EVENTS: A Historical Dataset of Top-Prestigious Events in Five Computer Science Communities

Said Fathalla[1,2] and Christoph Lange[1,3]

[1] Smart Data Analytics (SDA), University of Bonn, Germany
{fathalla,langec}@cs.uni-bonn.de
[2] Faculty of Science, University of Alexandria, Egypt
[3] Fraunhofer IAIS, Germany

Abstract Information emanating from scientific events, journal, organizations, institutions as well as scholars become increasingly available online. Therefore, there is a great demand to assess, analyse and organize this huge amount of data produced every day, or even every hour. In this paper, we present a dataset (EVENTS) of scientific events, containing historical data about the publications, submissions, start date, end date, location and homepage for 25 top-prestigious event series (718 editions in total) in five computer science communities. The dataset is publicly available online in three different formats (i.e., CSV, XML, and RDF). It is of primary interest to the steering committees or program chairs of the events to assess the progress of their event over time and compare it to competing events in the same field, and to potential authors looking for events to publish their work. In addition, we shed light on these events by analysing their metadata over the last 50 years. Our transferable analysis is based on exploratory data analysis.

Keywords: Scientific Events Dataset, Scholarly Communication, Digitization, Metadata Analysis

1 Introduction

Information emanating from scientific events, journal, organizations, institutions as well as private scholars become increasingly available online and being demanded electronically by all fields of science as well as industries. This places an excessive pressure on researchers who are working on scholarly communication to assess, analyse and organize this huge amount of data produced every day or even every hour. Digitization has made the preparation of manuscripts as well as the organization of scientific events, from submission to peer review to publication, considerably easier and efficient. Digitization is of crucial importance to all areas of scholarly communication. Therefore, over the last two decades, many organizations and institutes have begun to organize and establish new scientific events. This paper discusses some facts and figures representing 50 years of history of computer science events, where conferences, symposia, and workshops

are of paramount importance and a major means of scholarly communication. Consequently, several online platforms and services for conference management are widely used in this community, including *EasyChair*[4], *Easy Conferences*[5] and *ConfTool*[6]. A key question is: How does digitization affect scholarly communication in computer science? In particular, we address the following questions: *a*) What is the trend of submissions and acceptance rates? *b*) How did the number of publications change? *c*) Has the geographical distribution of events changed across various regions of the world? *d*) Is there an augmentation of publications of a computer science sub-community? *e*) Which events are more geographically diverse than others? In terms of events impact, we address the following questions: *a*) What are the high-impact events of computer science? *b*) Are high-impact events held around the same month each year? *c*) What is the ranking of high-impact events in available ranking services? *d*) Which country has hosted most high-impact computer science events?

We straighten some of these questions out by analysing comprehensive scholarly communication metadata from computer science events in the last 50 years. Our analysis methodology is based on exploratory data analysis, which aims at analysing data to explore the main characteristics, oftentimes with visual methods. Large collections of events metadata are nowadays publicly available on the Web. However, this data are not well-organised, distributed overs digital libraries and events websites, and not integrated. The existence of such data freely available online has motivated us to create the EVENTS dataset for computer science events. In addition, we analysed the key characteristics of scientific events over time, including their CORE[7], Qualis[8] and GII rankings[9], geographic distribution, average acceptance rate, time distribution over the year, submissions and publications. In total, we analysed 25 event series in computer science with regard to these indicators over a period of 50 years. These events provide different tracks such as research, industrial, and demo papers. The number of submissions and publications for each event involves all tracks publications and submissions. We selected five top-prestigious events in five CS communities derived from analysing the topics covered by each event series, then events mapped to the ACM Computing Classification System (CCS)[10]: Information systems (IS), Security and privacy (SEC), Artificial intelligence (AI), Computer systems organization (CSO) and Software and its engineering (SE). The selection of these events is based on several criteria such as CORE ranking, Qualis ranking, GII ranking and Google h5-index. EVENTS dataset closes an important gap in analysing the progress of a CS community in terms of submissions and publications over a long-time period. In addition, it measures some significant aspects of different computer science communities over time, e.g., changes of submission figures

[4] http://easychair.org
[5] http://easyconferences.eu/
[6] http://www.conftool.net/
[7] http://www.core.edu.au/
[8] http://qualis.ic.ufmt.br/
[9] http://valutazione.unibas.it/gii-grin-scie-rating/
[10] https://dl.acm.org/ccs/ccs.cfm

and the distribution of prestigious events in each community over the world. We believe that the EVENTS dataset will have a great impact on scholarly communication community [6], particularly for:

- *event organizers* – to trace their events' progress/impact,
- *authors* – identify prestigious events to submit their research results,
- *proceedings publishers* – to know the impact of the events whose proceedings they are publishing.

This article is organized as follows: Section 2 gives an overview of related work. Section 3 presents the main characteristics of the dataset. Section 4 explains the curation process of creating and evolving the dataset. Section 5 discusses the results of our analysis. Section 6 concludes and outlines our future work.

2 Related Work

A recent review of the literature found that most studies tended to focus on grabbing information about scholarly communication from bibliographic metadata. *DBLP* and *DBWorld*[11], 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. DBLP allows event organizers to upload XML data with bibliographic data for ingestion. The WikiCFP employs crawlers to track high-profile conferences. *SemanticScholar*[12] offers a keyword-based search facility that retrieves metadata about publications and authors on highly relevant hits with the possibility of filtering. It uses artificial intelligence methods in the back-end and retrieves results based on highly relevant hits with the possibility of filtering. *ScholarlyData*[13] provides RDF dumps for scientific events. *Springer LOD* and *ScholarlyData*[14] publish as Linked Open Data metadata of conference related to computer science collected from Springer's traditional publishing process. Ameloot et al. presented a comprehensive analysis of the Principles of Database Systems (PODS) conference series include word clouds of most PODS researchers and newcomers, the connectedness of PODS co-author graph, Histogram of the longest streaks and locations of PODS in the period 2002–2011 [2]. Se and Lee proposed a goodness measure for ranking events [11]. The *goodness* of events is the sum of the goodness of articles in it which is based on bibliographic properties, author information and readers evaluation, whereas our comparisons are based on metrics derived from the characteristics of the events themselves such as event h-index, available events ranking etc. Biryukov and Dong provided an in-depth exploration of Computer Science Communities in terms of publication growth rate, population stability and collaboration trends using DBLP dataset

[11] `http://dblp.uni-trier.de/`, `https://research.cs.wisc.edu/dbworld/`
[12] `https://www.semanticscholar.org`
[13] `http://www.scholarlydata.org/dumps/`
[14] `http://lod.springer.com/`, `http://www.scholarlydata.org/dumps/`

since 1970 [5]. Similarly, Aumüller and Rahm analysed affiliations of database publications using author information from DBLP [3]. Fathalla et al. provided an analysis of 40 computer science conference series in terms of continuity, time and geographic distribution, submissions and publications [7]. Hiemstra et al. analysed the SIGIR community in terms of authors countries, number of papers per year for each country and co-authorship [8]. Barbosa et al. analysed the metadata of 340 full papers published in 14 editions of the Brazilian Symposium on Human Factors in Computing Systems (IHC) [4]. Vasilescu et al. presented a dataset of eleven software engineering conferences, containing historical data about publications and program committees in the period 1994-2012 [10]. Agarwal et al. presented a bibliometric analysis of the metadata of seven ACM conferences (HT, JCDL, DOCENG, WEBSCI, CIKM, WSDM, UMAP) involve different CS fields such as information management, data mining, digital libraries and information retrieval .[1]

3 EVENTS Characteristics

In EVENTS dataset, we are interested in prestigious events in five computer science research communities. To this end, we propose a historical dataset containing information about 25 top-prestigious events in the last five decades, including the event full title, acronym, start date, end date, numbers of submissions, the numbers of accepted papers, city, state, country, event type, field and homepage.

Statistics. EVENTS contains two types of events since 1969: conferences and symposiums. Table 1 provides statistics for the 25 events in the five CS communities (IS, SEC, AI, CSO, and SE). The dataset contains metadata of 653 editions distributed with 15 attributes and available in three formats: CSV, XML, and RDF.

Table 1: EVENTS dataset statistics.

Metrics	**value**	**Metrics**	**value**
series	25	event types	2
editions	718	communities	5
entries	9,460	duration(years)	50
attributes	15	available formats	3

Reusability. Using this dataset, event organizers and chairs will be able to assess their selection process, e.g., keep acceptance rate stable even when the submissions increase, held the event around each time each year, and compare against other competing events. Furthermore, we believe this dataset will assist researchers who want to submit a paper to be able to decide to which events they could submit their work? (e.g., Which events have a high-impact in a particular CS filed? And when a specific conference held each year which helps them to

prepare their research in a specific time plan). Furthermore, EVENTS is used for analysing events impact over time. This analysis could be used to recommend top events in which researchers in different CS communities can publish their work. section 5 presents a part of the analysis that could be performed by using EVENTS dataset.

Extensibility. EVENTS can be extended to meet future requirements by 1) adding more events in each community, 2) adding events in other communities and 3) adding more attributes such as hosting university or organization, sponsors, and event steering committees or program committee chairs.

Availability. EVENTS is published at `https://saidfathalla.github.io/EVENTS-Dataset/EVENTS.html`. It is subject to the Creative Commons Attribution license, as documented at `https://saidfathalla.github.io/EVENTS-Dataset/EVENTS_Licence.html`. The RDF version has been validated using W3C Validation Service[15]. The following is an RDF individual of AAAI (Conference on Artificial Intelligence) conference in 2017:

```
<rdf:Description rdf:about="http://example.org/data/AAAI.csv#row=10">
  <Title>23rd Annual National Conference on Artificial Intelligence</Title>
  <Event_Acronym>AAAI</Event_Acronym>
  <Event_Field>Artificial Intelligence</Event_Field>
  <Event_Homepage>aaai.org/Conferences/AAAI/aaai17.php</Event_Homepage>
  <Event_Series>AAAI2017</Event_Series>
  <Event_Year>2017</Event_Year>
  <Event_Start_date rdf:datatype="&xsd;dateTime">2017-02-04</Event_Start_date>
  <Event_End_date rdf:datatype="&xsd;dateTime">2017-02-09</Event_End_date>
  <Event_City>San Francisco</Event_City>
  <Event_State>California</Event_State>
  <Event_Country>USA</Event_Country>
  <Event_Submission_Rate>24.2\%</Event_Submission_Rate>
  <Event_Submitted_papers rdf:datatype="&xsd;integer">2590</Event_Submitted_papers>
  <Event_Accepted_papers rdf:datatype="&xsd;integer">638</Event_Accepted_papers>
  <Event_Type>Conference</Event_Type>
</rdf:Description>
```

4 Data Curation

While we are collected the data for the dataset, several technical problems have been faced, such as eliminate irrelevant and redundant data, unify event name, complete missing data, and correct incorrect data. Therefore, data curation process is needed. *Data Curation* refers to processes and activities related to the organization, integration, annotation and publication of data collected from various sources [9]. In addition, the EVENTS dataset is maintained over time and remains available for reuse. Algorithm 1 summarizes EVENTS curation process. The curation of EVENTS dataset is an incremental process involving:

- *data completion* – add new available data, i.e. when the number of submission and publications become available online,
- *data update* – changing the name of the even, for instance, the name WWW event has been changed to *TheWeb* in 2018,

[15] `https://www.w3.org/RDF/Validator/`

– *redundancy elimination* – exclude duplicated data,
– *data augmentation* – add new events to the dataset.

4.1 Data Acquisition.

After identifying top events, metadata (raw data) of these events is collected either in structured or unstructured format. The metadata of selected events has been collected from various sources such as IEEE Xplore Digital Library[16], ACM Digital Libraries[17], DBLP[18], OpenResearch.org[19] and events websites. The selection is based on several criteria such as CORE ranking, GII ranking, GII ranking and Google h-index (the largest number h such that h articles published in the last 5 complete years have at least h citations each).

4.2 Data Preprocessing.

The main objective of data Preprocessing phase is to fill in missing data, identify and correct incorrect data, eliminate irrelevant data and resolve inconsistencies. In order to prepare the raw data for analysis, we carried out three preprocessing processes: *data cleansing*, *data integration* and *data transformation*.

Data integration. This process involves combining data from multiple sources into meaningful and valuable information. In addition, this process also involves eliminating redundant data which occur during the integration process.

Data cleansing. This process involves detecting and correcting incorrect or inaccurate records. For instance, we found several websites provide incorrect information about events submissions and accepted papers. We double check this information from events official websites or proceedings published in digital libraries.

Data transformation. This process involves converting cleaned data values from unstructured formats into a structured one. For instance, data collected from events websites as text (i.e. unstructured format) is transformed to CSV (i.e. structured format) and consequently to XML and RDF.

Event name unification. This process involves integrating all editions of an event which had changed their name since beginning under its most recent name because it is important for the researchers to know the recent name rather than the old name. The old name is also important for a researcher who wants to get an overview of the history of an event. Table 2 lists events which changed their names over time. For example, PLDI is the unified name of the *Conference on Programming Language Design and Implementation* which was named *Symposium on Compiler Construction* in the period 1979–1986, *Symposium on Interpreters and Interpretive Techniques* in 1987 and finally it has been changed to its recent name in the period 1989–2018, i.e., for 30 years.

With the completion of these steps, we are now ready to perform our exploratory data analysis.

[16] `http://ieeexplore.ieee.org`
[17] `https://dl.acm.org/`
[18] `http://dblp.uni-trier.de/`
[19] `http://openresearch.org/Main_Page`

Table 2: CS Sub-fields and Top Conferences

Acronym	Full title	Period
NIPS	Neural Information Processing Systems	1987
	Advances in Neural Information Processing Systems	1988–2018
PKDD	Principles of Data Mining and Knowledge Discovery	1997–2002
	Principles and Practice of Knowledge Discovery in Databases	2003–2018
PLDI	Conference on Programming Language Design and Implementation	1989–2018
	Symposium on Interpreters and Interpretive Techniques	1987
	Symposium on Compiler Construction	1979–1986
EuroCrypt	Workshop on Cryptography	1982
	Workshop on the Theory and Application of Cryptographic Techniques	1983–1994
	Conference on the Theory and Applications of Cryptographic Techniques	1995–2018

Algorithm 1 EVENTS curation

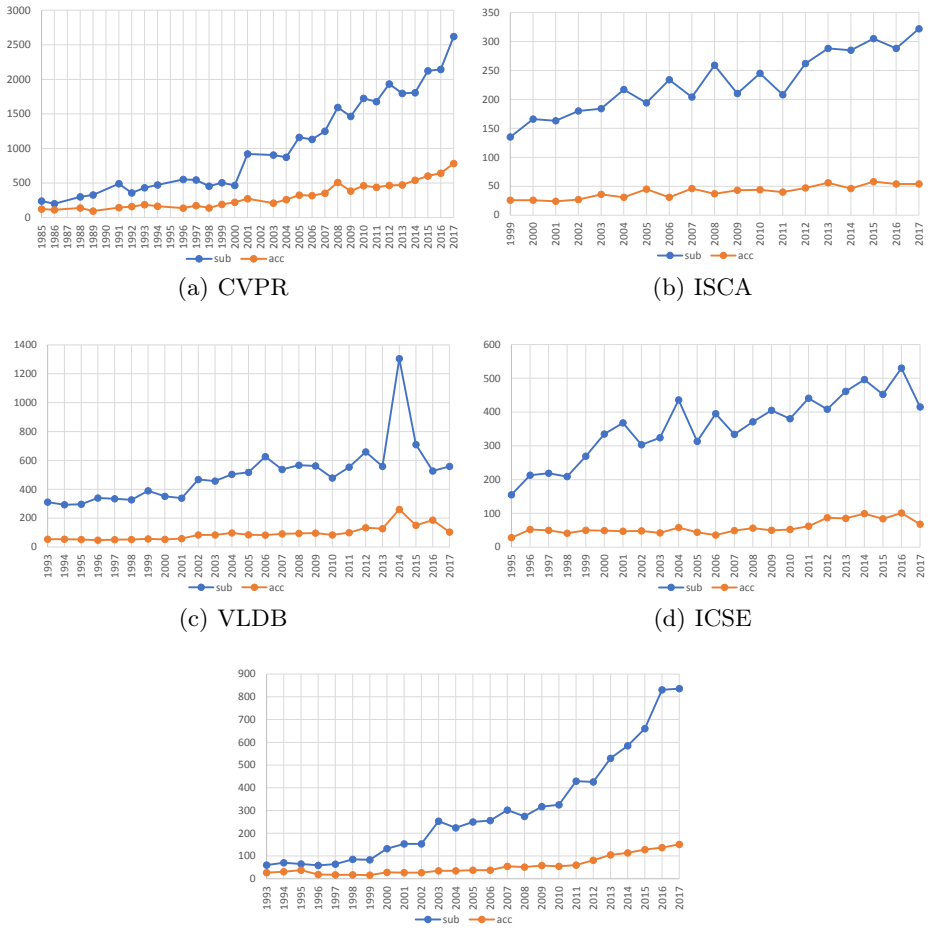
1: Identify prestigious events.
2: **while** there is missing data OR new data available **do**
3: *Data Acquisition*: from digital libraries and events websites.
4: *Data Preprocessing*:
5: - *Data integration*: combine data from multiple sources and eliminate redundant data.
6: - *Data cleansing*: detecting and correcting incorrect or inaccurate records.
7: - *Data Transformation*: convert cleaned data to a structured format.
8: - *Name unification*: unify all editions of an event to its most recent name.
9: *Dataset update*: update incorrect and fill in missing data.
10: **end while**
11: Publish a new version.

5 Data Analysis And Results

Over the last 50 years, we have analysed metadata of CS events in EVENTS dataset including h5-index, average acceptance rate, the number of editions of each event, most country hosts the event, in which month event is held each year (usual month), year of the first edition and proceedings publisher. We analysed events metadata in EVENTS dataset to answer the proposed research questions in section 1.

Submissions and publications. Figure 1 presents accepted and submitted papers measures for the top event in the five CS communities from 1985 to 2017. For CVPR conference, the number of submitted and accepted papers were very close in the first edition in 1985 and the gap between them begin to slightly increase till 2000, then the gap noticeably increased till the end of the time span, i.e. in 2017. In addition, CVPR has the largest number of submitted and the accepted papers and among the others, while ISCA has the smallest. Similarly, CCS the number of submitted papers and the accepted papers were very close in 1993 and the gap between them begin to increase till get the max in 2017. However, the gap between the number of submitted papers and accepted papers in VLDB remains the same during the whole time span except for 2014. Overall, we can see a clear upward trend in the number of submitted and accepted papers during the whole time span. The reason is that digitization makes more

research papers available to the whole community and submitting papers and even contacting papers reviewers become much easier and efficient.



(a) CVPR (b) ISCA

(c) VLDB (d) ICSE

(e) CCS

Figure 1: Variation of the number of submitted and accepted papers of the top event in each CS community since 1985.

Time distribution. We observed that the organizers of the prestigious events always try to keep holding their events around the same month each year, which makes researchers who want to submit their work except when will the event be in the next year. We observed that almost all editions of prestigious events involved in the study have been held in the same month of each year as shown in Figure 4. Namely, PLDI conference (Conference on Programming Lan-

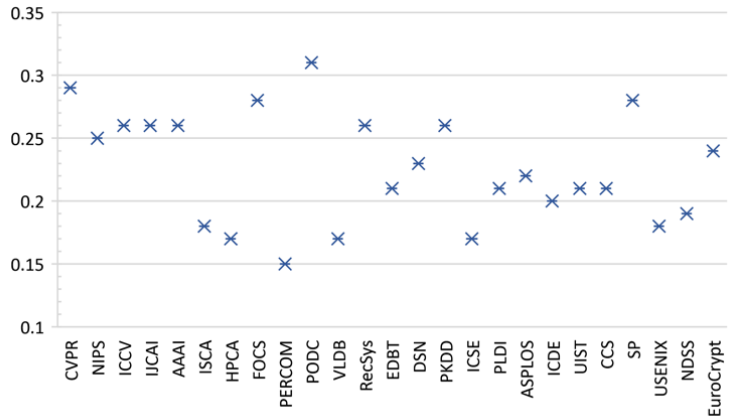


Figure 2: Average acceptance rate of all events

guage Design and Implementation) has been held 30 times (out of 36) in June and SP (IEEE Symposium on Security and Privacy) has been held 31 times (out of 39) since 1989 in May.

Acceptance rate. We analyzed the acceptance rate of the events involved in the study over the last 50 years. As shown in Figure 2, for each event, we compute the average of the acceptance rate of each event since beginning. Interestingly, we found that the average acceptance rate for all events falls into the range 15% to 31% in the time window of 50 years. The largest acceptance rate is the one of PODC (Symposium on Principles of Distributed Computing) of 31%, while PERCOM (International Conference on Pervasive Computing and Communications) has the smallest one of 15%.

H5-index. Figure 3 presents the h5-index of all events along with CORE 2018 ranking. The highest h5-index is the one of CVPR (IEEE Conference on Computer Vision and Pattern Recognition) of 158, while PODC (Symposium on Principles of Distributed Computing) has the smallest one of 25.

Geographical distribution. We analyse the geographical distribution of each event in the dataset. The key question is which countries hosted most events in the considered top events in EVENTS dataset and how frequently a country hosted an event during the last five decades? Figure 5 shows how frequent different countries around the world host a top event in the five CS communities considered in the study. We observed that USA leads by far for hosting the most editions of CVPR, ISCA, VLDB, ICSE, and CCS. Then, Canada comes to the second for hosting ISCA, VLDB, and ICSE.

Event impact. Table 3 shows the Scientometric profile of all events in EVENTS dataset in the five considered CS communities ordered by descending h5-index for each community. The greatest h5-index among all events is 158 for CVPR, while the least one is 31 for both PERCOM and PKDD. The AI community has the largest average h5-index (89.9%); SEC comes second with

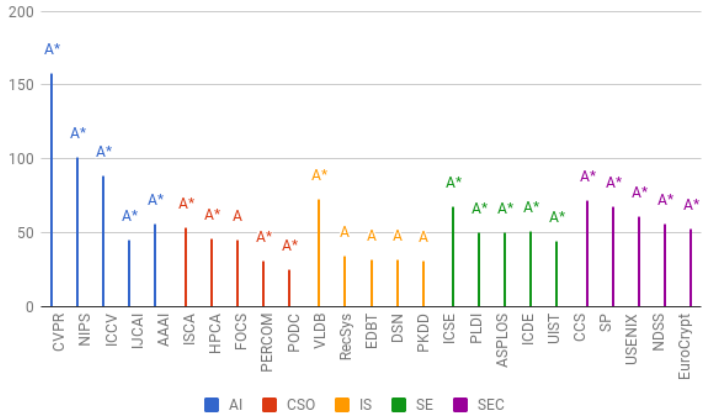


Figure 3: H5-index of all events along with CORE 2018 ranking

62%. Surprisingly, the Qualis ranking of RecSys as *B1*, the h5-index of RecSys is relatively high and is ranked as *A* by CORE and is ranked as *A-* by GII. Regarding publishers, we observed that ACM publishes most of the events proceedings, then IEEE comes next. However, we observed that some events such as NDSS and USENIX publish their proceedings on their own website. In terms of the number of editions, ISCA has the longest history of 45 editions since 1969, while RecSys is the newest one of the 12 editions since 2007. Although RecSys is a relatively new conference, it has a good reputation and it is highly-ranked in CORE, GII, and Qualis.

6 Conclusions and Future work

In this paper, we present a dataset (EVENTS) of conferences and symposiums, containing historical data about 25 top prestigious events in five computer science communities. We present our methodology of creating the dataset starting from identifying prestigious events, data acquisition and Preprocessing to finally publish the dataset. Up to our knowledge, it is the first time to publish a dataset containing metadata of top prestigious events in Information systems, Security and privacy, Artificial intelligence, Computer systems organization and Software and its engineering. In addition, we analyse the metadata of these events over the last 50 years in terms of numbers of submissions, the numbers of accepted papers, geographical distribution, time distribution, h5-index and events ranking in three scientific events ranking services. In summary, we made the following observations:

– There is not much information about events prior 1990 particularly, the number of submissions and accepted papers,

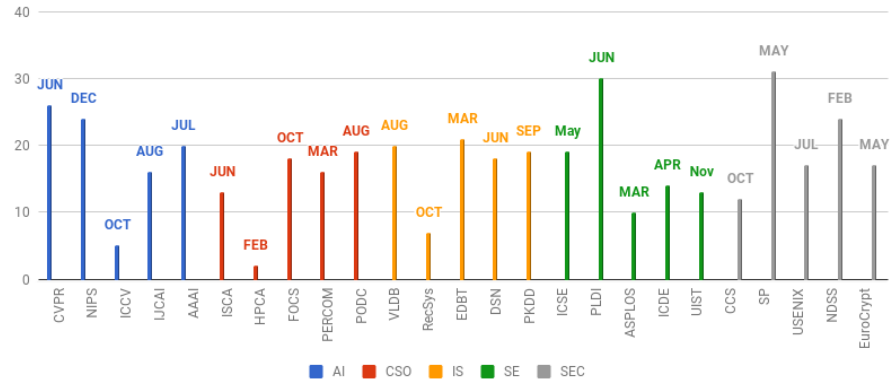
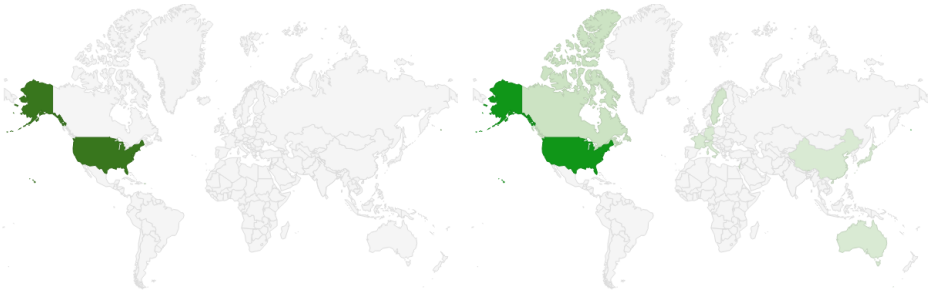


Figure 4: Time distribution of all events in terms of the most frequency of occurrence during the year

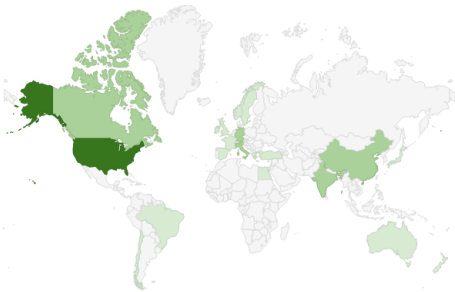
Table 3: Scientometric profile of all events in EVENTS dataset in five CS communities.

Acronym	Comm.	CORE 2018	GII	h5-index	Qualis	Editions 2018	Avg. AR	Most Country	Usual Month	Usual Month Freq.	Since	Publisher
CVPR	AI	A*	A+	158	A1	28	0.33	US	JUN	26	1985	IEEE
NIPS		A*	A++	101	A1	32	0.25	US	DEC	18	1987	NIPS
ICCV		A*	A++	89	A1	17	0.26	Japan	OCT	5	1987	IEEE
IJCAI		A*	A++	45	A1	27	0.26	US	AUG	16	1969	AAAI
AAAI		A*	A++	56	A1	32	0.26	US	JUL	20	1980	AAAI
ISCA	CSO	A*	A++	54	A1	45	0.18	US	JUN	27	1973	IEEE
HPCA		A*	A+	46	A1	24	0.20	US	FEB	17	1995	ACM
FOCS		A	A++	45	A1	30	0.28	US	OCT	25	1989	IEEE
PERCOM		A*	A+	31	A1	16	0.15	US	MAR	16	2003	IEEE
PODC		A*	A+	25	A1	37	0.30	Canada	AUG	19	1982	ACM
VLDB	IS	A*	A++	73	A1	33	0.18	US	AUG	20	1985	VLDB
RecSys		A	A-	34	B1	12	0.26	US	OCT	7	2007	ACM
EDBT		A	A	32	A2	21	0.20	Italy	MAR	21	1988	OP
DSN		A	A	32	A1	19	0.23	US	JUN	18	2000	IEEE
PKDD		A	A	31	A2	22	0.25	France	SEP	19	1997	ACM
ICSE	SE	A*	A++	68	A1	24	0.17	US	MAY	25	1975	ACM
PLDI		A*	A++	50	A1	33	0.21	US	JUN	33	1979	ACM
ASPLOS		A*	A++	50	A1	23	0.22	US	MAR	10	1982	ACM
ICDE		A*	A+	51	A1	34	0.20	US	FEB	14	1984	IEEE
UIST		A*	A+	44	A1	31	0.21	US	OCT	18	1988	ACM
CCS	SEC	A*	A++	72	A1	25	0.22	US	OCT	12	1993	ACM
SP		A*	A++	68	A1	39	0.28	US	MAY	31	1980	IEEE
USENIX		A*	A-	61	A1	27	0.19	US	AUG	17	1990	USENIX
NDSS		A*	A+	56	A1	25	0.20	US	FEB	24	1993	NDSS
EuroCrypt		A*	A++	53	A1	37	0.24	France	MAY	23	1982	Springer

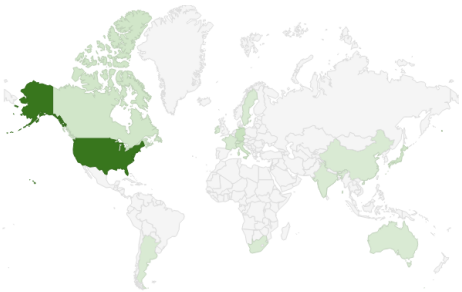


(a) CVPR

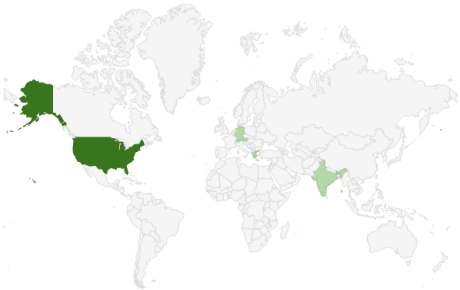
(b) ISCA



(c) VLDB



(d) ICSE



(e) CCS

Figure 5: Geographical distribution of the top event in each CS community since 1973.

– organizers of the prestigious events try to keep the events held around the same month each year,
– There is a clear upward trend in the number of submitted and accepted papers during the whole time span due to the digitization of scholarly communication. However, the digitization of scholarly communication also has negative impacts, most significantly the proliferation of submissions, which significantly increases the reviewing workload,
– Among all countries, USA hosted about 76% of the events in the dataset in the last five decades.

To further our research, we plan to systematically investigate review quality, update EVENTS by adding to meet future requirements by adding more events in each community and more attributes such as hosting university or organization, sponsors, and event steering committees or program committee chairs. We also plan to elaborate on the set of features that could be used to efficiently compare events in the same community, such as acceptance rate, h-index, and organizers reputation, i.e. their h-index and i10-index. Furthermore, we plan to perform more exploratory analysis by applying more metrics such as geographical distribution and publications by continents, event continuity, event progress rate and acceptance rate stability.

References

1. Agarwal, S., Mittal, N., Sureka, A. A glance at seven acm sigweb series of conferences. In: ACM SIGWEB Newsletter(Summer) (2016), p. 5.
2. Ameloot, T. J. et al. 30 Years of PODS in facts and figures. In: SIGMOD Record 40(3) (2011).
3. Aumüller, D., Rahm, E. Affiliation analysis of database publications. In: SIGMOD Record 40(1) (2011).
4. Barbosa, S. D. J., Silveira, M. S., Gasparini, I. What publications metadata tell us about the evolution of a scientific community: the case of the Brazilian human–computer interaction conference series. In: Scientometrics 110(1) (2017), pp. 275–300.
5. Biryukov, M., Dong, C. Analysis of computer science communities based on DBLP. In: *TPDL*. Springer. 2010.
6. Bryl, V. et al. What's in the proceedings? Combining publisher's and researcher's perspectives. In: *SePublica*. CEUR-WS.org 1155. 2014.
7. Fathalla, S. et al. Analysing Scholarly Communication Metadata of Computer Science Events. In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2017, pp. 342–354.
8. Hiemstra, D. et al. SIGIR's 30th anniversary: an analysis of trends in IR research and the topology of its community. In: *ACM SIGIR Forum*. Vol. 41. 2. ACM. 2007, pp. 18–24.
9. Sabharwal, A. Digital curation in the digital humanities: Preserving and promoting archival and special collections. Chandos Publishing, 2015.

10. Vasilescu, B., Serebrenik, A., Mens, T. A historical dataset of software engineering conferences. In: *Proceedings of the 10th Working Conference on Mining Software Repositories*. IEEE Press. 2013, pp. 373–376.
11. Yan, S., Lee, D. Toward alternative measures for ranking venues: a case of database research community. In: *JCDL*. ACM. 2007.