



# Predictive power of conference-related factors on citation rates of conference papers

Danielle H. Lee<sup>1</sup>

Received: 11 July 2018 / Published online: 10 November 2018  
© Akadémiai Kiadó, Budapest, Hungary 2018

## Abstract

This paper aims to determine the factors significantly predicting the future citation rates of conference papers. Whereas a large body of bibliometric studies has investigated the multiple factors predicting future citation rates, the attention has been paid mainly on journal articles. This study analyzes 43,463 papers from 81 conference series in the ‘Information Science’ and ‘Computer Science’ fields and examines the contributions of conference-related factors to the citation rates of the conference papers. More specifically, this paper assesses the following conference related factors as being potentially predictive factors of citation rates: longevity and names of the conference series, the number of presented papers at individual conferences, acceptance rates, the seasons of conferences, the content similarity of the presented papers at a conference, the degree of the authors’ international collaborations and the records of the best paper awards at conferences. The regression results illustrate that all of the factors were significantly predictive to the future citations of the conference papers. The factors that contributed the most to explain the citations of the conference papers include: the degree of the authors’ international collaborations at individual conferences, the records of best paper awards and the acceptance rates of individual conferences.

**Keywords** Citation analysis · Conferences · Information science · Bibliometric analysis

## Introduction

In computer-related fields, conferences are important channels used to disseminate scientists’ research findings, because of the considerable merits. In the fast-moving research environment, the cycle of publishing conference papers (i.e., from paper submission to editorial decision) is shorter than the cycle of publishing journal articles. Hence, conference papers are perceived as timelier (Eckmann et al. 2012; Freyne et al. 2010; Vardi 2010). Journal articles undergoing rigid review processes usually require a sophisticated research design and a much more careful experiment and analysis. On the other hand, conferences

---

✉ Danielle H. Lee  
suleehs@gmail.com

<sup>1</sup> Department of Software, Sangmyung University, 31, Sangmyungdae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do 31066, South Korea

offering various formats of presentations (e.g., presentations, posters, industry papers, discussions, position papers) open up opportunities to present less sophisticated and less mature, but more cutting-edge, ideas. Good conferences also bring scientists together. Hence, the attendees can reap the benefits of their participation by obtaining instant feedback about their studies from colleagues and seeking collaboration opportunities via on-site interactions (Freyne et al. 2010). Moreover, a good quality conference paper is commonly developed into a journal article in a dedicated issue in a journal related to the corresponding conference (Zhang and Jia 2013).

A series of bibliometric studies about computer-related disciplines has quantitatively specified the significance of conference papers. Among the articles indexed by the ACM in 2006, 39% were conference articles. 41% of the articles cited more than ten times were conference papers (Wainer et al. 2011). According to Scopus' statistics, published in 2013, 62.3% of articles in computing science were conference papers, while 32.8% were journal articles. Only 1.9% of the chemical science and 7.3% of the physics articles were from conferences (Scopus 2013). Laender et al. (2008) compared triennial publications produced by 30 computer science graduate programs in three regional areas (North-American, Europe, and Brazil). The analysis indicated that the ratios of conference papers to journal articles published by Brazilian programs, North-American programs and European programs were 2.9, 2.5 and 2.1, respectively (Laender et al. 2008).

One might claim that the large share of conference papers in the computer science discipline does not necessarily represent the significance. The large share may be merely caused by the highly selective review processes of journal publications and the increasing number of conferences. As an extended discussion of the speculations, another series of the existing literature examined the importance of conference papers in several bibliometric perspectives. This included the differences in the citations (Freyne et al. 2010; Vrettas and Sanderson 2015), the patterns of reproducing conference articles for journal publications (Montesi and Owen 2008; Wainer and Valle 2013), the patterns of citing references (Wainer et al. 2011), and the development of a widely adoptable quality evaluation and ranking system for conferences, which could be equivalent to the Impact Factor of journals (Li et al. 2018; Loizides and Koutsakis 2017; Souto et al. 2007). All of the studies were to prove the comparative value of conference papers in comparison with journal articles.

Freyne et al. (2010) empirically verified the relative importance of elite conferences. Papers from elite conferences garnered as many citations as mid-level journals and more citations than lower-level journals. Rahm and Thor (2005) compared two of the most representative conferences (i.e. 'SIGMOD' and 'VLDB') with three major journals ('TODS', 'VLDB Journal' and 'Sigmod Record') on the same topic (i.e., 'Database'). The paper found that the conference papers from the two major conferences earned substantially more citations than the journal articles (Rahm and Thor 2005). Vrettas and Sanderson (2015) also revealed similar results. Papers presented at leading conferences, which are ranked as the highest A\* in the CORE ERA (Computing Research and Education Association of Australasia) ranking, earned significantly more citations than articles published in journals of the same A\* ranking. The study also posited that computer scientists perceived the values of international conferences more highly than researchers from any other disciplines (Vrettas and Sanderson 2015).

Given the fact that the disciplinary boundary between 'computer science' and 'information science' is blurred and the fast pace of innovations in both disciplines coincides with each other, information scientists also accommodate the academic culture of computer scientists. Information scientists also place substantial weight on conference papers (Butler and Visser 2006; Glänzel et al. 2006; Larivière et al. 2012; Wersig 1993).

Despite the significant importance of conference papers in some disciplines, the bibliometric explorations of conference papers have not had much scholarly attention. The bibliometric literature about journal articles still holds supreme. The majority of the existing bibliometric literature has explored journal articles from a variety of perspectives (Shirakawa et al. 2012; Tahamtan et al. 2016; Waltman 2016). In the meanwhile, the studies targeting conference papers have been carried out with just two branches of perspectives. The first branch is, as explained previously, to prove the value of conference papers compared to journal articles (Freyne et al. 2010; Montesi and Owen 2008; Vrettas and Sanderson 2015; Wainer and Valle 2013).

The second branch has focused on a longitudinal analysis of conference proceedings (e.g. history, participants' characteristics, presented papers' characteristics, etc.) about a certain research topic, a country or a digital library, in which the conference proceedings are indexed. Bartneck and Hu (2009) examined the nature of the CHI (the SIGCHI conference on human factors in computing systems) conference proceedings over the past 25 years. The study analyzed the growth pattern of publications at the conference series, authors' affiliated countries and the types of authors' organizations (e.g. universities, institutes and companies) (Bartneck and Hu 2009). Chan et al. (2006) conducted an analysis of the citation patterns and references in the ICIS (the International Conference on Information Systems) conference proceedings from 2000 until 2002.

Some studies also conducted bibliometric analyses of conference proceedings about 'databases (Rahm and Thor 2005; Sakr and Alomari 2012)', 'international business studies (Wuehrer and Smejkal 2013)', 'software engineering (Vasilescu et al. 2014)', and 'recommendation systems (Kim and Chen 2015)'. One study (Barbosa et al. 2017) analyzed a Brazilian conference series to better understand how a certain research community has evolved in a certain country (i.e. Brazil). Some studies analyzing digital libraries where conference proceedings are indexed—e.g., IEEE (Shirakawa et al. 2012), Web of Science (Michels and Fu 2014), DBLP (Song et al. 2014)—determined the coverage of conference proceedings in the digital libraries or topical changes in a certain research area. The lack of bibliometric studies about conference papers and the limited research foci provided a fresh impetus to the direction of the current study. More specifically, this study aims to determine the significant factors predicting the citation numbers of conference papers.

The citation numbers of academic publications are an important indicator of the research impact and some critical criteria to quantitatively evaluate the research (Onodera and Yoshikane 2015; Waltman 2016). Because there is a general consensus among scientists that the citations of articles are not fully explained by the quality of the articles (Onodera and Yoshikane 2015), as an important branch of the bibliometric studies, a large literature investigated the significant factors affecting the changes in citation numbers (Tahamtan et al. 2016). According to Borgman and Furner's (2002) classification, the existing bibliometric studies about citation numbers have explored a number of factors including publication venues, articles, people and more. This study will primarily pay attentions to factors relevant to *conferences*. More specifically, this study will analyze the magnitude effect of various conference-related factors on the citations of conference papers.

How do the names of conference series predict the future citations of conference papers? Can papers presented at conferences with a longer history accrue more citations than papers presented at conferences with a shorter history, due to the cumulative reputation effect? Do the size of the conferences (in terms of the number of presented papers) and the degree of selectivity (in terms of the acceptance rates) predict the future citations of the presented papers? Does the seasonal accessibility of conferences entice good papers and lead to earning more citations? Can the overall content diversity of the presented papers

**Table 1** Factors and features of the conferences considered in this study

Levels	Factors
Conference series	Age of conference series ( <i>age</i> ) Name of conference series ( <i>name</i> ) <sup>a</sup>
Individual conferences	Number of papers presented at a conference ( <i>noPaper</i> ) Acceptance rate of a conference ( <i>acceptRate</i> ) Time of a conference ( <i>time</i> ) Content similarity among the presented papers at a conference ( <i>contentSim</i> ) Degree of international collaborations whose papers were presented at a conference ( <i>internationality</i> )
Individual papers	Record of the best paper award ( <i>bestPaper</i> )

<sup>a</sup>Factors excluded in the regression test

at a conference explain the citations of the papers? Can the international diversity of the participating authors at a conference explain the citations of the papers, as well? Do the award winning best papers, which resulted from the rigorous review process of a program committee, acquire more citations than non-award winning papers?

To answer the questions, 43,463 papers from 81 conference series held between 2009 and 2012, in the ‘Information Science’ and ‘Computer Science’ fields, served as the context of this study. The primary contributions of this study to the existing literature are two-fold. Firstly, this is one of the early attempts to investigate conference relevant characteristics, especially the effects on the citations of conference papers. Secondly, this is one of a few attempts to expand the prior research that limited its bibliometric target on journal articles to conferences articles.

The remainder of this study is organized as follows. In “[Factors under consideration](#)” section introduces detailed descriptions about the factors explored in this study. In “[Data collection](#)” section illustrates how to choose sample conferences and papers and how to collect and process the sampled conference data. In “[Results](#)” section explains the results. The conclusions and discussion are summarized in “[Conclusions and discussion](#)” section.

## Factors under consideration

The primary purpose of this paper is to determine the magnitude effect of various conference related factors on the citation rates of conference papers. The factors considered in this paper are presented in Table 1. The factors are classified into three levels: (1) conference series level, (2) individual conference level, and (3) individual conference paper level. The text in parenthesis at the end of each factor in Table 1 indicates the abbreviated code of the corresponding factor, which is being used throughout this study.

This study included two conference series level factors: (1) the longevity of the conference series, and (2) the names of the conference series. The longevity of the conference series determined how many prior conferences were held. Conference series with a longer history might build up more of a reputation than other conference series with a shorter history. On the other hand, newer conferences with a currently popular research topic could bring more scholarly attention than conferences with longer traditions and classical

research topics (Martins et al. 2010). Hence, this study explores whether conference longevity information explains the citations of the presented papers or not.

In relation to the names of the conference series, 81 *conferences series* were sampled. This included 297 *individual conferences* held from 2009 until 2012 and the papers presented at each individual conference. By using the papers from the same conference series in multiple years, this study examines whether the prior reputation of a conference series could have an increasing effect on the citations of the presented papers or not. A few studies demonstrated that the reputations of publication venues positively affect the citations of the presented articles. In the Internet studies discipline, Peng and Zhu (2012) concluded that, what matters the most in predicting the future citation rates of published articles is where you publish your articles; hence, they proved the halo effect of the published journals (Peng and Zhu 2012). Another study (Vanclay 2013), substantiated in environmental science, also made the same conclusion. However, both studies targeted journal articles, and the reputations were chiefly represented by the JCR Impact Factors, unlike the current study. This study will assess the reputation effect of conference series' names. Due to the excessive categorical values of the factor (i.e., 81 conference names) and to examine the reputation effect in details, the factor was excluded in the below regression test. Instead, a separate sub-section dedicated to it (i.e. "[Citation rates of conference papers and the reputation effects of conference names](#)") will assess the reputation effect of the conference names.

Five factors about individual conferences were then considered: (1) the number of presented papers, (2) acceptance rate, (3) time of conferences, (4) content similarity of presented papers and (5) international collaborations of authors. The number of presented papers is the measure used to estimate the size of each individual conference and to examine how the size of the individual conference could predict the future citations of the conference papers. There could be a few measures for the size of the conferences: the number of paper submissions, enrollments and attendees, the number of sponsors, and the conference profits. Good conferences usually attract a large number of good manuscripts, attendances and sponsors (Martins et al. 2010). Because these types of information are not readily available for every individual conference, the number of presented papers was chosen as a commonly available factor for every individual conference.

The acceptance rates of individual conferences were considered to measure the effect of the selectivity degree of individual conferences on the citations of conference papers. There is an ongoing debate about the influence of conferences' acceptance rates on conference papers. Freyne et al. (Freyne et al. 2010) correlated the acceptance/rejection rates of 15 conferences with the Google citation rates. They found no significant correlation between the rejection rates and the multiple years' citation rates of conferences (Freyne et al. 2010). On the other hand, in a study based on the ACM digital library data, Chen and Konstan (2010) found significant correlations between the acceptance rates of conferences and the citation rates of papers presented at the conferences. This study will make its own conclusion using about 43,400 sample conference papers.

The *time of a conference* represents one of the 12 months when individual conferences were held. The time of a conference tested whether the seasons of conferences affects the citation rates of conferences papers or not. Once a conference paper is accepted, at least one of the authors must attend the conference to present their work. When a conference is held during the middle of a school semester, faculty and student researchers may feel burdensome to cancel or skip classes. Therefore, if there is another alternative option (i.e. another relevant conference) that is held during the vacation season, the researchers may select that alternative.

The next two factors—content similarity among papers at a conference and the internationality of the authorship at a conference—are to examine whether the diversity of conferences affects the citations of conference papers or not. The topical scope of conferences varies significantly. While some conferences exhibit a wider or interdisciplinary scope of the research topics, other conferences concentrate on a narrower, or a distinct, discipline scope of topics. However, it is unknown whether the overall topical scope of a conference explains the citations of the presented papers or not. This study carries out an analysis to illustrate the predictive power of the topical diversity of a conference, which is represented by the content similarity values of all of the presented papers at a conference.

Drawing on the wealth of the bibliometric literature about international co-authorship (Han et al. 2014; Hoekman et al. 2010; Nomaler et al. 2013), international collaborations were found to increase dramatically since the 1990s. They are inclined to produce articles having higher scholarly impacts than domestic collaborations. However, the attention of this research direction has focused on journal articles. As far as the international co-authorship of conference papers is concerned, few researchers attempted to address the influence on the citations of papers. One exception is Elshawi and Sakr (2016), who investigated the co-authorship patterns of authors at three elite conferences about computer algorithms (i.e., SODA, ICALP, and SOCG). The study did not touch upon the patterns of international collaborations and further the impact of international collaborations on the citations of conference papers. This study calculated the ‘internationality’ of a given conference by counting the number of papers written by authors from more than one country to the whole papers presented at the conference. Given an assumption that international collaborations produce more impactful papers, this study assumes that conferences presenting more internationally co-authored papers would have more citations than other conferences having fewer internationally co-authored papers. Consequently, in “Results” section, the ratio of internationally co-authored papers for each conference was regressed on the citations of the presented papers.

The last investigated factor, and a sole factor considered in the level of individual conference papers, is the *records of the best paper awards*. This was included to reflect experts’ assessment of conference papers. No coherent conclusions have yet to be made in relation to if the best paper winners were destined to earn significantly more citations than non-awarded papers. For the CHI (the SIGCHI conference on human factors in computing systems) conference series, Bartneck and Hu (2009) compared the number of citations for the best paper winners and the nominees against the citations earned by not awarded or not nominated random sample papers. They failed to find any statistically significant citation differences among the three types of papers (the best paper winner papers, nominated papers and non-awarded papers). However, Wainer et al. (2015) presented the opposite result. In the study based on multiple-year records of 29 computer science conferences, the authors demonstrated that the best paper winners earned significantly more citations than the non-awarded papers. Moreover, among the sample papers, half of the top 10% and more than 60% of the top 20% of the most cited papers were recipients of the best paper awards (Wainer et al. 2015). In this study, multiple-year records for 81 conference series were used to re-examine whether the records of the best paper awards would impact the future citations of the conference papers or not.

## Data collection

### Selection of target conference papers

To collect sample conference papers, first the target conference series were chosen. This study selected 81 international conference series where faculty members from 13 schools and departments of ‘information science’ in the United States<sup>1</sup> had presented their research findings. Even though the collection of target conferences started from the ‘information science’ major, the topics of the conferences are not strictly limited to ‘information science’. The topics were also highly associated with ‘computer science’. The topics ranged from information retrieval, social computing, human computer interactions, information and knowledge management, data mining, and artificial intelligence to network management, network security, cloud computing and more. The target conferences were highly ranked by the CORE ERA ranking. Among the 81 conference series, 15 conference series were ranked A\*, 24 conference series were ranked A, 17 were ranked B and 7 were ranked C, respectively. The remaining 18 conference series were not listed in the ranking.

For the 81 conference series, individual conferences held between 2009 and 2012 were chosen for this study. The year span was decided to sample publications being published relatively recently, but accruing sufficient citations. Lisée and colleagues suggested 4.1 years as a half-life for computer science conference papers (Lisée et al. 2008). This means that it usually takes 4.1 years for computer science conference papers to earn 50% of the entire number of citations that they will ever receive. Given the suggestion that the disciplinary boundary between ‘computer science’ and ‘information science’ is blurred, this study concluded that 4 years of citations (from 2012 till the time of data collection, July 2016) were reasonable. 43,463 conference articles were published at the 81 conference series between 2009 and 2012; these were target articles of this study. The target conference series and the numbers of sample papers in each year’s individual conference are illustrated in Table 2.

### Data sources of citation rates and various factors

Scopus was the main source used for collecting the citation numbers of the current study’s sample papers. Notwithstanding the popularity of WoS (Web of Science) citation metrics, as a standard of faculty promotion, tenure evaluation, research funding and evaluation, several studies (Bornmann et al. 2012; Franceschet 2010; Meho and Yang 2007; Zahedi et al. 2014) demonstrated the biased citation coverage. The main coverage of the WoS is on journal publications (i.e., approximately 8700 journals). The coverage of conference proceedings is rather small. As another drawback, the coverage of the WoS metrics varies depending on the research fields (Zahedi et al. 2014). In their study based on 25 Library and Information Science faculty’s publications, Meho and Yang (2007) substantiated the importance of other citation metrics in the ‘Information Science’ discipline. The study

<sup>1</sup> Pennsylvania State University; Syracuse University; Temple University; University of California—Irvine; University of Florida; University of Illinois—Urbana—Champaign; University of Indiana; University of Maryland—College Park; University of North Carolina; University of Pennsylvania; University of Pittsburgh; University of Southern California; University of Washington (in alphabetic order).

**Table 2** The list of target conferences (in the right column, the numbers in parentheses are the conference years)

Conf	Full conference name	No. of presented papers (conference years)
ACNS	Conference on Applied Cryptography and Network Security	32 (2009), 32 (2010), 31 (2011), 33 (2012)
AIED	Conference on Artificial Intelligence in Education	129 (2011)
AMIA	The American Medical Informatics Association Annual Symposium	142 (2009), 188 (2010), 180 (2011), 159 (2012)
ANCS	Symposium on Architectures for Networking and Communications Systems	42 (2009), 37 (2010), 37 (2011), 33 (2012)
ASIS&T	Association for Information Science and Technology Annual Meeting	189 (2009), 181 (2010), 170 (2011), 201 (2012)
CCGRID	IEEE/ACM Symposium on Cluster, Cloud and Grid Computing	85 (2009), 114 (2010), 72 (2011), 137 (2012)
CCNC	IEEE Consumer Communications and Networking Conference	336 (2009), 286 (2010), 285 (2011), 209 (2012)
CCS	ACM Conference on Computer and Communications Security	58 (2009), 98 (2010), 62 (2011), 111 (2012)
CHI	Annual CHI Conference on Human Factors in Computing Systems	277 (2009), 302 (2010), 409 (2011), 369 (2012)
CIKM	ACM Conference on Information and Knowledge Management	332 (2009), 316 (2010), 422 (2011), 451 (2012)
CLOUD	IEEE International Conference on Cloud Computing	32 (2009), 77 (2010), 109 (2011), 139 (2012)
CloudCom	IEEE Conference on Cloud Computing, Technology and Science	75 (2009), 105 (2010), 113 (2011), 134 (2012)
Cluster	IEEE International Conference on Cluster Computing	85 (2009), 35 (2010), 77 (2011), 78 (2012)
CSCW	ACM Conference on Computer Supported Cooperative Work	58 (2010), 112 (2011), 266 (2012)
DC	Conference on Dublin Core and Metadata Applications	14 (2009), 20 (2010), 27 (2011), 14 (2012)
ECIR	European Conference on Information Retrieval	91 (2009), 72 (2010), 72 (2012)
ECRTS	Euromicro Conference on Real-Time Systems	26 (2009), 27 (2010), 24 (2011), 34 (2012)
EC-TEL	European Conference on Technology Enhanced Learning	88 (2009), 68 (2010), 52 (2011)
EDM	International Conference on Educational Data Mining	33 (2009), 52 (2010), 67 (2011)
eScience	IEEE International Conference on e-Science	51 (2009), 44 (2010), 54 (2011), 79 (2012)
GlobeCom	IEEE Global Telecommunications Conference	1099 (2009), 1283 (2010), 1056 (2011), 957 (2012)
GROUP	ACM International Conference on Supporting Group Work	54 (2009), 66 (2010), 53 (2012)
HICSS	Annual Hawaii International Conference on System Sciences	505 (2009), 484 (2010), 492 (2011), 586 (2012)
HIPC	International Conference on High Performance Computing	49 (2009), 40 (2010), 40 (2011), 55 (2012)
HPSR	IEEE Conference on High Performance Switching and Routing	28 (2009), 36 (2010), 44 (2011), 39 (2012)



**Table 2** (continued)

Conf	Full conference name	No. of presented papers (conference years)
Hypertext	ACM Conference on Hypertext and Hypermedia	73 (2009), 56 (2010), 35 (2011), 46 (2012)
i-Semantics	International Conference on Semantic Systems	83 (2009), 46 (2010), 38 (2011), 43 (2012)
i-Society	International Conference on Information Society	107 (2010), 89 (2011), 97 (2012)
ICC	IEEE International Conference on Communications	1005 (2009), 991 (2010), 1069 (2011), 1333 (2012)
ICCCN	Conference on Computer Communications and Networks	183 (2009), 127 (2010), 235 (2011), 140 (2012)
ICDIM	International Conference on Digital Information Management	40 (2009), 89 (2010), 59 (2011), 65 (2012)
ICDM	IEEE International Conference on Data Mining	139 (2009), 158 (2010), 148 (2011), 151 (2012)
ICIS	International Conference on Information Systems	298 (2011), 308 (2012)
ICTITST	Conference for Internet Technology and Secured Transactions	147 (2009), 83 (2010), 140 (2011), 139 (2012)
ICNP	IEEE International Conference on Network Protocols	36 (2009), 31 (2010), 56 (2011), 53 (2012)
iConference	iConference	167 (2011), 148 (2012)
ICPADS	IEEE International Conference on Parallel and Distributed Systems	138 (2009), 121 (2010), 153 (2011), 146 (2012)
ICPP	International Conference on Parallel Processing	71 (2009), 72 (2010), 81 (2011), 53 (2012)
ICSE	ACM/IEEE International Conference on Software Engineering	70 (2009), 212 (2010), 212 (2011), 230 (2012)
ICSSP	International Conference on Software and Systems Process	34 (2011), 35 (2012)
ICWSM	International AAAI Conference on Weblogs and Social Media	72 (2010), 111 (2012)
IMC	ACM SIGCOMM Internet Measurement Conference	41 (2009), 47 (2010), 42 (2011), 45 (2012)
Infocom	Conference on Computer Communications	378 (2009), 378 (2010), 416 (2011), 397 (2012)
IRI	IEEE Conference on Information Reuse and Integration	87 (2009), 79 (2010), 107 (2011), 106 (2012)
ISCC	IEEE Symposium on Computers and Communications	193 (2009), 211 (2010), 192 (2011), 155 (2012)
ISMIR	International Society for Music Information Retrieval Conference	124 (2009), 110 (2010), 133 (2011), 100 (2012)
ISSI	International Conference on Scientometrics and Informetrics	156 (2009)
ISWC	International Semantic Web Conference	66 (2009), 77 (2010), 82 (2011), 81 (2012)
ITBAM	Conference on Information Technology in Bio- and Medical Informatics	21 (2010), 16 (2011), 15 (2012)
ITS	International Conference on Intelligent Tutoring Systems	158 (2010), 137 (2012)

Table 2 (continued)

Conf	Full conference name	No. of presented papers (conference years)
IUI	ACM International Conference on Intelligent User Interfaces	73 (2009), 80 (2010), 88 (2011), 84 (2012)
iV	International Conference on Information Visualisation	101 (2009), 95 (2010), 100 (2011), 107 (2012)
JCDL	ACM/IEEE Annual Joint Conference on Digital Libraries	112 (2009), 66 (2010), 102 (2011), 99 (2012)
KDD	ACM SIGKDD Conference on Knowledge Discovery and Data Mining	140 (2009), 122 (2010), 163 (2011), 184 (2012)
LCN	Annual IEEE Conference on Local Computer Networks	168 (2009), 177 (2010), 163 (2011), 108 (2012)
MASCOTS	IEEE Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems	74 (2009), 62 (2010), 58 (2011), 61 (2012)
MobiCom	Annual ACM Conference on Mobile Computing and Networking	30 (2009), 34 (2010), 29 (2011), 53 (2012)
MobiHoc	ACM Symposium on Mobile Ad Hoc Networking and Computing	50 (2009), 26 (2010), 25 (2011), 35 (2012)
MSWiM	ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems	54 (2009), 51 (2010), 53 (2011), 50 (2012)
P2P	IEEE International Conference on Peer-to-Peer Computing	47 (2009), 45 (2010), 49 (2011), 32 (2012)
RecSys	ACM Conference on Recommender Systems	82 (2009), 74 (2010), 72 (2011), 76 (2012)
RTAS	Real-Time and Embedded Technology and Applications Symposium	32 (2009), 33 (2010), 29 (2011), 30 (2012)
RTSS	IEEE Real-Time Systems Symposium	44 (2009), 36 (2010), 33 (2011), 35 (2012)
SAINT	IEEE/IPSJ Annual Symposium on Applications and the Internet	66 (2009), 88 (2010), 99 (2011), 68 (2012)
SBAC-PAD	Symposium on Computer Architecture and High Performance Computing	21 (2009), 30 (2010), 26 (2011), 36 (2012)
SBP-BRIMS	Conference on Social Computing, Behavioral-Cultural Modeling and Prediction	48 (2010), 46 (2011), 43 (2012)
SECON	Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks	81 (2009), 88 (2010), 73 (2011), 93 (2012)
SenSys	ACM Conference on Embedded Networked Sensor Systems	80 (2009), 71 (2010), 75 (2011), 58 (2012)
SIGIR	ACM Conference on Research and Development in Information Retrieval	207 (2009), 198 (2010), 221 (2011), 210 (2012)
SIGMOD	ACM International Conference on Management of Data	119 (2009), 137 (2010), 139 (2011), 97 (2012)
SITIS	Conference on Signal Image Technology and Internet Based Systems	74 (2009), 52 (2010), 68 (2011), 137 (2012)
SocInfo	International Conference on Social Informatics	17 (2010), 38 (2011), 42 (2012)
TPDL	International Conference on Theory and Practice of Digital Libraries	64 (2011), 62 (2012)
UIST	Annual ACM Symposium on User Interface Software and Technology	33 (2009), 38 (2010), 67 (2011), 62 (2012)
UMAP	Conference on User Modeling, Adaptation, and Personalization	53 (2009), 43 (2010), 90 (2011), 43 (2012)

**Table 2** (continued)

Conf	Full conference name	No. of presented papers (conference years)
VL/HCC	IEEE Symposium on Visual Languages and Human-Centric Computing	161 (2009), 191 (2010), 169 (2011), 185 (2012)
VLDB	International Conference on Very Large Data Bases	46 (2009), 47 (2010), 53 (2011), 60 (2012)
WCNC	IEEE Wireless Communications and Networking Conference	539 (2009), 533 (2010), 374 (2011), 604 (2012)
WebSci	International Web Science Conference	28 (2011), 46 (2012)
WoWMoM	IEEE Symposium on World of Wireless, Mobile and Multimedia Networks	99 (2009), 110 (2010), 134 (2011), 125 (2012)
WWW	International World Wide Web Conference	198 (2009), 241 (2010), 242 (2011), 380 (2012)

results revealed that Google Scholar and Scopus covered four times and two times more conference papers than the WoS, respectively (Meho and Yang 2007).

The debate about the comparative advantage between Google Scholar and Scopus is still open. According to the results of a recent study (Harzing and Alakangas 2016) in the Engineering and Applied Science field, the coverage of Google Scholar is distinctly wider than the coverage of Scopus (for instance, 42% larger in the number of citations and 25% higher in h-index). However, it is also known that Google Scholar still has some room to improve (e.g., duplicated citations, inconsistency, errors in author names and published journal/conference names) (Franceschet 2010). Based on the longitudinal and cross-disciplinary comparison, Harzing and Alakangas (2016) showed that Scopus has a sufficiently stable coverage of citations. Hence, one can safely assume that Scopus is a reliable source for collecting the citations of conference papers. Moreover, most of the sample conference proceedings were published and indexed in the Scopus database. The Open API of Scopus<sup>2</sup> enabled to extract the citations of the current study's target papers on a mass scale.

Among the various conference-related factors of the sample publications, the names of the conference series, the number of papers presented at individual conferences, the content-related textual metadata of the target papers (i.e. titles, authors' keywords, and abstracts) and the author-related metadata of the target papers (i.e. authors' names, their affiliations, and the countries) were automatically extracted using Scopus's Open API. The remaining information (e.g., longevity of conference series, acceptance rates of individual conferences, months of individual conferences and the records of the best paper awards) was collected manually. The conference websites, the prefaces of the conference proceedings, authors' personal websites and publishers' information (e.g. IEEE conference information pages) about conference series were referred to.

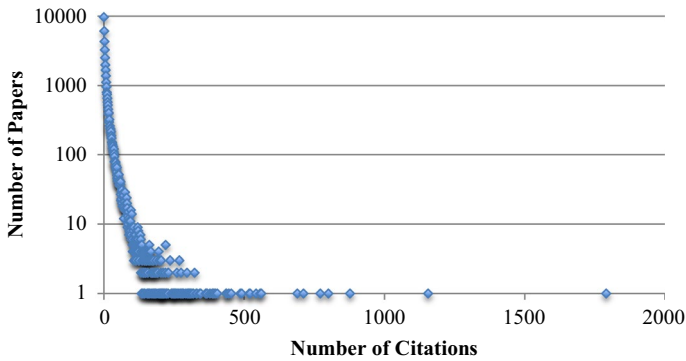
To determine the content similarity values among all of the papers presented at a conference (i.e., '*contentSim*'), this study took into account three types of textual metadata of the sample conference papers—titles, authors' keywords, and abstracts. All of the textual metadata of each sample conference paper was aggregated into one individual bag without any weight. After applying stemmer and stop word removal to each bag, the term frequency-inverse document frequency (TF-IDF) was calculated for each bag to weigh the keywords in the bag by their importance. The weighted bags of keywords allowed for the measuring of the similarity/dissimilarity among papers by computing the cosine similarity values. Lastly, the calculated similarity values of all presented papers at a conference were averaged out to illustrate the overall content similarity values of one individual conference. For detailed information about this content similarity calculation, refer to Lee and Brusilovsky (2017). Finally, to calculate the international collaborations of the participating authors at a conference (i.e. '*internationality*'), the ratio of internationally co-authored papers to the total number of presented papers for each individual conference was counted.

## Results

This paper aims to assess the predictive power of various conference-related factors on the citation rates of conference papers. To that end, the citation rates were identified as the dependent variable of this study. This study's target papers were published from 2009

---

<sup>2</sup> <https://dev.elsevier.com/>.



**Fig. 1** Distribution of papers by the citation rates

through 2012; earlier published papers have a longer duration to garner more citations than more recent papers. To neutralize the time effects and normalize the citation rates, as a dependent variable of this study’s statistical analysis, a 5-year citation window was applied to the citation rates of the conference papers. For instance, for a sample paper published in 2009, the citations to this paper from 2009 until the fifth-year 2013 were counted; for another sample paper published in 2012, the citations to this paper in 2012–2016 were counted. The various conference-related factors introduced in the “[Factors under consideration](#)” section were identified as the independent variables. The variables were used in a regression model.

**Citation rates of conference papers and the reputation effects of conference names**

Prior to the analysis based on a regression model, the general distributions of the citation rates were examined. Again, since the sampled papers were published from 2009 to 2012, the numbers of citations up to 5 years since the publications were taken into account. Figure 1 illustrates how this study’s sample papers are distributed by the citation rates; the majority of papers earned a small number of citations. Specifically, 24.2% of the papers ( $n = 10,536$ ) were never cited during the 5 years; 78.7% of the sampled papers ( $n = 34,203$ ) including the zero-cited papers of 24.2% were cited fewer than ten times in approximately 5 years since they were published. The total number of citations earned by the lower 78.7% of this study’s sample papers occupied 23.3% of the citation rates of all sample papers. On the other hand, the citation rates of the top 5% most cited papers ( $n = 2296$ ) occupied 42.9% of the entire citation rates of all sample papers. The average number of citations earned by the top 5% most cited papers was 63.5 ( $\sigma = 58.3$ ).

This study examined the citation rates affected by the conference series’ names to examine whether the reputation effect of a conference series exists or not. The first analysis was to answer the following question: *How do we measure the reputation effects of a conference series?* The most primitive way to answer this question is to count the raw citation numbers accruing to each conference series from 2009 until 2012. However, this study’s sample has a wide range of conference sizes: 14 to 1333 papers each year. Naturally, the conference series that has more papers tends to garner larger sums of citations. However, when taking into account the citation rates earned by the individual papers, the big scale conferences do not necessarily entail a large number of impactful publications. For

instance, some conference series presented more than 1000 papers every year, but more than one-fifth of the presented papers were never cited in 5 years. In addition, fewer than 10% of the presented papers were cited 10 times<sup>3</sup> or more.

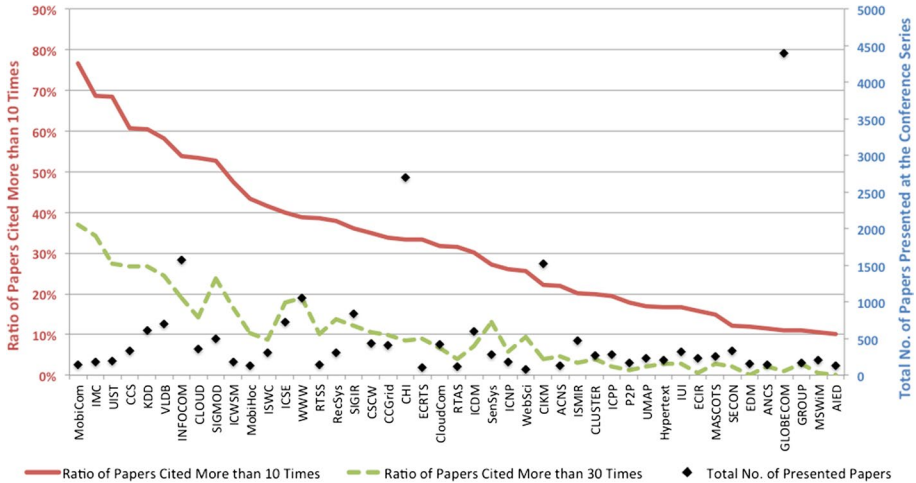
Various formats and lengths of presentations (e.g. presentation, poster, industry paper, discussions, etc.) in each conference series may cause the unexpectedly low citations of the presentations. Shorter publications (e.g., posters, industry papers, discussions) tend to carry less mature, but more innovative, content. These types of shorter publications are appropriate for on-site discussions, not citations. Moreover, specific presentation types—i.e., industry papers, symposium abstracts or position papers—intend to share knowledge beyond academia; they target not only academia, but also industry and governments. When a conference series offers a lot of posters, industry sessions, and discussion sessions, the shorter papers occupy a considerable proportion of the conference publications, which leads to fewer citations overall.

Because the format of each presentation offered by every conference series is unavailable, this study examined how the average number of pages of papers presented at each conference series is correlated with the average citations of each conference series. The analysis revealed a statistically positive correlation ( $r=.30$ ,  $p<.001$ ). On average, the longer papers are cited more frequently than the shorter papers. The results bear a resemblance to the existing studies, postulating that longer papers have a better chance to get cited than shorter papers, since longer papers convey an adequate depth of content as references (Tahamtan et al. 2016; Vrettas and Sanderson 2015). The correlations between the average number of pages and the ratio of highly cited papers were also computed; the average number of pages is positively correlated with the ratio of papers cited more than 10 times ( $r=.34$ ,  $p<.001$ ) and the ratio of papers cited more than 30 times ( $r=.33$ ,  $p<.001$ ). That is, the conference series with more shorter papers earned fewer citations overall and vice versa.

The results illustrate that, because of the highly skewed (i.e., lots of zero citations) distribution of citations, the size differences of the conference series and the diverse formats of the presentations, it is inappropriate to consider the raw numbers of the citations, the average number of citations or the sum of the raw citation rates per conference series, as a measure to assess the overall reputation effect of the conference series. Rather, because the major interest of this study was on how the *reputation effects* of a conference series bring more impactful papers, focus was placed on the ratio of highly cited papers per conference series.

Figure 2 shows the conference series where at least 10% of the presented papers were cited more than 10 times. Specifically the graph depicts two types of ratios: the ratio of papers cited more than 10 times and the ratio of papers cited more than 30 times. The Fig. 2 indicates that some of the elite conferences—such as ‘KDD’, ‘INFOCOM’, ‘VLDB’, ‘ICSE’, ‘WWW’, ‘SIGIR’, ‘CIKM’ and ‘CHI’—produced a large number of papers and a good proportion of presented papers with a high impact. Moreover, another group of notable conferences—for instance, ‘MobiCom’, ‘IMC’, ‘UIST’ and ‘CCS’—were rather small in size, but produced a large proportion of notable papers. Over 60% of the presented papers were cited more than 10 times, while about 30% were cited 30 times or more after 5 years of publication. The Fig. 2 also illustrates that the sizes of the conferences may not

<sup>3</sup> In the line of work by Wainer et al. (2011), which substantiated the citation rate changes in the same discipline with this study, the citation number ‘10’ was chosen as the threshold to determine if a given paper is highly cited or not.



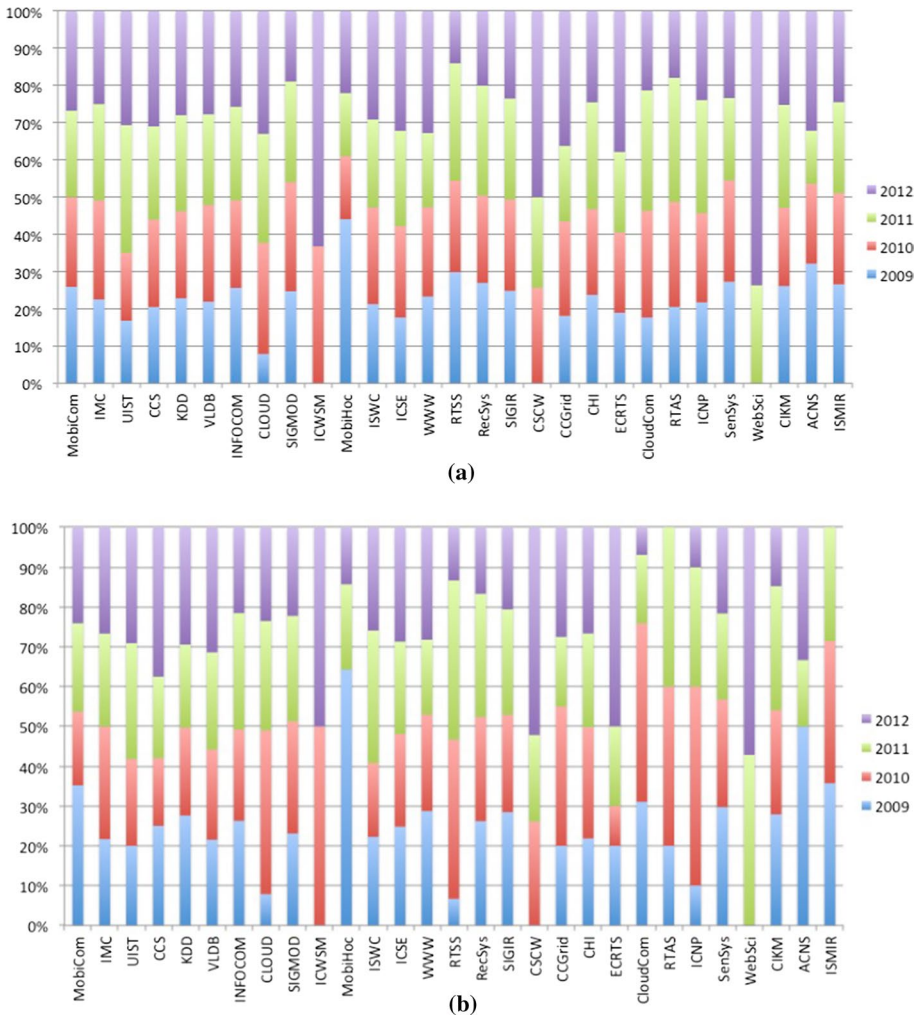
**Fig. 2** Conference series sorted by the ratio of highly cited papers (conference series of which the ratio of the papers cited more than 10 times is less than 10% are omitted.)

be quite relevant enough to the reputation effects of the conference series to garner more citations for the presentations. To ensure this pattern statistically, in the next section, the effect of the conference sizes on the future citations of the presented papers will be tested using a regression test.

The Fig. 2 reveals that some elite conferences attract readers’ substantial attention to cite the publications of the elite conferences as references. However, *whether the readers’ attention on a conference series, if any, is ephemeral for a certain year’s conference or lasts over the years* is still in question. To answer this question, for the top 30 conference series, with respect to the ratio of papers cited more than 10 times, the patterns of the yearly ratio of highly cited papers were described in Fig. 3. Except for the conference series having 2 or 3 years’ records in this study’s sample—such as ‘ICWSM’, ‘CSCW’, ‘WebSci’, the yearly ratios of highly cited papers of the top conferences were quite evenly distributed. Therefore, the top 30 conferences did steadily produce highly cited papers. Hence, these top conferences were sustaining their reputation as elite conferences for years.

Next, testing was conducted on *how the changes of the conference sizes are correlated with the year-over-year growth of the citation rates*. It is likely that reputable conferences frequently receive good submissions, and the program committees try to include as many good papers as possible in their proceedings. This results in the size of the conferences (i.e., the number of presented papers) increasing. In this way, good contributions in the proceedings have a higher chance of being cited. Therefore, the yearly growth of the conference proceedings may be a positive sign indicating the improved reputation of the conference series, and furthermore, the increased chances of the presentations being cited. This study tested this presumption.

The relative changes in conference sizes (i.e., increased/decreased number of presented papers at an individual conference) in a given year were compared with the previous year’s conference sizes. The equation  $(x - y)/y$  was used, where  $x$  is the number of papers in a given year and  $y$  is the number of papers in the previous year. The relative changes in citation rates (i.e., increased/decreased number of the entire citation rates at an individual conference) were computed in the same way. The correlation between the changes in the

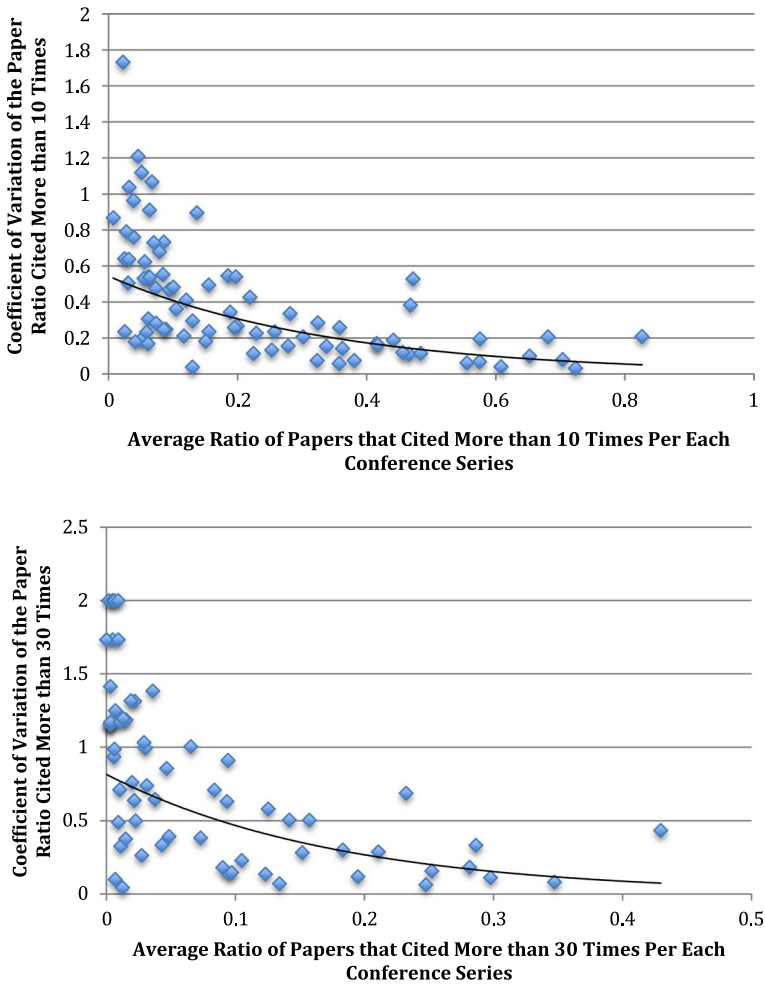


**Fig. 3** Patterns of the yearly ratio of highly cited papers for top 30 conference series. **a** Ratio of papers cited more than 10 times and **b** ratio of papers cited more than 30 times

conference sizes and the changes in the citation rates were computed. This study failed to find any significant correlation between the changes in the conference sizes and the changes in the citation rates ( $r = -.09, p = .21$ ). That is, the changes in the conference sizes does not immediately increase or decrease the impact made by the papers presented at the corresponding year’s conference.

Finally, depending on the pre-existing reputation of a conference series, this study examined how much the reputation effect of a given year’s individual conference would fluctuate. When a conference series has already built a robustly noticeable reputation (i.e., constantly high ratios of impactful papers), the impact contributed by a subsequent year’s conference papers would rarely decrease. On the other hand, for conference series having a weak or unstable pre-existing reputation, the changes in the citations earned by a certain year’s individual conference would be conspicuous.





**Fig. 4** The average ratio of highly cited papers and the coefficient of variation

To examine this speculation, for each conference series, the coefficient of variation was computed for the ratios of highly cited papers from 2009 until 2012. The coefficient of variation  $C_v$  is to measure the dispersion of a distribution or frequency distribution in a standardized format (Radicchi and Castellano 2013). In this study, the coefficient of variation represents how disperse the ratios of highly cited papers over the years are in a conference series. To that end, this study first calculated the mean  $\mu$  and the standard deviation  $\sigma$  of the ratios of the highly cited papers from 2009 until 2012 for each conference series. Then, the standard deviation was divided by the mean ( $C_v = \sigma/\mu$ ).

As Fig. 4 illustrates, for each conference series, *how much the ratios of highly cited papers have changed for years ( $C_v$ )* was plotted against *the average ratio of highly cited papers at the corresponding conference for multiple years ( $\mu$ )*. This was conducted to determine how much the ratio of highly cited papers has changed in a conference series can convey different meanings, depending on the average ratio of the highly cited paper. For instance, there is a conference series where the average ratio of the highly cited papers from 2009 until 2012 is about

10%, which is relatively low, and the coefficient of the ratios' variation is 1.0, which is relatively high. There is another conference series in which the average ratio of the highly cited papers for years is about 50%, which is relatively high, and the coefficient of the ratios' variation is 0.11, which is relatively low. The conference series, like the former example, did not build up their reputation yet, so the ratio of the highly cited papers irregularly fluctuated on a year-on-year basis. Contrastingly, the conference series, like the latter example, already established their reputation; as a result, the papers presented at the conference series were cited steadily often. The impact made by the latter conference series has been stable over the years.

The patterns depicted in the Fig. 4 are similar to the examples described above. The smaller the average ratio of the highly cited papers is, the more fluctuating and disperse are the distributions of the ratio changes. The correlation tests are also equated with the patterns in the Fig. 4. The average ratio of papers cited more than 10 times or 30 times were significantly correlated with the coefficient of variation ( $r = -.57, p < .001$  for the ratio of papers cited more than 10 times;  $r = -.54, p < .001$  for the ratio of papers cited more than 30 times). That is, whereas relatively lower impacts of less reputable conference series tend to rise and lower constantly, the higher degree of impacts made by elite conference series robustly continues.

The results exhibit coherent evidence to help us better understand the existing reputation effect of a conference series and the influence on the future citations of conference papers. The analyses demonstrated that the reputation effects of conference series do exist, because some elite conferences established their reputation over the years and constantly produced a greater proportion of referable papers. Conference series do not achieve their improved reputation in a moment. When a conference series successfully builds up their reputation, the fame does not easily wane.

### The results of the regression test on future citation numbers

The primary purpose of this research is to determine the conference-related factors that have significant predictive power on the future citations of conference articles. To accomplish this purpose, the highly skewed and over-dispersed distribution of citation rates necessitates careful consideration of a regression model. Two regression models were tested. First, for the over-dispersed distribution of citations ( $\mu = 8.01, \sigma = 20.12$ ), the negative binomial multiple regression (i.e. NBMR), which is widely used in bibliometric studies (Onodera and Yoshikane 2015; Thelwall and Wilson 2014), was considered. Second, for the citation rates containing a lot of zeros (i.e. 24.2% of the whole samples), as proposed by Thelwall and Wilson (2014), the Ordinary Least Squares regression (i.e. OLS), with log normalization after adding one to the citations, was considered. Prior to conducting the analysis, among the seven factors, the factors of the numeric values were converted to Z-scores to standardize the different ranges and the variances (Keith 2014, p. 541). For both regression models, all the factors were simultaneously entered into each regression test. Following the work of Onodera and Yoshikane (2015), to compare the goodness of the fit between the NBMR and OLS tests for this study's sample data, the mean squares of the relative residuals (MSRR) of the two regression tests (e.g., Eq. 1) were compared.

$$\text{MSRR} = \sum_{i=1}^n \left( \frac{p_i - \bar{p}_i}{\bar{p}_i} \right)^2 / n \quad (1)$$

Here,  $p_i$  and  $\bar{p}_i$  are the observed and predicted values of the dependent variable for the  $i$ th paper, respectively. The MSRR value for the NBMR test was 3.19. The MSRR value for

**Table 3** Ordinary least square (OLS) regression analysis summary

Variable	<i>B</i>	SEB	$\beta$	$sr^2$
<i>Age</i>	.06**	.01	.12	.00
<i>noPaper</i>	-.07**	.01	-.12	.00
<i>acceptRate</i>	-.84**	.05	-.18	.01
Season				
January	.04*	.02	.02	.00
February	.08*	.03	.02	.00
March	.16**	.02	.06	.00
April	.18**	.01	.12	.00
May	.10**	.01	.05	.00
June	-.06**	.02	-.03	.00
July	.00	.02	.01	.00
August	.10**	.02	.05	.00
September	-.07**	.02	-.03	.00
October	-.03*	.02	-.02	.00
November	.02	.01	.01	.00
<i>contentSim</i>	1.07**	.18	.05	.00
<i>internationality</i>	.86**	.05	.14	.01
<i>bestPaper</i>	.47**	.03	.11	.01

\*\* $p < .01$ ; \* $p < .05$

the OLS test was 0.53. The MSRR for the OLS test is almost one-sixth of the value for the NBMR. Hence, the OLS test has a better fit than NBMR, and hence, was chosen for this study.

The citation rates of the conference papers were regressed on the seven independent variables (refer to “Factors under consideration” section). The results of the OLS test are presented in Table 3. Table 3 represents the unstandardized regression coefficient (*B*), the standard errors (SEB), the standardized regression coefficient ( $\beta$ ) and the squared semi-partial correlation ( $sr^2$ ). The seven variables, altogether, significantly contributed to the regression model ( $F = 222.55, p < .001$ ) and accounted for 13.6% of the variance in the citations of the conference papers ( $R^2 = .14$  and adjusted  $R^2 = .14$ ). Each of the seven independent variables had statistically significant effects on the citations of the conference papers.

As the sole factor in the level of the conference series entered into the regression test, additional years of conference series increased the citation rates of the presented papers, controlling for other independent variables ( $B = .06, t = 11.30, p < .001$ , adjusted  $sr^2 = .005$ ). Therefore, the longevity of the conference series helps to garner more citations for the presentations.

As the first factor about individual conferences of a year, the size of an individual conference represented by the number of presented papers was also a significantly explanatory factor for the number of citations of conference papers. When controlling the other independent variables, the fewer the number of papers presented at a conference, the more the presented papers got cited ( $B = -.07, t = -10.04, p < .001$ , adjusted  $sr^2 = .004$ ). In addition, low acceptance rates resulting from the high selectivity of a conference committee help to increased citation rates of the presented papers; this result was statistically significant ( $B = -.84, t = -17.71, p < .001$ , adjusted  $sr^2 = .011$ ). The time of the conferences was also statistically significant in predicting future citations. On average, when other independent

variables were controlled for, the papers presented in the earlier months of the year (e.g., January, February, March, April, May, June) earned more citations than papers presented later in the year (e.g., September, October, November, December). One exception is papers presented in 'August'. The papers presented in August received more citations than the papers presented during the latter part of the year and an equivalent level of citations with the earlier months. In American Universities, for instance, the last month of summer vacations is August. Therefore, the academic achievements presented at the beginning of the year and at the end of the vacation tend to receive more citations.

The next two factors (i.e., the overall content similarity values of papers presented at an individual conference and the degree of international co-authors participated in an individual conference) are about the diversity of the individual conferences. The diversity of the conferences was favorable overall for the increase in citations. The content similarity values of an individual conference significantly and positively predicted the future citations of the presented papers ( $B = 1.06$ ,  $t = 6.03$ ,  $p < .001$ , adjusted  $sr^2 = .001$ ). That is, the larger content similarity value is, the more diverse the content of papers presented at an individual conference is; conferences presenting papers with more diverse content tended to earn more citations. The international collaborations of participating authors at individual conferences also significantly contributed to the changes in paper citations ( $B = .87$ ,  $t = 18.38$ ,  $p < .001$ , adjusted  $sr^2 = .012$ ). Conferences where many international scholars collaborated were beneficial for authors to receive more citations for their presentations.

Finally, as the factor about individual papers, the best paper awards also had a significant predictive power for future citations. The papers that won the best paper awards substantially contributed to earn more citations ( $B = .47$ ,  $t = 17.45$ ,  $p < .001$ , adjusted  $sr^2 = .011$ ). As far as the magnitude of each factor's effect, among the seven factors considered in this paper, the international collaborations of the authors that participated in an individual conference (i.e. *internationality*) was the most contributing factor to explain the changes in the number of citations. This was followed by the best paper awards of individual conferences and the acceptance rates of individual conferences.

## Conclusions and discussion

This study contributes to a better understanding of the predictive power of conference-related factors on the future citations of conference papers. The literature about the factors affecting the citations of journal articles is well established. However, the research about the factors predicting the citations of conference papers is still in an enfant stage. This study particularly focused on how much the factors about a conference, per se, could explain the citations of the conference papers. In particular, this study investigated the three levels of factors about conferences: conference series, individual conferences and individual conference papers.

The longevity of a conference series and the names of a conference series constitute the factors considered at the conference levels. Both factors significantly predict the citations of the conference papers. The results of the regression showed that the longer history a conference series has, the more citations the presented papers have garnered. Moreover, in light of the analyses illustrated in "[Citation rates of conference papers and the reputation effects of conference names](#)" section, the reputation effects attached to the names of conferences do exist. The results indicated that elite conference series illustrate a constant preponderance of producing highly impactful papers for

years. The reputation established by a conference series did not quite change, according to the increase/decrease in the sizes of a conference series. Besides, once a conference series earned its name as a noticeable conference, its scholarly impacts contributed by the individual conferences held in each year were rather consistent and stable.

All five factors considered in the level of individual conferences also significantly contributed to the citations of the conference papers. More specifically, the fewer the papers presented at conferences, the more citations the papers tended to earn. The higher selectivity of individual conferences (i.e., lower acceptance rates of conference papers) also helped to produce more impactful papers. This paper also found that the seasonal accessibility of individual conferences—the time of a conference—significantly contributed to explaining accruing citations of conference papers. The papers presented at the beginning of the year, from January until June, or at the end of vacation season (i.e., August) received more citations than the papers presented during the other months. Conferences on more diverse topics turned out to be more favorable to collect more citations of the presented papers. Conferences having more papers written by international collaborations were beneficial to earn more citations as well. As the sole factor considered in the level of individual papers, the best paper awards also have a significant predictive power on future citations. Among the seven factors entered into the regression tests, the overall internationality of the co-authorship of individual conferences, the best paper awards and the acceptance rates contributed the most to explain the citations of the conference papers.

The contributions of this study are two-fold: (1) this is an early study to primarily target conference papers to determine the factors significantly predicting the future citations of conference papers on a relatively large scale; and (2) this study is the first attempt to provide detailed analyses of the explanatory power of various conference-related factors on the citation rates of conference papers.

With the purpose of examining the recent and sufficient citation rates of the conference papers, the collection of the sampled articles was limited to papers published between 2009 and 2012. The collection of the sampled articles was initialized from the ‘Information Science’ perspective. This is one of the limitations of the current study; hence, it is recommended that future studies increase the sample size, time span and topics. This study also has a very limited way to calculate international collaborations. It simply relied on the ratio of internationally co-authored papers to the total number of presented papers for each individual conference. However, there are a number of ways to calculate international collaborations (Gargouri et al. 2010; Ibáñez et al. 2013; Sin 2011). Therefore, in future, authors’ collaboration patterns will be explored in more diverse ways. For instance, the number of difference countries per each target paper can be calculated. Authors’ degree of international or interdisciplinary collaborations before and after publishing target papers can be also calculated. Another recommended future research direction is to perform the topic modeling of conference papers and explore how the semantic topics of conference papers explain the citations of conference articles. The academic motivations to catch up with the recency of their research topics attract many scientists’ participation in conferences (Onodera and Yoshikane 2015). As such, it is recommended that future studies analyze the recency of the research topics on conference articles and various content-related factors of conference papers, as well as the ways to predict the research impacts of the topics and content properties.

**Acknowledgements** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (NRF-2018R1C1B6002434).

## References

- Barbosa, S. D. J., Silveira, M. S., & Gasparini, I. (2017). What publications metadata tell us about the evolution of a scientific community: The case of the Brazilian human–computer interaction conference series. *Scientometrics*, *110*(1), 275–300. <https://doi.org/10.1007/s11192-016-2162-4>.
- Bartneck, C., & Hu, J. (2009). *Scientometric analysis of the CHI proceedings*. Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems, Boston, MA, USA.
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, *36*(1), 2–72. <https://doi.org/10.1002/aris.1440360102>.
- Bornmann, L., Marx, W., Gasparyan, A. Y., & Kitas, G. D. (2012). Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International*, *32*(7), 1861–1867.
- Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics*, *66*(2), 327–343. <https://doi.org/10.1007/s11192-006-0024-1>.
- Chan, H. C., Hee-Woong, K., & Chee, T. W. (2006). Information systems citation patterns from International Conference on Information Systems articles. *Journal of the American Society for Information Science and Technology*, *57*(9), 1263–1274. <https://doi.org/10.1002/asi.20413>.
- Chen, J., & Konstan, J. A. (2010). Conference paper selectivity and impact. *Communications of the ACM*, *53*(6), 79–83. <https://doi.org/10.1145/1743546.1743569>.
- Eckmann, M., Rocha, A., & Wainer, J. (2012). Relationship between high-quality journals and conferences in computer vision. *Scientometrics*, *90*(2), 617–630. <https://doi.org/10.1007/s11192-011-0527-2>.
- Elshawi, R., & Sakr, S. (2016). On analyzing the impact of authors and their collaboration patterns in the major computer algorithms research conferences. *COLLNET Journal of Scientometrics and Information Management*, *10*(1), 155–173. <https://doi.org/10.1080/09737766.2016.1177951>.
- Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, *83*(1), 243–258. <https://doi.org/10.1007/s11192-009-0021-2>.
- Freyne, J., Coyle, L., Smyth, B., & Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*, *53*(11), 124–132. <https://doi.org/10.1145/1839676.1839701>.
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, *5*(10), e13636. <https://doi.org/10.1371/journal.pone.0013636>.
- Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, *68*(3), 457–473. <https://doi.org/10.1007/s11192-006-0124-y>.
- Han, P., Shi, J., Li, X., Wang, D., Shen, S., & Su, X. (2014). International collaboration in LIS: Global trends and networks at the country and institution level. *Scientometrics*, *98*(1), 53–72. <https://doi.org/10.1007/s11192-013-1146-x>.
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>.
- Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, *39*(5), 662–673. <https://doi.org/10.1016/j.respol.2010.01.012>.
- Ibáñez, A., Bielza, C., & Larrañaga, P. (2013). Relationship among research collaboration, number of documents and number of citations: a case study in Spanish computer science production in 2000–2009. *Scientometrics*, *95*(2), 689–716. <https://doi.org/10.1007/s11192-012-0883-6>.
- Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. London: Routledge.
- Kim, M. C., & Chen, C. (2015). A scientometric review of emerging trends and new developments in recommendation systems. *Scientometrics*, *104*(1), 239–263. <https://doi.org/10.1007/s11192-015-1595-5>.
- Laender, A. H. F., de Lucena, C. J. P., Maldonado, J. C., de Souza e Silva, E., & Ziviani, N. (2008). Assessing the research and education quality of the top Brazilian Computer Science graduate programs. *SIGCSE Bulletin*, *40*(2), 135–145. <https://doi.org/10.1145/1383602.1383654>.
- Larivière, V., Sugimoto, C. R., & Cronin, B. (2012). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology*, *63*(5), 997–1016. <https://doi.org/10.1002/asi.22645>.
- Lee, D. H., & Brusilovsky, P. (2017). Improving personalized recommendations using community membership information. *Information Processing and Management*, *53*(5), 1201–1214. <https://doi.org/10.1016/j.ipm.2017.05.005>.

- Li, X., Rong, W., Shi, H., Tang, J., & Xiong, Z. (2018). The impact of conference ranking systems in computer science: A comparative regression analysis. *Scientometrics*, *116*(2), 879–907. <https://doi.org/10.1007/s11192-018-2763-1>.
- Lisée, C., Larivière, V., & Archambault, É. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology*, *59*(11), 1776–1784. <https://doi.org/10.1002/asi.20888>.
- Loizides, O.-S., & Koutsakis, P. (2017). On evaluating the quality of a computer science/computer engineering conference. *Journal of Informetrics*, *11*(2), 541–552. <https://doi.org/10.1016/j.joi.2017.03.008>.
- Martins, W. S., Gonçalves, M. A., Laender, A. H. F., & Ziviani, N. (2010). Assessing the quality of scientific conferences based on bibliographic citations. *Scientometrics*, *83*(1), 133–155. <https://doi.org/10.1007/s11192-009-0078-y>.
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105–2125. <https://doi.org/10.1002/asi.20677>.
- Michels, C., & Fu, J.-Y. (2014). Systematic analysis of coverage and usage of conference proceedings in web of science. *Scientometrics*, *100*(2), 307–327. <https://doi.org/10.1007/s11192-014-1309-4>.
- Montesi, M., & Owen, J. M. (2008). From conference to journal publication: How conference papers in software engineering are extended for publication in journals. *Journal of the American Society for Information Science and Technology*, *59*(5), 816–829. <https://doi.org/10.1002/asi.20805>.
- Nomaler, Ö., Frenken, K., & Heimeriks, G. (2013). Do more distant collaborations have more citation impact? *Journal of Informetrics*, *7*(4), 966–971. <https://doi.org/10.1016/j.joi.2013.10.001>.
- Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, *66*(4), 739–764. <https://doi.org/10.1002/asi.23209>.
- Peng, T.-Q., & Zhu, J. J. H. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology*, *63*(9), 1789–1803. <https://doi.org/10.1002/asi.22649>.
- Radicchi, F., & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, *97*(3), 627–637. <https://doi.org/10.1007/s11192-013-1027-3>.
- Rahm, E., & Thor, A. (2005). Citation analysis of database publications. *SIGMOD Record*, *34*(4), 48–53. <https://doi.org/10.1145/1107499.1107505>.
- Sakr, S., & Alomari, M. (2012). A decade of database conferences: A look inside the program committees. *Scientometrics*, *91*(1), 173–184. <https://doi.org/10.1007/s11192-011-0530-7>.
- Scopus. (2013). Content coverage guide. Retrieved from <https://files.sciverse.com/documents/pdf/ContentCoverageGuide-jan-2013.pdf>.
- Shirakawa, N., Furukawa, T., Nomura, M., & Okuwada, K. (2012). Global competition and technological transition in electrical, electronic, information and communication engineering: Quantitative analysis of periodicals and conference proceedings of the IEEE. *Scientometrics*, *91*(3), 895–910. <https://doi.org/10.1007/s11192-011-0566-8>.
- Sin, S.-C. J. (2011). International coauthorship and citation impact: A bibliometric study of six LIS journals, 1980–2008. *Journal of the American Society for Information Science and Technology*, *62*(9), 1770–1783. <https://doi.org/10.1002/asi.21572>.
- Song, M., Heo, G. E., & Kim, S. Y. (2014). Analyzing topic evolution in bioinformatics: Investigation of dynamics of the field with conference data in DBLP. *Scientometrics*, *101*(1), 397–428.
- Souto, M. A. M., Warpechowski, M., & de Oliveira, J. P. M. (2007). *An ontological approach for the quality assessment of computer science conferences*. Berlin: Springer.
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225. <https://doi.org/10.1007/s11192-016-1889-2>.
- Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, *8*(4), 963–971. <https://doi.org/10.1016/j.joi.2014.09.011>.
- Vanclay, J. K. (2013). Factors affecting citation rates in environmental science. *Journal of Informetrics*, *7*(2), 265–271. <https://doi.org/10.1016/j.joi.2012.11.009>.
- Vardi, M. Y. (2010). Revisiting the publication culture in computing research. *Communications of the ACM*, *53*(3), 5. <https://doi.org/10.1145/1666420.1666421>.
- Vasilescu, B., Serebrenik, A., Mens, T., van den Brand, M. G. J., & Pek, E. (2014). How healthy are software engineering conferences? *Science of Computer Programming*, *89*, 251–272. <https://doi.org/10.1016/j.scico.2014.01.016>.

- Vrettas, G., & Sanderson, M. (2015). Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology*, 66(12), 2674–2684. <https://doi.org/10.1002/asi.23349>
- Wainer, J., Eckmann, M., & Rocha, A. (2015). Peer-selected “best papers”—Are they really that “good”? *PLoS ONE*, 10(3), e0118446. <https://doi.org/10.1371/journal.pone.0118446>.
- Wainer, J., Przibiszczki de Oliveira, H., & Anido, R. (2011). Patterns of bibliographic references in the ACM published papers. *Information Processing and Management*, 47(1), 135–142. <https://doi.org/10.1016/j.ipm.2010.07.002>.
- Wainer, J., & Valle, E. (2013). What happens to computer science research after it is published? Tracking CS research lines. *Journal of the American Society for Information Science and Technology*, 64(6), 1104–1111.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>.
- Wersig, G. (1993). Information science: The study of postmodern knowledge usage. *Information Processing and Management*, 29(2), 229–239. [https://doi.org/10.1016/0306-4573\(93\)90006-Y](https://doi.org/10.1016/0306-4573(93)90006-Y).
- Wuehrer, G. A., & Smejkal, A. E. (2013). The knowledge domain of the academy of international business studies (AIB) conferences: A longitudinal scientometric perspective for the years 2006–2011. *Scientometrics*, 95(2), 541–561. <https://doi.org/10.1007/s11192-012-0909-0>.
- Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications. *Scientometrics*, 101(2), 1491–1513. <https://doi.org/10.1007/s11192-014-1264-0>.
- Zhang, Y., & Jia, X. (2013). Republication of conference papers in journals? *Learned Publishing*, 26(3), 189–196. <https://doi.org/10.1087/20130307>.