30 Years of PODS in Facts and Figures*

Tom J. Ameloot† Maarten Marx ‡ Wim Martens § Frank Neven ¶

Justin Van Wees ‖

1 Introduction

The present paper discusses some facts and figures representing 30 years of history of the ACM Symposium on Principles of Database Systems (PODS). The analysis was done for the occasion of the PODS Anniversary Event held in Athens on June 12, 2011. All data corresponds to the period from 1982 up to and including 2011 unless explicitly mentioned otherwise and treats invited papers as regular papers. As the PODS Pages [1] are actively maintained, we do not repeat information here which can easily be found there. Also, supplementary material including word clouds, a snap shot of the PODS co-author graph, harvested data, spreadsheets, scripts, and the blueprint for the original PODS 30 year anniversary T-shirt can be found there as well.

2 Papers and People

Over 30 years 961 papers were published at PODS by 990 distinct authors. Table 1 depicts a more detailed overview of the distribution of papers per author. A large majority of PODS authors, that is, 64%, only published once while 90% of the authors published at most four papers. Still, there is a core of almost 100 hardliners that published 5 or more papers. Some of them are real devotees. Their papers show up at PODS every year. Record holder is Leonid Libkin whose longest uninterupted period of consecutive publications spans a period of 14 years. He is followed at a safe distance by by Moshe Y. Vardi, Dan Suciu, Georg Gottlob, Jan Van den Bussche, and Ron Fagin who all published PODS papers in 9 consecutive years. Figure 1 gives an overview of the longest streaks. Similarly, one can consider the longest period of absence at PODS, that is, the maximal time period between two PODS publications in which no article was published at PODS. Figure 2 provides a histogram displaying the longest period of absence. It is interesting to point out that although Ron Fagin published papers in PODS for 9 successive years, he was also absent for 12 years. The overall record here is 18 years of absence. Word clouds visualizing the devotees and the absentees can be found on the PODS Pages. In Figure 3, we display authors by their PODS score where each n-author paper attributes a score of $\frac{1}{n}$ to each co-author.

Number of papers	Percentage of authors
1	64%
2	17%
3	5%
4	4%
5 or more	10%

Table 1: Distribution of the number of papers per author.

3 Co-Authors

Figure 4 displays the distribution of the number of authors per paper. Surprisingly, about 25% of the papers are single-authored, while about 80% of all papers have three authors or less. Figure 5 shows the evolution of the number of single-author papers, displaying a decreasing trend. The latter is in line with the observation that over time the average number of authors of a PODS paper increased from 1.9 to 2.8. The maximum number of authors on a paper is 8. The paper was presented at PODS 2009 as an invited talk entitled “A web of concepts”. The reg-

*We acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under the FET-Open grant agreement FOX, number FP7-ICT-233599.

†Hasselt University & Transnational University of Limburg; Diepenbeek, Belgium. Email: tom.ameloot@uhasselt.be. PhD Fellow of the Fund for Scientific Research, Flanders (FWO).

‡Informatics Institute, Universiteit van Amsterdam; Amsterdam, The Netherlands. Email: maartenmarx@uva.nl

§Technical University of Dortmund; Dortmund, Germany. Email: wim.martens@udo.edu

¶Hasselt University & Transnational University of Limburg; Diepenbeek, Belgium. Email: frank.neven@uhasselt.be

‖Informatics Institute, Universiteit van Amsterdam; Amsterdam, The Netherlands. Email: justin@vwees.net

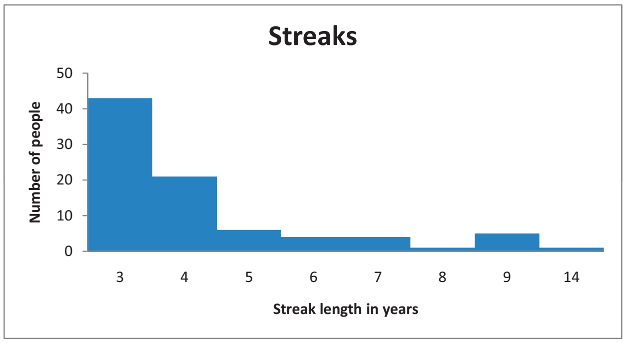


Figure 1: Histogram of the longest streaks (uninterrupted period of publishing by one author) at PODS.

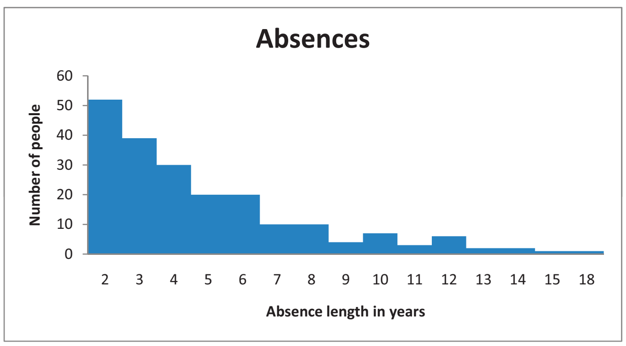


Figure 2: Histogram of the longest time periods between two successive PODS papers.

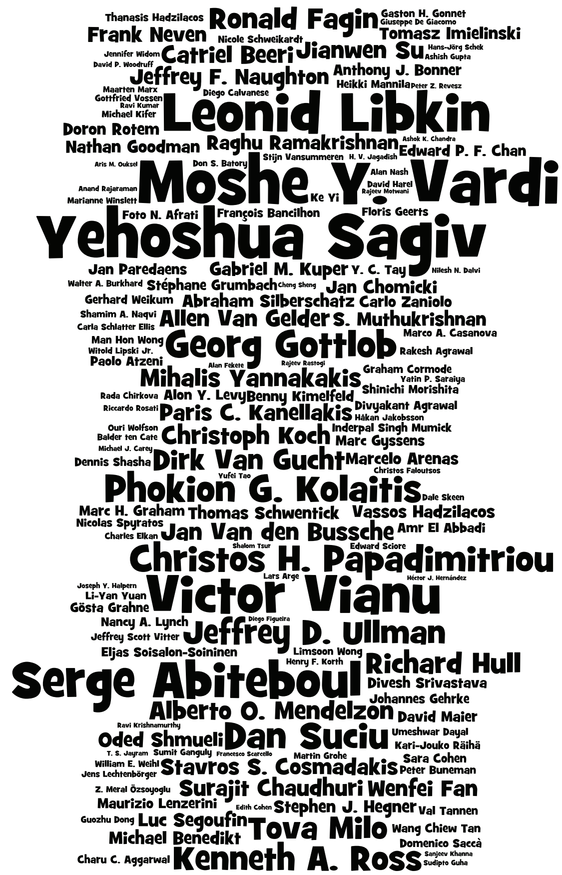


Figure 3: Most podsy researchers.

ular paper with most authors (7) was presented at PODS 2006 and was entitled "Achieving anonymity via clustering".

4 Community

Next, we try to provide insight into the community through an analysis of the structure of the co-author graph. We will use a number of measures from social network theory [2].

4.1 Betweenness centrality

Key nodes in a traffic network are those who lie on many shortest paths. The betweenness centrality of a node is the percentage of all shortest paths in a network which pass through that node. For each PODS year, we calculated this measure on the *complete* DBLP co-author graph restricted to the PODS authors, and restricted to publications up until that year. Thus, two PODS authors are connected in year x if they have a joint publication in DBLP (thus also in venues besides PODS) before or in year x. Referring to the most central person as the winner, Table 2 lists the most central persons per year and the number of times they won. The value of the centrality score of the winner has decreased over time. For instance, in 1982, 24% of all shortest paths went through Papadimitriou. The PODS co-author graph consisted then of 58 nodes. About ten years later, in 1991, Jeff Ullman won with 11%, while the winner from 2001 onwards, Serge Abiteboul, remained stable at 6% for 10 years and climbed to 7% in 2011. Table 3 lists the top 10 most central researchers in 2011 together with their score. The top 10 lists for all 30 years can be found at the PODS pages.

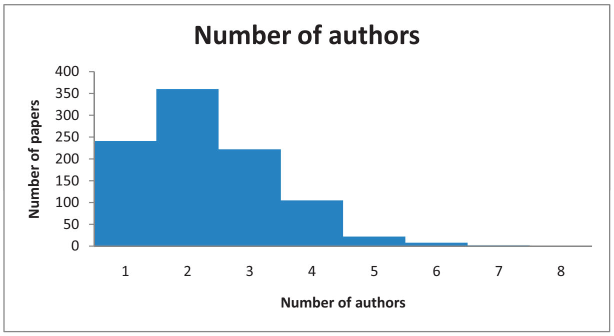


Figure 4: Distribution of the number of authors per paper.

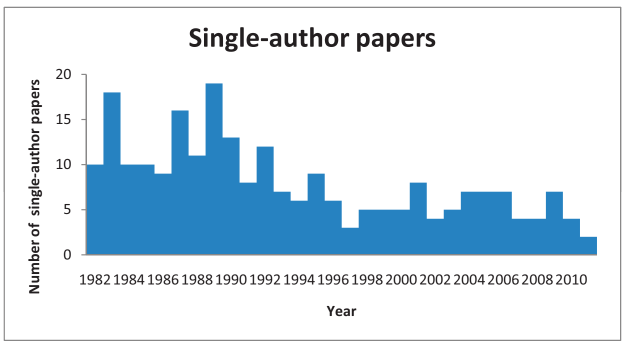


Figure 5: Evolution of single-author papers.

4.2 Connectedness

How well-connected is the PODS co-author graph, when restricting to only PODS publications? For starters, the PODS graph contains 990 nodes. The largest component grows steadily with an average of 26 new nodes yearly (over the last 20 years) and reaches 64% of all nodes in 2011 (7% in 1982, 36% in 1991 and 57% in 2000). Thus almost two thirds of all PODS authors are connected through a chain of PODS co-authorships. Now, what is the probability that two of your co-authors are co-authors themselves (a network transitivity measure)? This goes from 68% in the first PODS year, to 53% in 1985, to 37% in 1994, to a value between 32% and 35% (with mean 34%) in all later years. The community becomes larger, there are more nodes with high degree and thus it becomes harder to "know the friend of your friend". The average degree (number of researchers with whom you have at least one joint PODS paper) rises slowly from 1.21 (in 1982) via 2.23 (in 1991) and 2.83 (in 2001) to the current 3.38.

name	count	most central in
S. Abiteboul	14	89, 97, 99, 2001-2011
J. D. Ullman	7	87, 90–94, 98
C. H. Papadimitriou	6	82–85, 95, 2000
D. S. Parker Jr.	1	86
C. Beeri	1	88
A. Silberschatz	1	96

Table 2: Most central persons per year.

name	between. centrality
Serge Abiteboul	7%
Kenneth Ross	4%
S. Muthukrishnan	3%
Raghu Ramakrishnan	3%
Victor Vianu	3%
Christos H. Papadimitriou	3%
Jeffrey D. Ullman	3%
Abraham Silberschatz	3%
Dan Suciu	3%
Georg Gottlob	3%

Table 3: Ten most central PODS authors in 2011.

4.3 PODS communities

The set of PODS authors can be viewed as a Russian doll consisting of tighter and tighter communities. In a few steps, we reach a small core consisting mostly of prolific authors who were around since 1982. The obvious candidate for the first doll is the largest component (613 authors). For the next set of dolls we use a measure that has turned out to be very good in selecting subcommunities. It is a relaxation of the notion of a clique, called k-clique community [3]. Two k-cliques are considered adjacent if they share $k-1$ nodes. A community is then defined as the maximal union of k-cliques that can be reached from each other through a series of adjacent k-cliques.

The largest 3-clique community inside the largest component contains 125 nodes.[1] It has two largest 4-clique communities in it, both of size 15. These are disjoint, except for Ron Fagin. One of these components contains many prominent PODS authors. We have chosen this one as the next doll. This doll of 15 authors consists of a 5-clique and a 6-clique, both arising from a single paper[2] and four other authors.

On the PODS pages, we show a figure displaying

[1] The results in this section are calculated on the PODS 2010 graph which excludes the papers from PODS 2011.

[2] `http://www.informatik.uni-trier.de/~ley/db/conf/pods/pods96.html#AbiteboulKPV96` and `http://www.informatik.uni-trier.de/~ley/db/conf/pods/pods86.html#AfratiPPRSU86`.

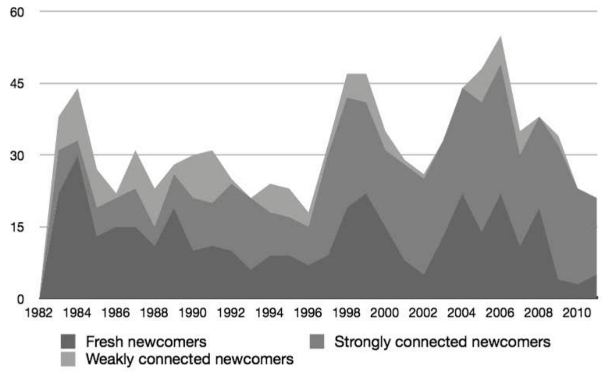


Figure 6: Number of newcomers per year.

Figure 7: Newcomers for the years 2005-2011 according to their PODS score.

the 125-node 3-clique community in blue with the two largest 15-node 4-clique communities it contains in pink and green. It also shows the next two largest 3-clique communities in green and orange. Looking at the names inside communities one cannot help but notice that cultural–geographic origin is a strong community–builder: Flanders and Greece–Israel provide strong binding force.

5 Newcomers

The number of newcomers per year is displayed in Figure 6. Newcomers are divided into three groups depending on their connection to the PODS graph of the year before. We have newcomers without any link to a PODS author ("fresh"), newcomers who are linked to a PODS author in the largest connected component ("strongly connected"), and those who are only connected to nodes outside the largest component ("weakly connected"). The average number of newcomers per year is 32 (over all years). Of these, on average, 40% are fresh, and 47% and 13% are strongly and weakly connected, respectively. Of these 32 newcomers, on average, 7.3 manage to get another PODS paper in the first three years after their first PODS paper. Just over half of these (3.9) even have at least two PODS papers in the first 5 years after their PODS-benjamin.

Which of the three groups of newcomers has the largest chance of reappearing in PODS? In other words, if you are a PODS newcomer, and intend to become a PODS regular, with whom should you write your first PODS paper? The results in Table 4 show that your chance or reemergence doubles when you, instead of starting fresh, join forces with a well-established PODS-member.

To get a feeling for the coming generation, Figure 7 displays the newcomers since 2005 according to their PODS score. On the PODS pages you can find a movie displaying the newcomers per year over the entire history of PODS.

6 Research Fields

PODS is influenced by and (hopefully) influences other fields of research. Such influence could be measured by examining the conference venues in which PODS-researchers publish. We give an overview how the relationship with other fields changes over time. To keep the analysis manageable, we restricted attention to the 100 venues in which PODS authors publish the most. Specifically, let p denote a conference paper by a PODS author. By $\mathit{year}(p)$ and $\mathit{field}(p)$, we denote the year of publication and the research field associated to p, respectively. Research field names were assigned uniformly for all papers in the same conference. For a research field f, a start year y_1 and an end year y_2, the influence of the PODS authors on (or from) field f during the period spanned by the interval $[y_1, y_2]$, denoted $\mathit{influence}(f, y_1, y_2)$, is defined as

$$|\{\text{paper } p \mid y_1 \leq \mathit{year}(p) \leq y_2 \wedge f = \mathit{field}(p)\}|\,.$$

Figure 8 displays the influence for three 10-year periods of time. While the method of computing influence can be criticized, the figures do confirm some

	All	Strongly Connected	Weakly Connected	Fresh
1 paper in 3 years	22%	30%	20%	15%
2 papers in 5 years	12%	16%	12%	7%

Table 4: Chance of reappearance of PODS newcomers, with at least one paper in the three years after first PODS publication and at least two papers after five years. Averages taken over the years 1983 to 2008 (3 years) and 1983 to 2006 (5 years).

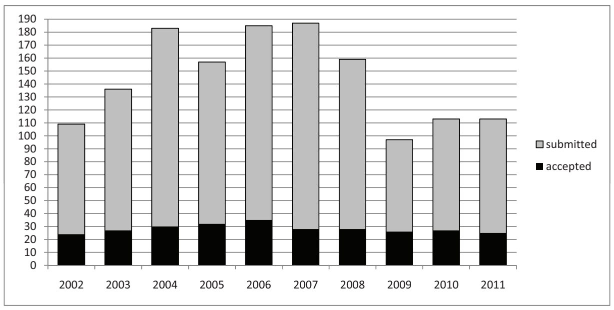


Figure 9: Number of PODS submissions and accepted papers per year for the period 2002–2011.

Figure 10: Locations of PODS in the period 2002–2011. The more submissions, the larger the font of the location.

global trends which on an intuitive level correspond to the reality. First of all, over time, the community became much more diverse as can be seen by the explosion of related research fields. Secondly, PODS researchers publish increasingly more papers outside the database theory community. Indeed, while the number of PODS-papers (or ICDT-papers for that matter) did not drastically change over the years, the database theory pie decreases from 15% to 5%, in favor of for instance, database systems which increased from 23 to 32%.

7 Conference statistics

Figure 9 displays the number of submissions as well as the number of accepted papers over the last ten years.

The average acceptance rate is 20%. Even though these figures are discussed at every PODS business meeting, there does not seem to be a satisfying explanation to why they are what they are. To provide an alternative view, Figure 10 relates submission numbers to conference locations.

8 Greek Gods of Python

During a divine intervention at the anniversary session two Greek gods suddenly materialized. The hymn they performed is depicted in Figure 11.

9 Outlook

There were several things we would liked to have investigated but where held back by constraints like lack of time and absence of data. Although the presented overview is mostly biased towards publications, we should not forget that the success of a community is also reflected by the number of people attending its main event even when they do not have a paper. It would be interesting to know how successfull our community is in this respect. Of course, as PODS merged with SIGMOD in 1996, the task no longer reduces to matching registrations to papers. We did not consider citations. Interestingly, DBLife [3] maintains a list of the top-300 most cited PODS-papers.

[3] http://dblife.cs.wisc.edu/

When your paper got rejected
by a review that corrected
your main result that turned out to be wrong.
Don't feel down, you're blessed!
It's Science that progressed.
So let's stand up and burst into a song.

And... Always look on the bright side of life...
Always look on the light side of life...

If your review pile just grows
to heights nobody knows
and the definition sections are unclear.
Don't let it make you mad.
Just think: "It's not so bad!
At least I didn't write the crap that's here."

And... Always look on the bright side of life...
Always look on the light side of life...

If you cannot comprehend
the talks that you attend
'cause they're just throwing formulas at you
and wireless is down.
Don't worry 'bout your frown.
No one knows that you don't have a clue

So... Always look on the bright side of life...
Always look on the light side of life...

If a student is so kind
to find a proof of just one line
for the central theorem of your PhD.
Don't feel bad or dumb
or that your life has gone:
You still were first, chronologically.

And... Always look on the bright side of life...
Always look on the light side of life...

If a talk you just gave
causes a big wave
of questions why your work is really new.
Wasn't it proved in 1960
by someone called Büchi?
Then at least you know that his result is true.

Always look on the bright side of life...
Always look on the right side of life...
Always look on the bright side of life...
Always look on the bright side of life...

Figure 11: Lyrics of "Always Look on the Bright Side of Life" (originally by Eric Idle), adapted for the 30th Anniversary of PODS.

Acknowledgements

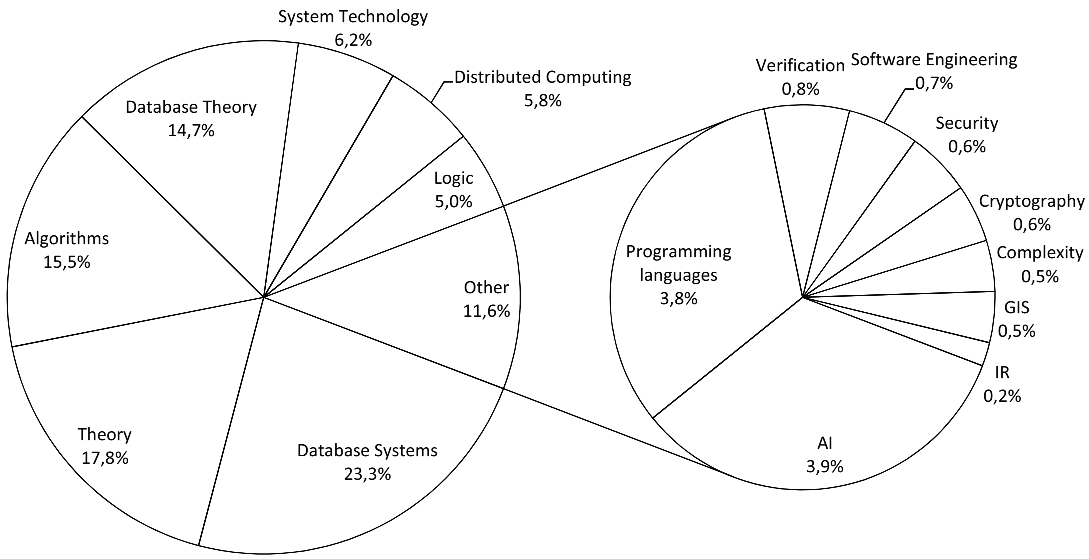
This analysis was only possible with the limited manpower we had because of the unsurpassed DBLP database created and maintained by Michael Ley.

References

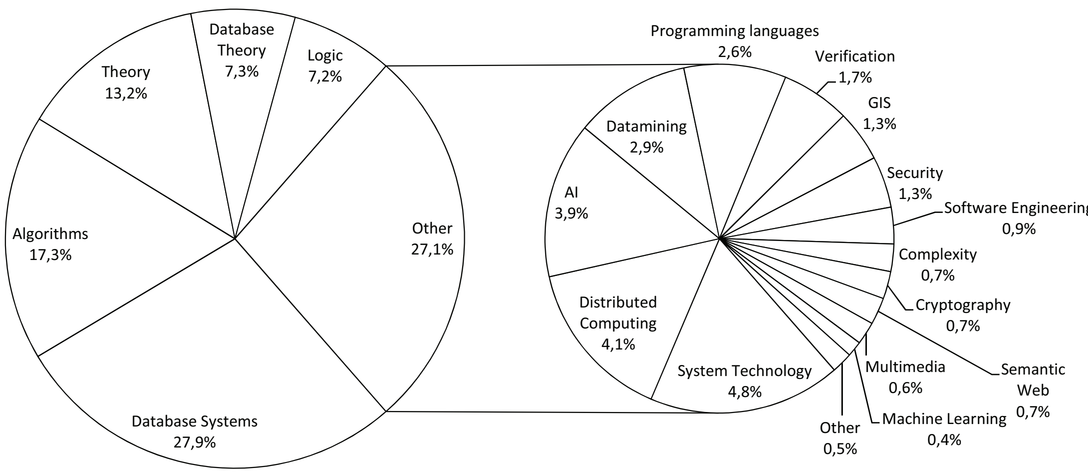
[1] The Pods Pages. Available at `http://www.sigmod.org/the-pods-pages`.

[2] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

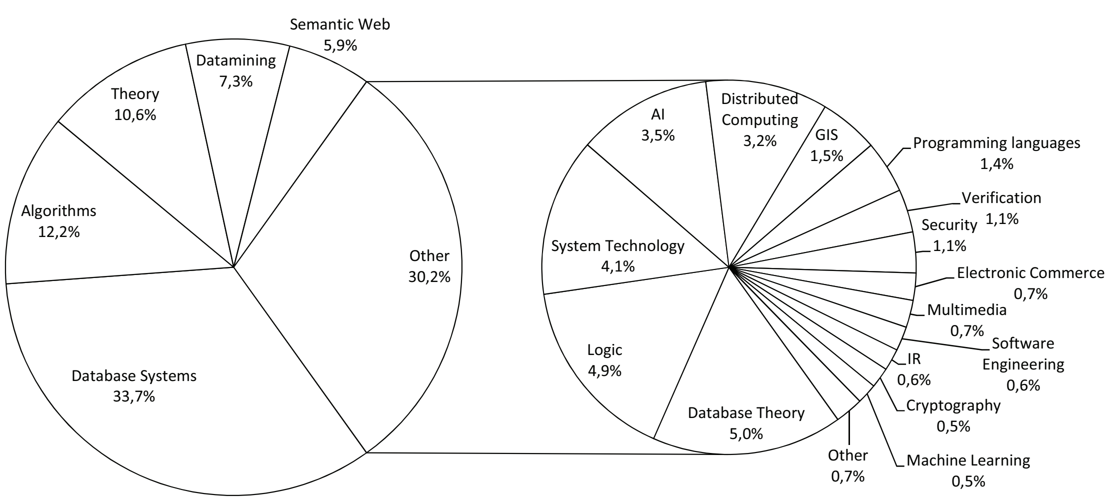
[3] Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *NATURE*, 435:814–818, 2005.



(a) Period 1982-1991



(b) Period 1992-2001



(c) Period 2002-2011

Figure 8: Influence on research fields.