Analysing Scholarly Communication Metadata of Computer Science Events

Said Fathalla[1,3(✉)], Sahar Vahdati[1], Christoph Lange[1,2], and Sören Auer[4,5]

[1] Smart Data Analytics (SDA), University of Bonn, Bonn, Germany
{fathalla,vahdati,langec}@cs.uni-bonn.de
[2] Fraunhofer IAIS, Sankt Augustin, Germany
[3] Faculty of Science, University of Alexandria, Alexandria, Egypt
[4] Computer Science, Leibniz University of Hannover, Hannover, Germany
[5] TIB Leibniz Information Center for Science and Technology, Hannover, Germany
soeren.auer@tib.eu

Abstract. Over the past 30 years we have observed the impact of the ubiquitous availability of the Internet, email, and web-based services on scholarly communication. The preparation of manuscripts as well as the organisation of conferences, from submission to peer review to publication, have become considerably easier and efficient. A key question now is what were the measurable effects on scholarly communication in computer science? Of particular interest are the following questions: Did the number of submissions to conferences increase? How did the selection processes change? Is there a proliferation of publications? We shed light on some of these questions by analysing comprehensive scholarly communication metadata from a large number of computer science conferences of the last 30 years. Our transferable analysis methodology is based on descriptive statistics analysis as well as exploratory data analysis and uses crowd-sourced, semantically represented scholarly communication metadata from OpenResearch.org.

Keywords: Scientific events · Scholarly communication · Semantic publishing · Metadata analysis

1 Introduction

The mega-trend of digitisation affects all areas of society, including business and science. Digitisation is accelerated by ubiquitous access to the Internet, the global, distributed information network. Data exchange and services are becoming increasingly interconnected, semantics-aware and personalised. Further trends are crowd-sourcing and collaboration, open data as well as big data analytics. These developments have profound effects on scholarly communication in all areas of science. We particularly focus on computer science, where conferences and workshops are of paramount importance and a major means of scholarly communication. Online platforms and services such as *EasyChair*[1] or

[1] http://easychair.org.

J. Kamps et al. (Eds.): TPDL 2017, LNCS 10450, pp. 342–354, 2017.
DOI: 10.1007/978-3-319-67008-9_27

CEUR-WS.org[2] automate and optimise scholarly communication workflows. A key question now is: What were the measurable effects of digitisation on scholarly communication in computer science? Of particular interest are the following questions: (*a*) Did the number of submissions increase? (*b*) Is there a proliferation of publications? (*c*) Can we observe popularity drifts? (*d*) Which events are more geographically diverse than others? We shed light on some of these questions by analysing comprehensive scholarly communication metadata from computer science conferences of the last 30 years. Large collections of such data are nowadays publicly available on the Web. Research has recently been conducted to browse and query such data [6,7], with a focus on authors, publications and research topics [4].

We analysed the evolution of key characteristics of scientific events over time, including frequency, geographic distribution, and submission and acceptance numbers. We analysed 40 conference series in computer science with regard to these indicators over a period of 30 years. Our analysis methodology is based on descriptive statistics analysis, exploratory data analysis and confirmatory data analysis. This article is organised as follows: Sect. 2 gives an overview on related work. Section 3 presents the methodology we used. Section 5 discusses the results of our evaluation. Section 6 concludes and outlines future work.

2 Related Work

Conference metadata and bibliography services. A lot of research has been performed to reveal information about scholarly communication from bibliographic metadata. *DBLP* and *DBWorld*[3], the most widely known bibliographic databases in computer science, provide information mainly about publications and events but also consider related entities such as authors, editors, conference proceedings and journals. *WikiCFP*[4] is a popular service for publishing calls for papers (CfPs). *Springer LOD* and *ScholarlyData*[5] publish as Linked Open Data metadata of conference related to computer science collected from Springer's traditional publishing process.

Conference series analysis. For various conference series, analyses similar to ours were performed by steering committee members or other members of the community. They often include the analysis of bibliographic data of each edition and rarely comprise comparisons with other events or editions of the same event series. A comprehensive analysis of the Principles of Database Systems (PODS) conference series includes detailed author analyses such as the distribution of the number of papers per author, which, for example, shows that two thirds of the authors are only involved in a single PODS publication (e.g., PhD students) but 10% are involved in 5 or more (e.g. active supervisors) [1]. It includes a relatively

[2] http://ceur-ws.org.
[3] http://dblp.uni-trier.de/, https://research.cs.wisc.edu/dbworld/.
[4] http://www.wikicfp.com/.
[5] http://lod.springer.com/, http://www.scholarlydata.org/dumps/.

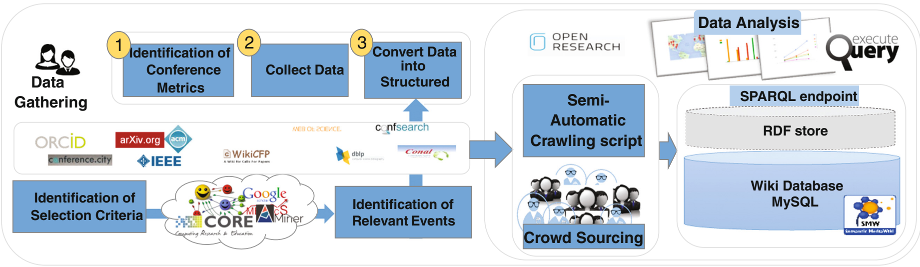


Fig. 1. Overall workflow of this study

short analysis of submission and acceptance rates for 10 years (2002–2011) that shows an increasing number of submissions in the beginning of the period, while they reduced in the last four years.

Literature Overview. Se and Lee proposed a list of alternative measures for ranking events [9]. The *goodness* of events (conferences and journals) is defined as the goodness of the articles published in these events. Biryukov and Dong addressed collaboration patterns among a research community using information of authors, publications and conferences [3]. Similarly, Aumüller and Rahm analysed affiliations of database publications using author information from DBLP [2]. A pilot study with a different focus analysed submissions to top technical conferences in computer science [5], while our analyses are about the quality of events considering different metrics than only metadata about publication and authors.

3 Method

The overall workflow of this study (see Fig. 1) comprises four steps: (1) identification of relevant events, (2) data gathering, (3) ingestion into the OpenResearch.org semantic scholarly communication data curation platform, and (4) data analysis.

Identification of Relevant Events. To identify a subset of high qualified events to which we can apply our evaluation, we collected all the metrics which are used by most of the well-known services. The analysis focuses only on conferences because of the high impact to the research community. However, all these metrics can be applied for more types of events. Depending on availability and re-usability of the metrics, the following set of criteria is finalized to be used in this study (see Table 1):

h-index Google Scholar Metrics (GSM)[6] provides ranked lists of conferences and journals by scientific field based on a 5-year impact analysis over the Google

[6] https://scholar.google.com/intl/en/scholar/metrics.html.

Table 1. Identification of relevant event criteria

Ranking metric	TPDL	WWW	PERCOM	COLT	EuroCrypt	CHI	CAV	PLDI
CORE rank	A*	A*	A*	A*	A*	A*	A*	A*
H5 index	74	66	28	22	50	83	39	45
Qualis	B1	A1	A2	A2	A1	A1	A1	A1

Scholar citation data. The ranking is based on the two metrics h5-index and h5-median. *Qualis*[7] uses the h-index as performance measure for conferences. Based on the h-index percentiles, the conferences are grouped into performance classes that range from `A1` (best), `A2`, `B1`, ..., `B5` (worst).

Mix of indicators. The *Computing Research and Education Association of Australasia* (CORE)[8] provides a ranking method for major conferences in computing. The ranking is determined by a mix of indicators including citation rates, paper submission, acceptance rates and the visibility and research track record of the key people hosting and managing the conference. Based on these metrics an event can be categorised into six classes `A*`, `A`, `B`, `C`, `Australian`, and unranked. The portal shows international event series in the first four categories.

Data Gathering. Data gathering is the process of collecting data from a variety of online sources in an objective and unbiased manner. We collected metadata about 40 conference series in different computer science sub-fields from different sources of metadata including title, series, sub-field, start date, end date, homepage, country and Twitter account. This information is available as Linked Data in the case of DBLP, and other structured forms, or semi-structured and unstructured in the case of WikiCFP, the ACM digital library[9], or conference.city[10]. The *OpenResearch.org* wiki[11] serves us both as an additional source of semantically structured data, and as a tool to support data analysis. At the time of writing, OpenResearch contains crowd-sourced metadata about more than 5000 conferences, 900 workshops and 350 event series. OpenResearch supports researchers in collecting, organising, sharing and disseminating information about scientific events, tools, projects, people and organisations in a structured way [8].

Data Preprocessing. In this step, we carried out several preprocessing tasks including:

Data Integration/Transformation: This step starts with identifying inadequate, incorrect, inaccurate or irrelevant data and then filling in missing data, deleting

7 http://www.conferenceranks.com/.
8 http://www.core.edu.au/.
9 http://dl.acm.org/.
10 http://www.conference.city/.
11 http://openresearch.org.

the dirty data, and resolving inconsistencies. In the data integration process, we combine data from multiple sources into meaningful and valuable information. Transformation is the conversion of cleaned data values from unstructured formats into a structured format.

Conference Name Unification: Looking into the collected data we found that some events have changed their names once or more since they had been established. This led us to perform a unification process before beginning to analyse the data. The unification process integrates all events of a series with multiple names under its most recent name because it is important for the researchers

Table 2. Conference title and acronym evolution for Some Conferences.

Unified acronym	Acronym	Full conference title	Time span
IEEE VR	IEEE VR	IEEE Virtual Reality	1999–2017
	VRAIS	Virtual Reality Annual International Symposium	1993–1998
ASE	ASE	Automated Software Engineering	1997–2017
	KBSE	Knowledge-Based Software Engineering Conference	1990–1996
ISWC	ISWC	International Semantic Web Conference	2002–2017
	SWWS	Semantic Web Working Symposium	2001
FOCS	FOCS	Annual Symposium on Foundations of Computer Science	1975–2017
	SWAT	Annual Symposium on Switching and Automata Theory	1966–1974
	SWCT	Annual Symposium on Switching Circuit Theory and Logical Design	1960–1965
ISMAR	ISMAR	International Symposium on Mixed and Augmented Reality	2002–2017
	ISAR	International Symposium on Augmented Reality	2000–2001
	IWAR	International Workshop on Augmented Reality	1999
ISSAC	ISSAC	International Symposium on Symbolic and Algebraic Computation	1988–2017
	SYMSAC	Symposium on Symbolic and Algebraic Manipulation	1966,1971,1976,1981 and 1986
	EUROSAM	International Symposium on Symbolic and Algebraic Computation	1974, 1979, 1982 and 1984
SPLASH	SPLASH	Systems, Programming, Languages and Applications: Software for Humanity	2010–2017
	OOPSLA	Conference on Object-Oriented Programming, Systems, Languages, and Applications	1986–2009

who want to submit their work to know the recent name rather than the name that had been in use for the longest time, as shown in Table 2. For example, SPLASH is the unified name of a Conference on Object-Oriented Programming, Systems, Languages, and Applications which was named SPLASH from 2010 to 2017 and previously OOPSLA from 1986 to 2009, i.e., for 24 years.

Ingestion into OpenResearch. The collected data can be ingested into OpenResearch.org in several ways using either single or bulk import. For single import, one should use semantic forms. The required steps for bulk import are: (a) Create a spreadsheet with the important information, (b) Export the spreadsheet to CSV, (c) import CSV file using OpenResearch's *ImportCSV* service.

4 Data Analysis

The heart of our work is an explorative analysis of the metadata of selected computer science conferences over the past 30 years.

Metrics and Analysis Tools. We first defined metrics, then chose suitable tools for computing them and evaluating the results of the computation.

We defined statistical metrics over numeric values, as well as metrics having other complex datatypes, focusing on conferences because of their high impact on research communities. We chose spreadsheets as the main tool to compute statistical metrics over numeric values; the evaluation of the results is supported by charts. OpenResearch provides further components for visual analytics, in particular for displaying non-numeric results (e.g., the conferences with the highest number of submissions). Even though spreadsheets are, in principle, based on the relational data model, they practically lack support for joins across sheets. Joins may be required for connecting information about events to information about related entities, such as persons participating in events. The SPARQL query language for RDF, which is supported by OpenResearch, facilitates such join computations. However, while SPARQL also supports basic statistical analysis via aggregate functions, this type of analysis is better supported by spreadsheets.

Statistical Analysis. Acceptance rate is defined as the ratio between submitted and accepted articles.

Continuity refers to how continuously a conference has been held over its history. We propose a formula $C = \min\{100\%, (E * R)/A\}$ to calculate the percentage of continuity for a specific conference where C stands for continuity, E for the number of editions of the event, R for the regularity of the event editions (1 for 'every year', 2 for 'every two years'), and A for the age, counting the number of years since the first time the event was established. Year is the granularity for this metric.

Geographical Distribution: Every event is held in a geographical *Location*. We consider it as a triple of City, Country, Continent. From the extension of this metric to event series, one can derive the number of distinct locations visited by an event. We map every distinct location to the number of times the event has taken place there (by city, country or continent). We can thus classify event series by their most frequent location, e.g., as a "German" or "European" series. Geographical Distribution of an event series increases the awareness of researchers about the existence of the event and its covered topics.

Time Distribution: Every event is held in a certain period of time each year. It is important for a researcher interested in a particular conference to know when this conference will be held in the year to know when to prepared and present their work.

Sub-field Popularity: In the sub-field popularity metric, we divided conferences into five groups, each of which is labelled with the sub-field of computer science they belong to. We considered two time intervals: three 10-years periods for accepted papers but three 5-years periods for submitted papers due to the difficulty to obtain information about the number of submitted papers for many conferences. Table 3 shows research communities and corresponding conferences investigated.

Table 3. CS Sub-fields and top conferences

Acronym	CS sub-field	Conferences
GRA	Computer Graphics	ACMMM, EuroGraphics, IEEE VR, SIGGRAPH
SEC	Computer Security	CCS, CRYPTO, EuroCRYPT, ASIACRYPT
PROG	ProgrammingLanguages	ICFP, PLDI, POPL, SPLASH
SE	Software Engineering	ICSE, FSE, ASE, FASE
DB	Database Systems	PODS, SIGMOD, ICDT, VLDB

Field Productivity: Field Productivity reveals how much interest there is in a computer science sub-field in a given year within the past 30 years. The Field Productivity (FP) for a sub-field (f) in a year (y), where $C_{i,y}^{f}$ is the number of publications for a conference i in year y and n is the number of conferences belonging to sub-field *f*, and m is the number of years in the time span of the study, is defined as $FP_y^f = \frac{\sum_{i=1}^{n} C_{i,y}^f}{\sum_{k=1}^{m}\sum_{i=1}^{n} C_{i,y_k}^f}$.

Entity-Centric Visual Analytics. In contrast to spreadsheets and their charting facilities, OpenResearch makes it easy to generate visualisations that focus on entities rather than numbers. Besides geographical maps and ranked tables or lists, timelines are a prominent example of entity-centric visualisations. The input for a timeline is provided by a query in the MediaWiki expression language. The following code, for example, defines a timeline of events with upcoming submission deadlines:

```
{{#ask: [[Category:Event]]
        [[submission deadline::>{{CURRENTYEAR}}...]]
        [[Category:{{#urlget:field}}]]
 | ?title = Name        | ?abstract deadline
 | ?submission deadline | ?notification
 | ?Category:Conference = Conference
 | ?Category:Workshop   = Workshop
 | format=timeline      | sort=submission deadline}}
```

Similar types of queries that we have implemented in OpenResearch include: – event series in a given field and their average acceptance rates, – countries with a high number of events in a given field, – fields with decreasing numbers of accepted papers over years, In addition to querying the data inside the OpenResearch wiki, queries to external SPARQL endpoints can be embedded into wiki pages using the *LinkedWiki*[12] extension for MediaWiki.

Joins Across Entity Types. OpenResearch currently focuses on semantic representation of CfPs as one wiki page per event, but including semantic relations to related entities, e.g., to document the role that a person had in the organisation of an event. A concrete use case for querying this data is supporting the research community in taking decisions on what conference to submit one's results to, or whether to accept invitations for assuming certain roles in the organisation of a certain conference. Such queries often require joins across multiple entity types. Simple queries of this kind can be implemented in the MediaWiki expression language introduced in Sect. 4, more complex one require SPARQL. The output of both kinds of queries can be a table, list, map, timeline, etc.

Consider, for example, finding all roles that a person has ever had in events; this requires joins between person and event entities:

```
SELECT ?event ?person ?hasRole WHERE {
  ?e       rdfs:label         ?event .
  ?e       ?hasRole           ?person .
  ?hasRole rdfs:subPropertyOf property:Has_person .
  ?person  rdfs:label        "PERSON NAME" .}
```

Geographical distribution and affiliation changes of persons in the role of general chairs of events related to a certain field over last 10 years can be shown on a map or graph by embedding a SPARQL query as follows into the wiki page representing a certain field (i.e., in MediaWiki, a *category* page):

```
{{#sparql: SELECT ?event ?country ?person WHERE {
  ?e  a     category:Semantic_Web .
  ?p  property:Has_location_country ?country .
  ?p  property:Has_affiliation  ?organization .
  [...]
  MINUS{ ?e  property:Has_general_chair  :person . }
  FILTER (?startDate >= "2007-01-01"^^xsd:date && ?endDate   <  "2017-01-01"^^xsd:date )
  } LIMIT 10  | format=maps}}
```

[12] https://www.mediawiki.org/wiki/Extension:LinkedWiki.

5 Observations

In this section we report detailed analysis results for 40 conference series over a period of 30 years according to six statistical analysis dimensions. The complete raw data is available at https://goo.gl/vnsXRe.

5.1 Statistical Analysis

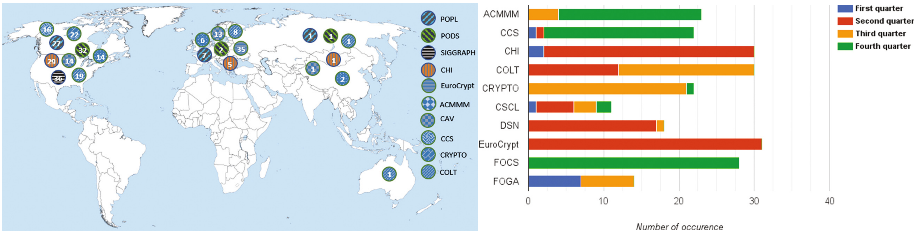
Acceptance Rate. We have selected Fig. 3(a) shows the average acceptance rate for a sample of 10 conferences from different CS sub-fields in five consecutive 5-year periods from 1992 to 2016.

In all three periods, the average acceptance rate for all series falls into the range 17% to 26% in the time window of 25 years. The greatest acceptance rate ever was the one of COLT in the second period (45%), but decreased it to 36% in 2016. The average acceptance rate of EuroCrypt had increased to 33% by 2011 before decreasing to final 23%. The average acceptance rate of CCS dramatically decreased. The number of submissions to this series increased over time; however, the acceptance rate remained approximately the same. Only the average acceptance rate of EuroCrypt significantly increased to 33% in 2007–2011 and then decreased again to 20% in 2012–2016. A reason for decreasing acceptance rate can be increasing submissions, with the number of presentation slots at a conference being more or less constant over time.

Continuity. The continuity of conferences is calculated using the proposed formula in Sect. 4. For example, the continuity of CCS (ACM Conference on Computer and Communications Security) is 92% where it was held every year from 1993 except for two years in 1995 and 2003. Moreover, the continuity TPDL (The International Conference on Theory and Practice of Digital Libraries) is 100% where it is occurring every year since the first year of establishment. For illustration, the continuity of five conferences are shown in Table 4; for the others, the continuity is 100%. Overall we observed a very high continuity among the renowned conferences.

Table 4. Continuity of five conference series

Conference	Age	Editions	Regularity	Continuity (C)
ACMMM	23	22	1	96%
CCS	24	22	1	92%
CHI	35	34	1	97%
FOGA	27	13	2	96%
TPDL	21	21	1	100%



(a) Geographical distribution (b) Time distribution

Fig. 2. Geographical and time distribution of events.

Geographical Distribution. The EUROCRYPT conference series has been held in a different country every year since 1987 but always in Europe. This is mostly related to the organisation committee in this series since it is a European committee. For the same reason, the SIGGRAPH series has been held every year since 1974 in different North American countries (mostly in the US).

The FOCS series has been held 26 times in the US, every year since 1989 in North America, and in Europe only for one edition in 2004. On the contrary, ISSAC has been moving between different countries of different continents such as Japan, Canada, Germany, etc., since its first edition. Figure 2(a) shows the Geographical Distribution of a sample of ten conference series randomly selected. The most geographically diverse conference series are EUROCRYPT (diversity by country in Europe). The most static conference is FOCS series that has been held 26 times only in the US for the past 25 years.

Time Distribution. Most editions of top conference series are held around the same month of each year; see Fig. 2(b). Namely, the PERCOM conference (IEEE International Conference on Pervasive Computing and Communications) has been held every year since 2003 in March and POPL (ACM SIGACT Symposium on Principles of Programming Languages) has been held every year since 1994 in January. Furthermore, almost all conferences in the study have been established around the same month. For example, EuroCrypt is always held in April or May and SIGGRAPH always held in July or August.

Sub-field Popularity. There are five groups labelled: Computer Security (SEC), Computer Graphics (GRA), Database Systems (DB), Programming languages (PROG) and Software Engineering(SE) each of which contains four top conferences belonging to this sub-field. Table 5 compares five CS communities in terms of the number of accepted and submitted papers. GRA communities made the largest number of submission in the whole time span, even though GRA submissions began to decrease since 2005 until they reached their minimum value in the last period. The average number of accepted papers (Fig. 3(c)) in GRA doubled in the first time frame and increased to almost 150% in the past 10 years, similarly in SE. The average number of accepted papers in DB slightly increased

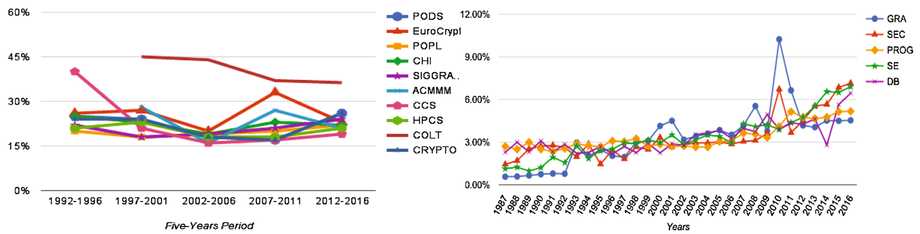
Table 5. Accepted and submitted papers measures for five CS sub-fields over three 10-years and three 5-years intervals respectively

		Computer Graphics			Computer Security			Programming Languages			Software Engineering			Database Systems		
		Avg.	Min	Max	Avg.	Min	Max	Avg.	Min	Max	Avg.	Min	Max	Avg.	Min	Max
Accepted Papers	1987–1996	78.9	33	172	106.2	68	135	103.8	90	119	65.1	36	101	135.7	117	166
	1997–2006	198.4	113	261	130.1	86	157	110.6	102	125	116.1	105	130	161.9	121	206
	2007–2016	302.2	219	593	235.2	144	337	170.3	128	199	189.5	144	256	240.7	151	347
Submitted Papers	2000–2004	927	535	1,182	633	513	849	522	481	576	709	585	879	905	718	1,207
	2005–2009	1,188	1,090	1,454	855	607	988	594	568	635	904	803	1,038	1,250	1,166	1,348
	2010–2014	1,304	1,017	1,786	1,122	936	1,264	754	676	827	992	837	1,170	973	548	1,109

in the first period and then again increased in the last 10 years by 50%. Over all three periods, the GRA community has attracted most, and PROG has attracted the least submissions. Overall, there is an increasing number of submissions for all CS sub-fields we considered (Fig. 3(d)

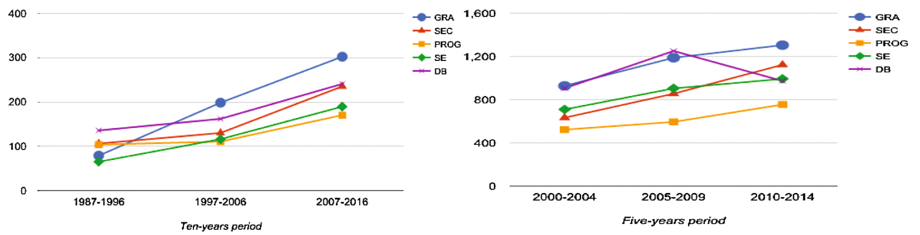
Field Productivity. We calculated the Field Productivity for the five sub-fields in the study. The results are shown in Fig. 3(b).

We found that PROG and DB remained at the same FP with some ups and downs from 1987 to 2010 and then saw a slight increase. At the end of the 1980s and the early 1990s, GRA had the lowest FP with less than 1% until it began to increase to around 3% by 1993 and continued increasing to around 10% before decreasing to only 4% by the end of the period.



(a) Avg. Acceptance Rate for Ten Top Conference Series

(b) Field Productivity for five CS sub-fields



(c) Avg. No. of Accepted papers of five CS sub-fields

(d) Avg. No. of Submitted papers of five CS sub-fields

Fig. 3. Visualisation of observations

Moreover, all fields had an FP around 3% from 1987 till 2006. For instance, FP of SE varied between 1.13% and 2.97%. In addition, GRA reached the maximum FP in 2010 with 10% and DB reaches the maximum FP with 6.45% in 2016. Overall, GRA has the highest FP with 5,795 publications over the other fields; the PROG community has the lowest FP with 3,707 publications. The DB community ranks second with 5,383 publications, followed by SEC with 4,715 publications and then PROG with 3,847 publications.

6 Conclusions

We presented a method for analysing scholarly communication metadata of scientific events. We combined descriptive and exploratory analysis with regard to a broad set of metrics, supported by spreadsheets, charts and queries in the OpenResearch semantic wiki. Up to our knowledge for the first time, we were able to empirically validate the often raised concern of a proliferation of submissions to major conferences. Also, we were able to calculate and demonstrate with our method a number of other indicators, such as a new way to calculate conference continuity, the popularity of different sub-fields, a new way to calculate field productivity or the geographic distribution of conferences. In addition to efficiency gains, the digitisation of scholarly communication also has negative impacts, most significantly the proliferation of submissions, which significantly increases the reviewing workload with an already noticeable knock-on effect on reviewing quality (one of the core features of peer-review). We plan to systematically investigate review quality in future.

In summary, we made the following observations:

- With the number of submissions to the top conferences having tripled on average in the last three decades, acceptance rates are going down slightly.
- Most of those conferences that are A- or A*-rated today have a long continuity.
- Geographical distribution is not generally relevant; some good conferences take place in the same location; others cycle between continents.
- Good conferences always take place around the same time of the year. This might mean that the community got used to them being important events.
- Some topics have attracted increasing interest recently e.g., database topics thanks to the 'big data' trend. This might be confirmed by further investigations into more recent, *emerging* events in such fields.

In further research, we aim to expand the analysis to other fields of science and to smaller events. Also, it is interesting to assess the impact of digitisation with regard to further scholarly communication means, such as journals (which are more important in fields other than computer science), workshops, funding calls and proposal applications as well as awards. Although large parts of our analysis methodology are already automated, we plan to further optimise the process so that analysis can be almost instantly generated from the OpenResearch data basis.

Acknowledgement. This work has been supported by the European Union for the Horizon 2020 project OpenBudgets.eu (GA no. 645833).

References

1. Ameloot, T.J., et al.: 30 Years of PODS in facts and figures. SIGMOD Rec. **40**(3), 54–60 (2011)
2. Aumüller, D., Rahm, E.: Affiliation analysis of database publications. SIGMOD Rec. **40**(1), 26–31 (2011)
3. Biryukov, M., Dong, C.: Analysis of computer science communities based on DBLP. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) ECDL 2010. LNCS, vol. 6273, pp. 228–235. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15464-5_24
4. Bryl, V., et al.: What's in the proceedings? Combining publisher's and researcher's perspectives. In: SePublica, vol. 1155 (2014). CEUR-WS.org
5. Committee on Global Approaches to Advanced Computing, Board on Global Science and Technology, Policy and Global Affairs Division. Appendix F: Pilot Study of Papers at Top Technical Conferences in Advanced Computing. In: The New Global Ecosystem in Advanced Computing: Implications for U.S. Competitiveness and National Security. National Academies Press (2012)
6. Diederich, J., Balke, W.-T., Thaden, U.: Demonstrating the semantic growbag: automatically creating topic facets for faceteddblp. In: ACM (2007)
7. Osborne, F., Motta, E., Mulholland, P.: Exploring scholarly data with rexplore. In: Alani, H., Kagal, L., Fokoue, A., Groth, P., Biemann, C., Parreira, J.X., Aroyo, L., Noy, N., Welty, C., Janowicz, K. (eds.) ISWC 2013. LNCS, vol. 8218, pp. 460–477. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41335-3_29
8. Vahdati, S., Arndt, N., Auer, S., Lange, C.: OpenResearch: collaborative management of scholarly communication metadata. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) EKAW 2016. LNCS, vol. 10024, pp. 778–793. Springer, Cham (2016). doi:10.1007/978-3-319-49004-5_50
9. Yan, S., Lee, D.: Toward alternative measures for ranking venues: a case of database research community. In: JCDL. ACM (2007)