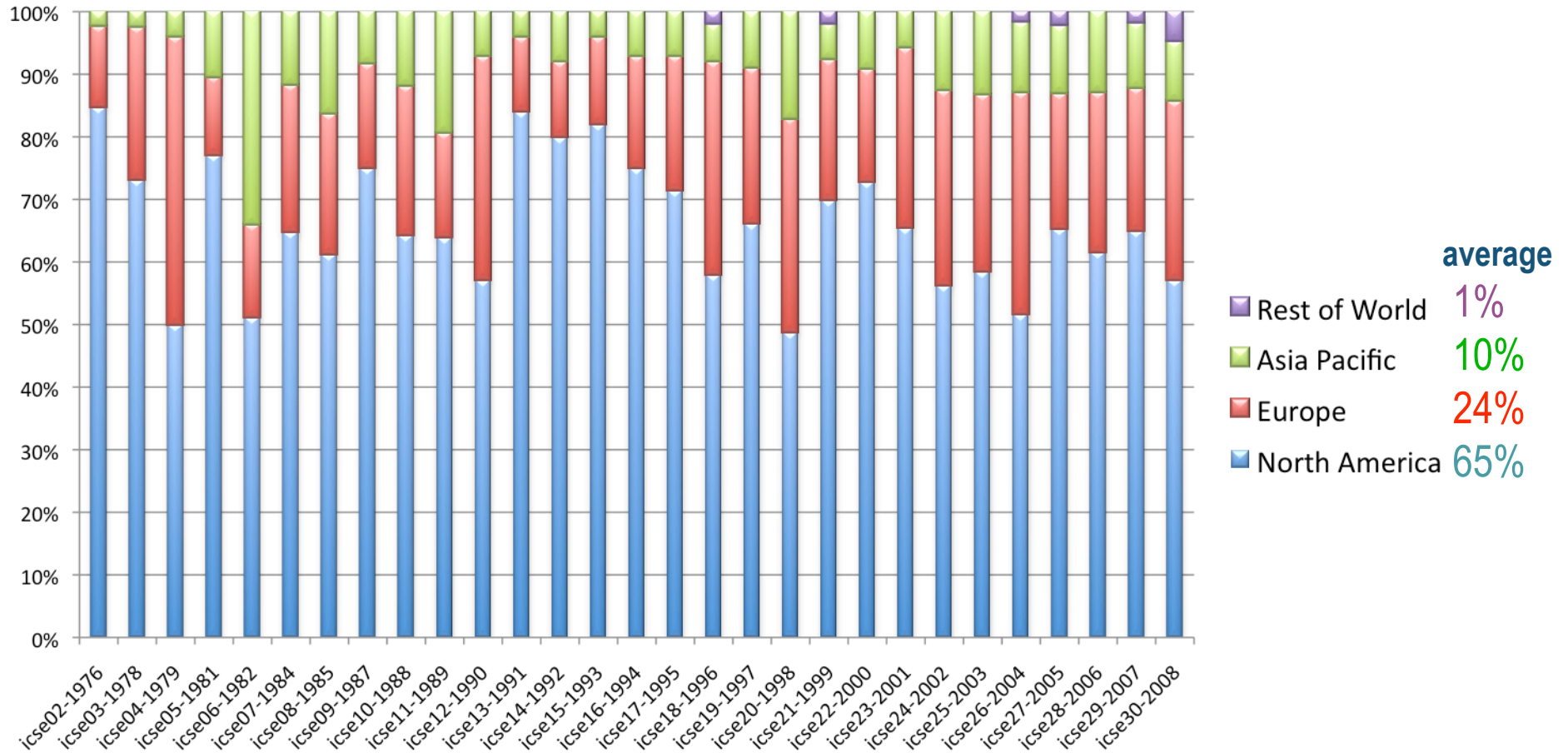
Total number of accepted papers



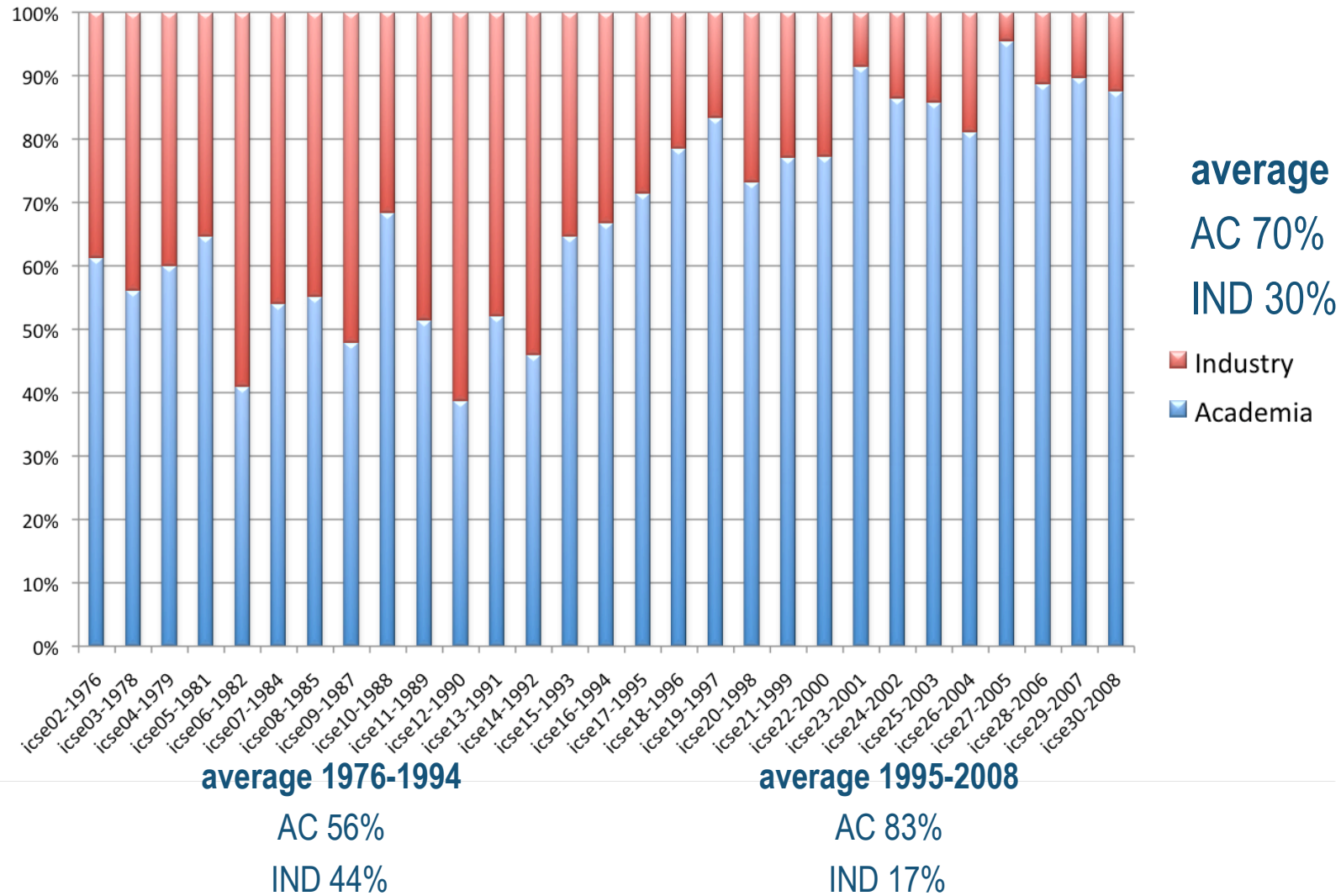
Submitted/accepted



ICSE papers: geographical distribution



Papers: industry vs. academia



Preliminary findings

ICSE is nurtured by an international community

PC		NorthAmerica	Europe	AsiaPacific	RestOfWorld
all		55%	32%	12%	1%
-5 yrs		46%	40%	11%	3%
papers		NorthAmerica	Europe	AsiaPacific	RestOfWorld
all		65%	24%	10%	1%
-5 yrs		60%	27%	11%	2%


High number of submissions, high selection rate

There is a consensus that ICSE is a prestigious venue

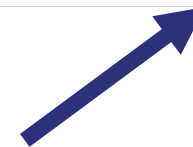
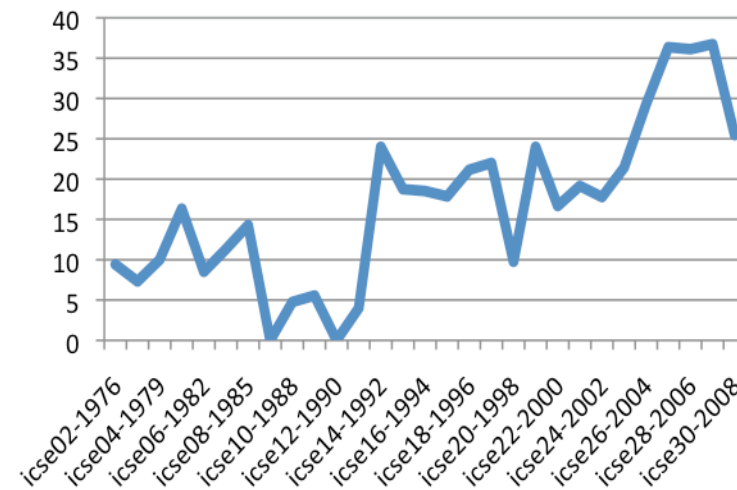
Preliminary findings

- Industry participation decreased over time
 - research labs of large corporations disappeared
 - ICSE became more research oriented
 - initially: tool fair
 - then: tutorials
 - now: workshops
 - now mostly co-authored papers (academia/industry)

Topics

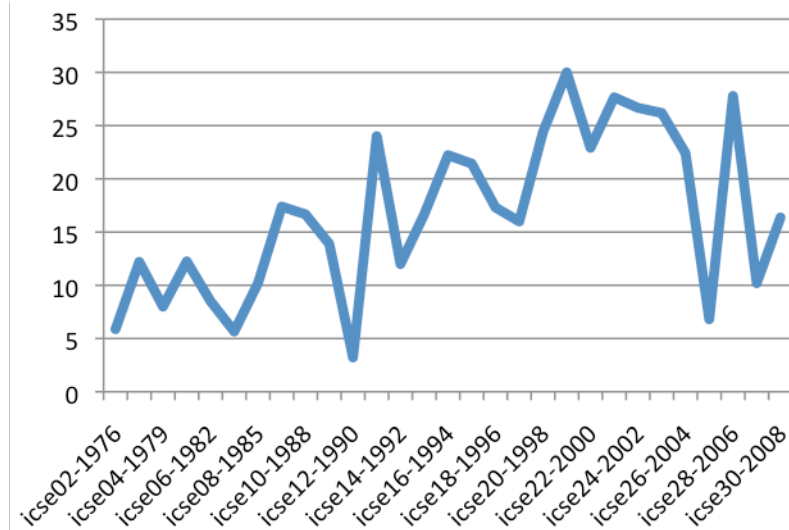


Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



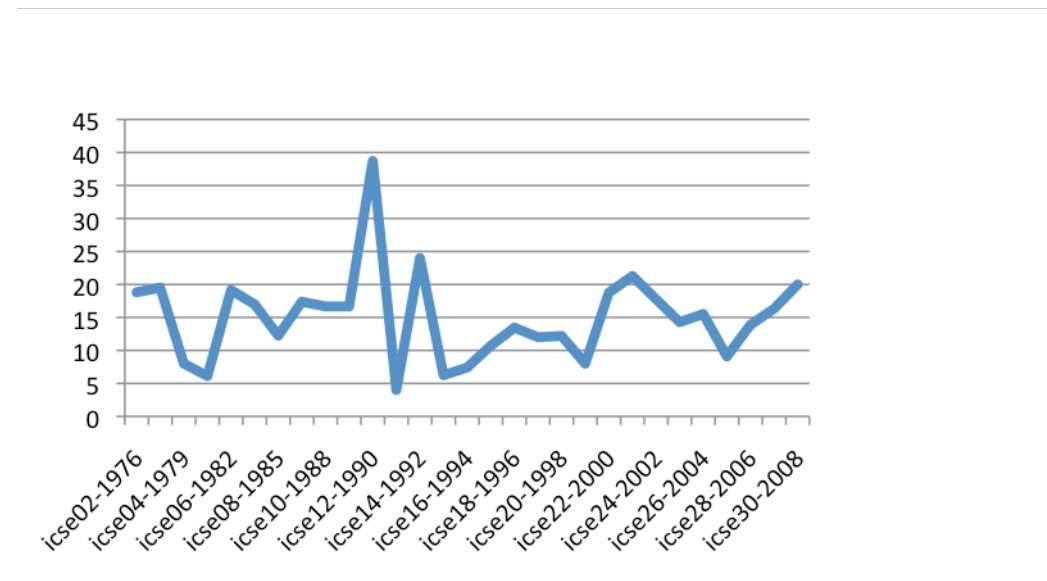
Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



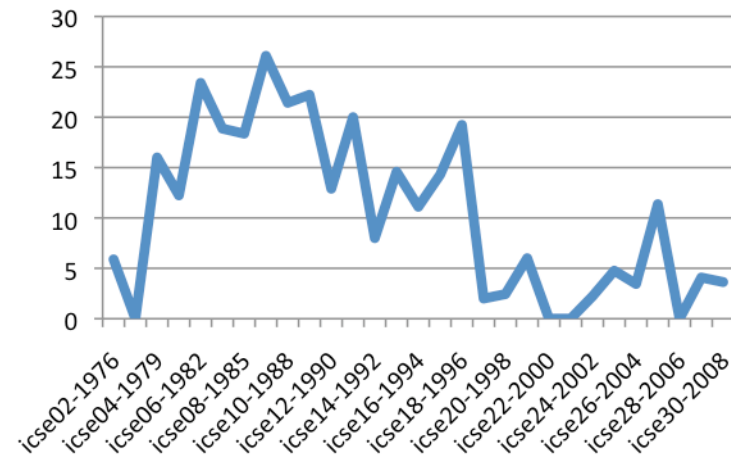
Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



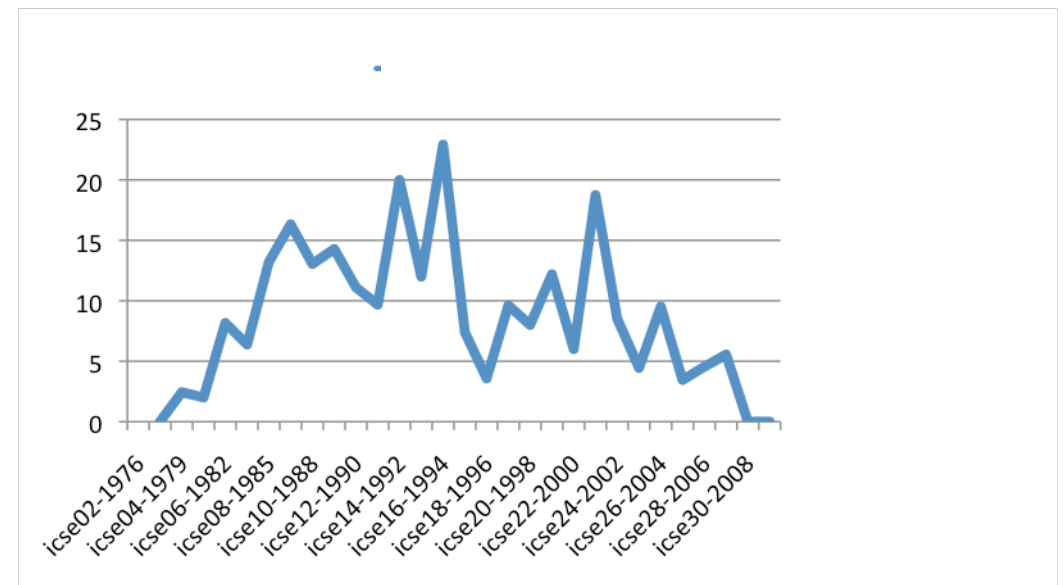
Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



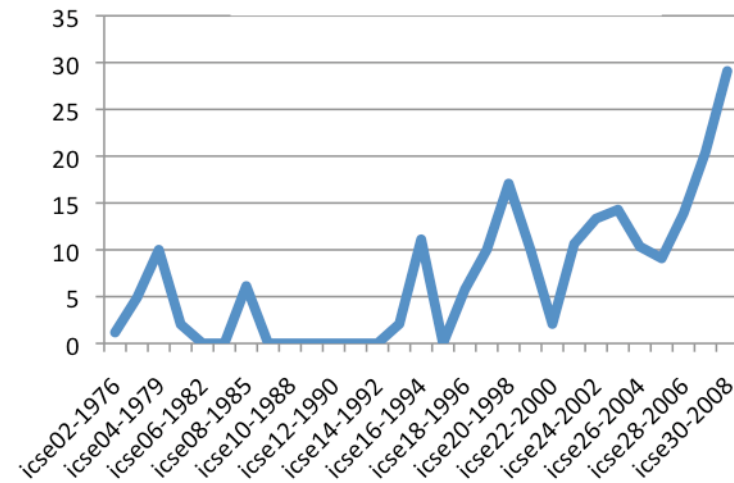
Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



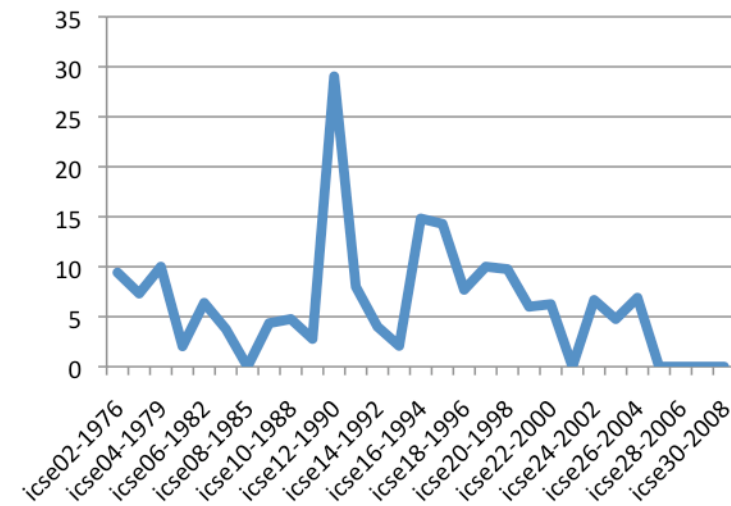
Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18

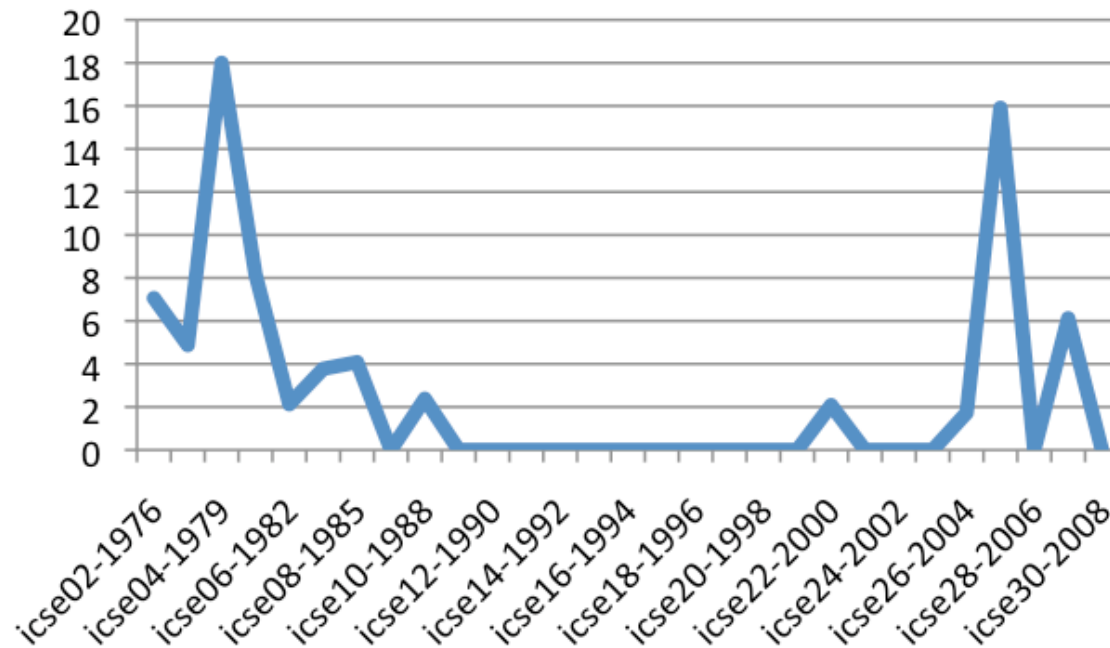


Topics

Topic	%
Testing&Analysis	17
Design&Software Architecture	17
Specification&Verification	15
Tools&SDEs	11
Management&Costs	9
Empirical Studies	7
Existing Systems	6
Others (13 topics)	18



Programming languages



Some findings

- Large variety of topics
- Trend from "methodology oriented", informal approaches to systematic and formally-based approaches
- Increased attention on empirical evaluations
- Still little on paradigmatic examples or domain-specific approaches

Can we identify our best products?

- Which criteria to follow?
 - Quantitative
 - Citation count
 - Downloads count
 - Others?
 - Qualitative
 - Most influential -10 years?
 - Others?

Quantitative criteria

- Citation count
 - date of count Jan-March 2009
 - source Google Scholar
 - version (conference/journal) both (if j expands c)
 - precision ???
- Number of downloads
 - date, period Jan-March 2009, 12 months
 - source ACM digital library
 - version conference
 - precision ???

Outstanding products based on citations

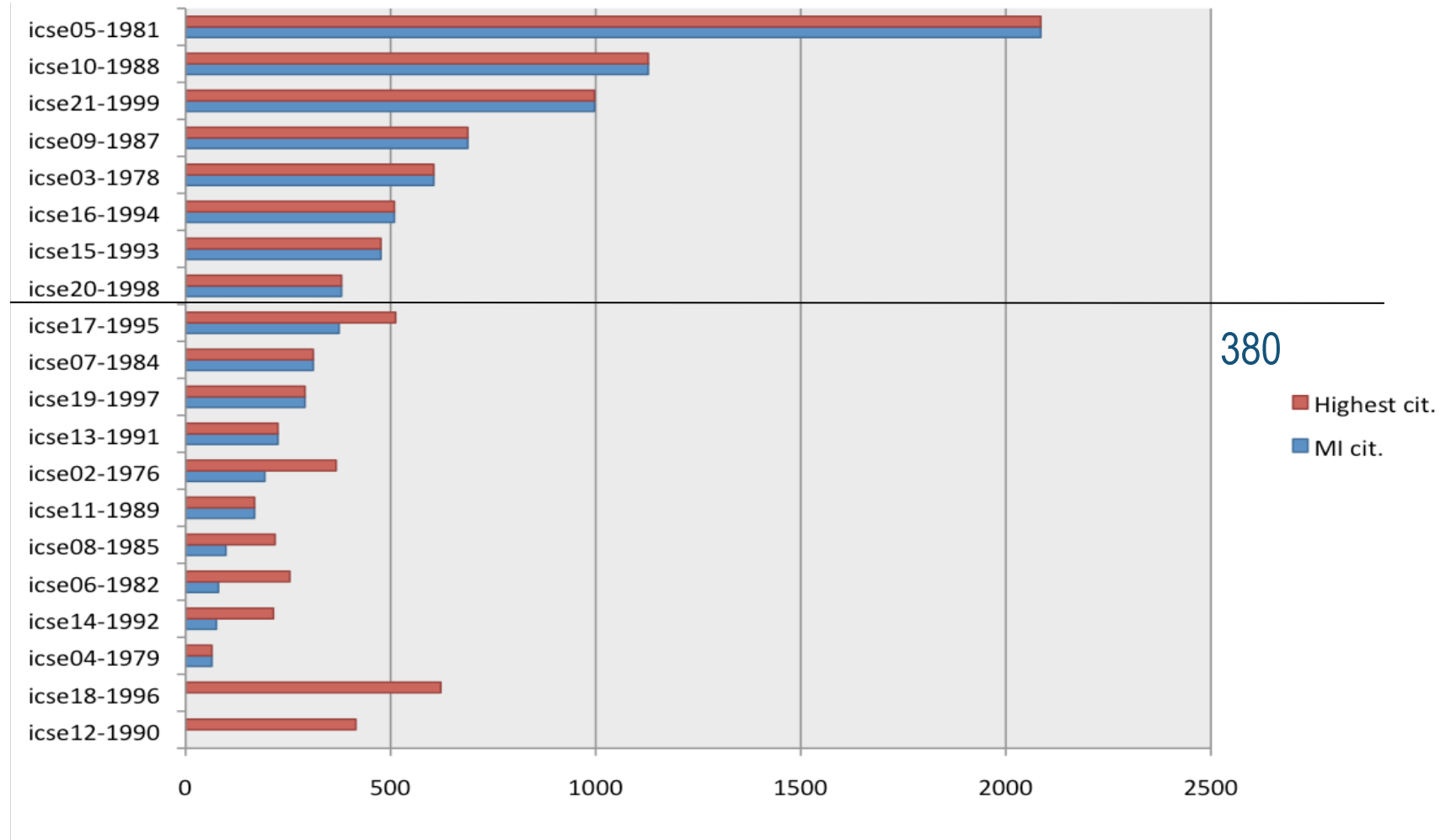
1. Program slicing (ICSE'81) **2120**
2. Statemate: a working environment for the development of complex reactive systems (ICSE'88) **1170**
3. N degrees of separation: multi-dimensional separation of concerns (ICSE'99) **997**
4. Bandera : extracting finite-state models from Java source code (ICSE'00) **791**
5. Software processes are software too (ICSE'87) **688**
6. Managing the development of large software systems: concepts and techniques (ICSE'87) **667**
7. Executable object modeling with statecharts (ICSE'96) **622**
8. Designing software for ease of extension and contraction (ICSE'78) **605**

Citations vs expert judgment

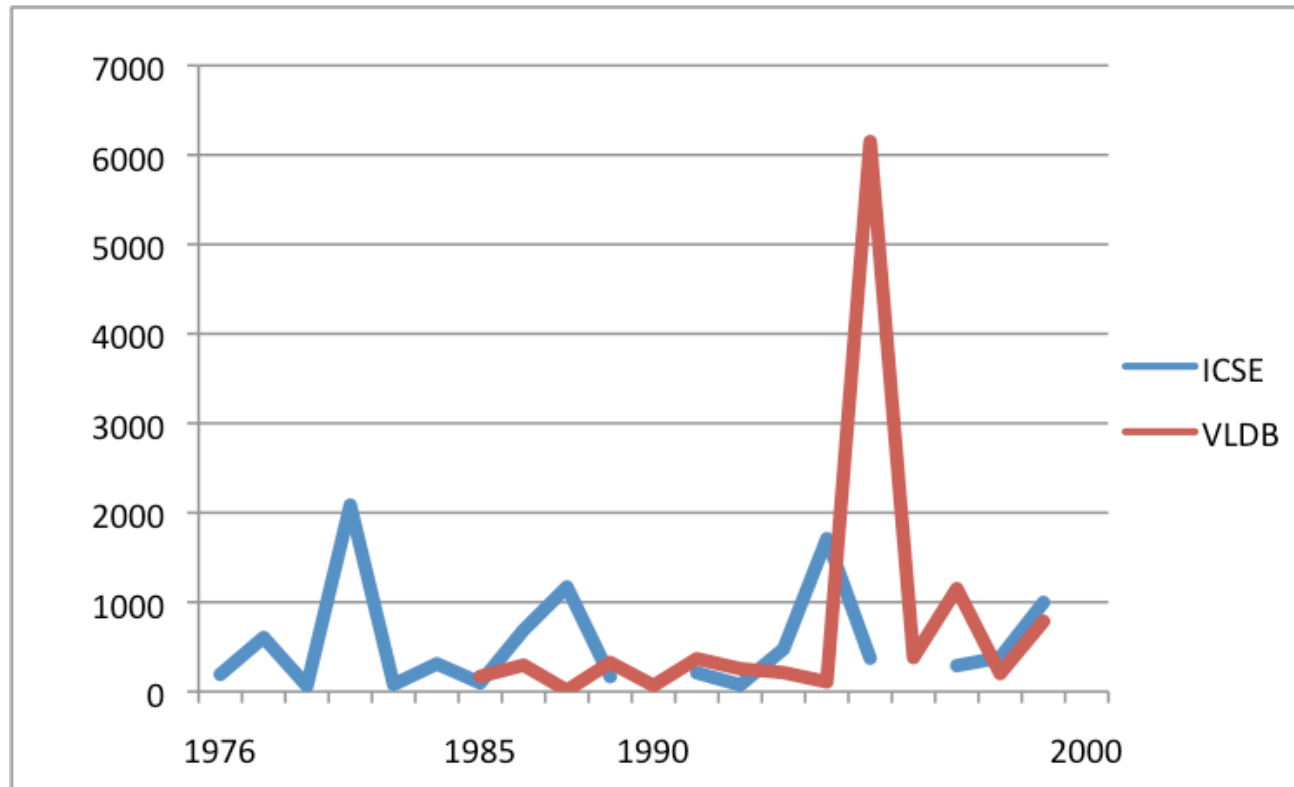
most influential papers

1. Program slicing (ICSE'81) **2120**
2. Statemate: a working environment for the development of complex reactive systems (ICSE'88) **1170**
3. N degrees of separation: multi-dimensional separation of concerns (ICSE'99) **997**
4. Bandera : extracting finite-state models from Java source code (ICSE'00) **791**
5. Software processes are software too (ICSE'87) **688**
6. Managing the development of large software systems: concepts and techniques (ICSE'87) **667**
7. Executable object modeling with statecharts (ICSE'96) **622**
8. Designing software for ease of extension and contraction (ICSE'78) **605**

MIP vs citations



How do we compare with others?



Citations of most influential papers

#Downloads (-12 months)?

From ACM DL

1. Automated Support for Development, Maintenance, and Testing in the Presence of Implicit Control Flow (ICSE'04): **781**---**13 cit**
2. Managing the development of large software systems: concepts and techniques (ICSE'87): **705**---**667 cit**
3. New directions on agile methods : a comparative analysis (ICSE'03): **567**---**161 cit**
4. Program slicing (ICSE'81): **520**---**2120 cit**
5. Quantitative evaluation of software quality ICSE'76: **495**---**212 cit**
6. Static detection of cross-site scripting vulnerabilities ICSE'08---**396**---**8 cit**
7. Analysis and testing of Web applications ICSE'01: **375**---**195 cit**
8. Aspect-oriented programming and modular reasoning: ICSE'05: **300**---**32cit**

Findings: #citations vs. #downloads

- Time factor (both #citations and #downloads)
 - my Jan-March numbers are different from the current ones
 - how can you compare #cit of a 20 years and a 2 years old paper?
- No significant correlation between citations and downloads
- #citations and #downloads express different merits
 - example: a highly downloaded paper may be useful in practice, but does not inspire further research
- #downloads demonstrates longevity of ICSE contributions
 - 3 out of the top 5 highest #downloads are papers over 20 years old, one is more than 30

Findings: #citations vs. expert judgment

- Citations and expert judgment are better correlated, according to the ICSE experience
 - most highly cited papers have been recognized as influential
 - ...but several influential papers have low citations

What do the numbers tell us?

Do they indicate impact?

- Does the number of paper citations indicate impact of a piece of work?
- Do citations of papers in a journal/proceedings indicate impact of a venue?
- Do citations of one's research measure impact of that research?



Part 2: A side road on numbers/citations



Joint Committee on Quantitative
Assessment of Research

Report CITATION STATISTICS

Intl Mathematical Union (IMU)

in cooperation with Intl Council of
Industrial and Applied Math (ICIAM)
and the Institute of Mathematical
Statistics (IMS), 6/12/2008

R. Adler, J. Ewing, P. Taylor (Eds.)

A side road on numbers/citations



Informatics Europe

Report RESEARCH EVALUATION
FOR COMPUTER SCIENCE

Viewpoint article in CACM, April 2009

Eds. B. Meyer, C. Choppy, J.
Staunstrup, J. van Leewen (Eds.)

Also D. Parnas, CACM nov 2007

STOP THE NUMBERS GAME

Facts

- Numbers increasingly used to assess/rank
 - papers
 - journals and conferences---the (in)famous *impact factor*
 - individuals
- Motivations
 - different stakeholders need to evaluate research
 - peer review can be subjective
 - peer review is expensive
 - ~~numbers are simple and objective~~ **simplistic and misleading**

Findings

- Much of the modern bibliometrics is flawed (statistics improperly used)
- Objectivity and accuracy illusory
 - *the meaning of a citation can be even more subjective than peer review*
- **Sole** reliance on citation data provides incomplete and shallow understanding of research
 - *only valid if reinforced by other judgments*

numbers are not inherently superior to and cannot substitute complex judgement

Why are citations counted?

From Thomson Scientific

"The value of information is determined by those who use it...the better way to measure the quality of work is by measuring the **impact** it makes on the community at large"

The statement makes an implicit equivalence assumption

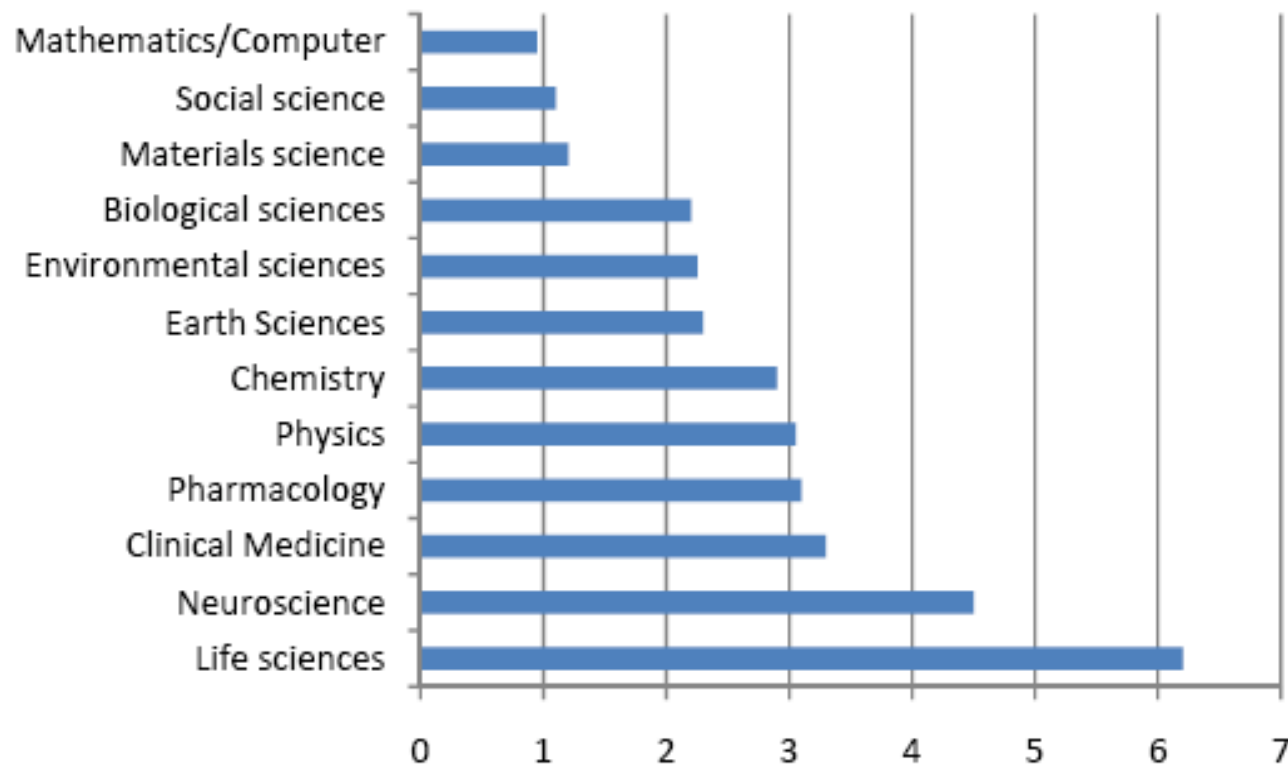
#citations = impact

Citations vs impact

- Is the equality
 #citations = impact
 justified? (When) does it work?
- It presupposes a highly rational and uniform model of reference giving, but this is NOT true

The complex sociology of citations

Average citations per article



citation practices differ substantially among disciplines

The complex sociology of citations

- Citations are of two main kinds
 - REWARD
 - intellectual dependency on the cited
 - RHETORICAL
 - a way to carry out a scientific conversation
 - reference included to show that topic is of interest to someone else or to prove that the author knows the literature
 - the cited explains some result, not necessarily of the cited author
 - the cited represents another approach, or is an example of...

The complex sociology of citations

- Most citations are rhetorical
- Reward citations can be of many kinds
 - currency, negative credit, operational information, persuasiveness, positive credit, reader alert, social consensus
- Obliteration effect
 - incorporated into other work, which is cited

An example

- The impact of MESA on system design, by HC Lauer, EH Satterthwaite, ICSE'79
 - 60 cit (highest of '79), 19 downloads ... one of the few papers on MESA

BUT

- Its impact evidenced by indirection
 - Implementing remote procedure calls, by Birrell&Nelson, ACM TOCS 1984, 1840 citations
 - explicitly states its debt to MESA

If impact is the question, are (citation) numbers the answer?

- Citation data and statistics provide some valuable information
 - very high citation numbers tell more than smallerbut only a limited and incomplete view of research quality
- Statistics derived from them often poorly understood and misused
- **Expert judgment cannot be eliminated**
- **Research too important to measure only with a coarse tool**

Lost in the journey?



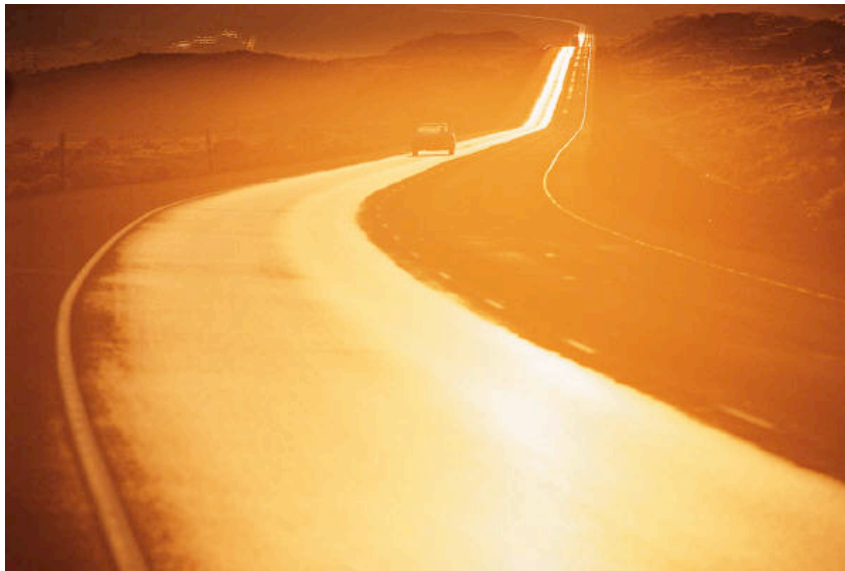
- We started by looking back at SE research through ICSE as a magnifying lens
- We tried to assess what we accomplished by identifying our impactful results
- We got trapped by the number game

Should we look for better assessment methods? Why?

Should we care?

- Our social responsibility is to care about *impact* of research
 - *understanding*
 - *measuring*
 - *improving*
- If we don't, others will do
 - governments, funding bodies
 - accountability, "return" to taxpayer or donor
 - universities, schools, departments
 - competition (CS versus other sciences, SE versus other CS areas)

Part 3: Where do we go from here?



Towards

- understanding
- measuring
- improving

impact of SE research

Do we understand what impact is? How can we measure it?

- Not quite
- A notable counter-example
 - IMPACT project
 - aims at demonstrating **impact of SE research on practice**
 - focuses on specific research areas in SE
 - an initiative of ACM SIGSOFT
 - [the SIGSOFT Impact project.webarchive](#)
 - area leaders responsible for research
 - backed by sound science history methodology



Published results

- Osterweil, L.J., Ghezzi, C., Kramer, J., Wolf, A.L. Determining the Impact of Software Engineering Research on Practice . IEEE Computer, March 2008.
- Estublier, J., Leblang, D., Hoek, A., Conradi, R., Clemm, G., Tichy, W., and Wiborg-Weber, D. Impact of software engineering research on the practice of software configuration management. ACM TOSEM, Oct. 2005.
- Ryder, B. G., Soffa, M. L., and Burnett, M. The impact of software engineering research on modern programming languages. ACM TOSEMI, Oct. 2005.
- Emmerich, W., Aoyama, M., Sventek, J. The Impact of Research on the Development of Middleware Technology. ACM TOSEM, August 2008.
- Clarke, L. A. and Rosenblum, D. S. Historical Perspective on Runtime Assertion Checking in Software Development. SIGSOFT SEN, March 2008.
- Rombach, D., Ciolkowski, M., Jeffery, R., Laitenberger, O., McGarry, F., Shull, S. Impact of research on practice in the field of inspections, reviews and walkthroughs: learning from successful industrial uses. In ACM SIGSOFT SEN, November 2008.

Some key findings/confirmations

- SE research has had impact on SE practice
- Maturation needs time
 - 15-20 years between first publication of an idea and widespread availability in products
- Substantially different mechanisms have been successful at causing impact
 - people are key to knowledge transmission
 - people movement
 - almost all impact traces lead back to PhD theses

Problems with impact definition/measurement

- Scholarly assessment of impact of SE research on practice is a difficult and expensive
 - it is *research*
- Substantially different methods are needed to assess impact of
 - journals/conference
 - individual researchers/papers

Obstacles to impact

- The dominant reward/recognition system based on publication count does not favor impact
 - from D. Parnas, CACM 2007
 - "it encourages superficial research"
 - "it encourages small, insignificant results"
 - "it rewards publication of half-baked ideas"
 - "it slow scientific progress: to increase their score, researchers must avoid tackling the tough problems and problems that will require years of dedicated work and instead work on easier ones"

Obstacles to impact

- Building on top of others' work too often too difficult or impossible
 - TOSEM papers 2001-2006
 - 60% refer to a tool, only 20% **installable**
- ICSE'06 paper by Zannier, Melnik, Maurer evaluates ICSE empirical studies
 - complete absence of replicated studies
 - self-evaluations dominate empirical studies

Conclusions: on impact

- Research on impact of SE research must continue and should become a community effort
- As a community we need to get to a common understanding and articulate methods for definition and measurement of impact for
 - journals/conferences
 - individual research
 - define impactful research products beyond paper count
- We must be aware of risks
 - be rigorous, avoid being incestuous

Conclusions: on enhancing research flow

- Define community challenges and support repeated experiments
 - catalogs of case-studies on which competing approaches can be applied and compared
- Favor research products backed by tools, but verify that the claimed results of their use can be replicated
- Use Internet facilities for community support
 - research-as-service
- Reward successful reuse of one's research rather than paper citation

The end of the journey



In 40+ years we have gone far

Beginning of a new journey



We have been successful, but we can do more and better

Towards impact-aware research

Can the ICSE community take the lead of this?

Acknowledgements

- L. Osterweil, J. Kramer, A. Wolf and IMPACT project members for their insights into impact of research
- Andrea Mocci, Mattia Monga, for helping me in mining data

Thank you!!!