This is the peer reviewd version of the followng article:

Does the venue of scientific conferences leverage their impact? A large scale study on Computer Science conferences / Bedogni, L.; Cabri, G.; Martoglia, R.; Poggi, F.. - In: LIBRARY HI TECH. - ISSN 0737-8831. - 41:2(2023), pp. 501-515. [10.1108/LHT-09-2021-0305]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

05/05/2024 01:08

(Article begins on next page)

Does the Venue of Scientific Conferences Leverage their Impact? A Large Scale study on Computer Science Conferences

January 17, 2022

Abstract

Purpose: Conferences bring scientists together and provide one of the most timely means for disseminating new ideas and cutting-edge works. The importance of conferences in many scientific areas is testified by quantitative indicators. The main goal of this paper is to investigate a novel research question: is there any correlation between the impact of scientific conferences and the venue where they took place?

Approach: To measure the impact of conferences we conducted a large scale analysis on the bibliographic data extracted from 3,838 Computer Science conference series and over 2.5 million papers spanning more than 30 years of research. To quantify the "touristicity" of a venue we exploited indicators about the attractiveness of a venue from reports of the World Economic Forum, and we have extracted 4 country-wide touristic indicators and 2 city-wide, which measure the attractiveness and the touristicity of any country or city.

Findings: We found out that the two aspects are related, and the correlation with conference impact is stronger when considering country-wide touristic indicators, achieving a correlation value of more than 0.5 when considering the average citations, and more than 0.8 when considering the total citations. Moreover the almost linear correlation with the Tourist Service Infrastructure index attests the specific importance of tourist / accommodation facilities in a given country.

Originality: This is the first attempt to focus on the relationship of venue characteristics to conference papers. The results open up new possibilities, such as supporting conference organizers in their organization efforts.

Keywords: Citation analysis; Conferences; Information science; Bibliometric analysis; Bibliometric indicators; Research impact; Correlation analysis.

Article classification: Research Paper.

1 Introduction

Conferences play an important role in many disciplines, in particular in scientific areas. Given the evolution speed of these fields, conference papers are perceived as one of the most timely means for disseminating new ideas since the cycle of publication is much shorter than journal ones. Moreover, conferences open up opportunities to bring scientists together and introduce cutting-edge works, even if less mature and sophisticated. As a consequence in specific disciplines like Computing and Information Science, for instance, 62.3% of the articles published on Scopus are conference papers, while only about one-third (i.e. 32.8%) were journal papers, as reported in Scopus (2020).

As researchers, our main goal is to provide our scientific communities with novel contributions that increase our body of knowledge or enable us to see the world under a different point of view. However our contributions must be validated and spread by appropriate means such as international conferences. Of course, researchers choose the most suitable venue/conference depending on the specific contribution, but we wonder whether there are other factors that can influence the choice. There are studies that have shown that such factors exist for conferences (see Section 2) but no work has focused on the *touristic appeal* of the conference venue.

Our initial idea was that a conference venue able to attract tourists is also attractive for researchers to submit their paper. Of course our claim is not that the venue is the "only" or "most important" factor, but we claim that it is a factor considered by researchers, as hosting a conference in a more touristic place also offers better experiences for participants outside of the conference sessions. However, as scientists, we are aware that a feeling is not enough and an idea must be supported by data and their analysis. So, we decided to explore the correlation between the attractiveness of a conference and the "touristicity" of its venue from a data-driven point of view.

First of all, we had to better define the two concepts to be correlated.

Intuitively, the attractiveness of a conference is defined by the number of submissions it receives; unfortunately, this number is not so easy to retrieve: while the number of the accepted papers is public, the number of the submitted papers is not and only some conferences publish it. So we decided to rather focus on the conference *impact*, and we did it by exploiting another indicator, the number of the citations of the papers published in the conference proceedings: the two aspects go indeed hand in hand, as an attractive conference with a lot of submissions can select better papers that will have eventually more citations, therefore having a higher impact on the research community.

In the same way, the "*touristicity*" of a venue can be very intuitive but is not straightforward to be formally defined. To this purpose, we exploited some touristic indicators that represent the "venue factors" for both cities and countries, such as the size of the Wikipedia page for the city hosting the venue, the number of annual tourist arrivals and other officially defined indexes from

the Travel and Tourism competitiveness report of the World Economic Forum[1].

Our main research questions are therefore the following:

- **RQ1**: is there any correlation between the impact of a scientific conference and the venue where it takes place?
- **RQ2**: if yes, which are the venue indicators that are more involved in the correlation?

To answer these questions, we focused on the conferences in the field of Computer Science and we built a large dataset comprising nearly 4000 conference series and over 2.5 million papers spanning more than 30 years.

We grouped and aggregated the data following different criteria and producing data visualizations that enabled us to analyze the context from different points of views. We computed the required indicators from such views and, more specifically, we computed the correlation between the impact indicators and the touristicity indicators using three different correlation measures. We present the obtained results including the discussion of some representative examples, and we show that a correlation exists between the number of citations and the indexes from the Travel and Tourism competitiveness report; in particular, there are two indexes (TTCI – Travel & Tourism Competitiveness Index – and TSI – Tourist service infrastructure) which show almost a linear correlation.

This paper exploits our past expertise in bibliometric analyses (BLIND (2019b), BLIND (2019a), BLIND (2020)) and focuses on scientific conferences. As mentioned, the novelty of our work is that no other research analyzed specifically the correlation between the impact of a conference and the venue where it took place, as far as we know. Previous works investigated what are the main factors that influence conference participation decision-making (e.g. Terzi et al. (2013); Borghans et al. (2010); Lee (2019)) and agree that, among the factors considered by the authors when deciding whether to attend a conference, many concern the destination with its cultural and infrastructural potential. However, only a few recent works studied the relation between these factors and the impact of conferences (e.g. Goel et al. (2015); Lee (2020, 2019)), and none of them analyzed the relation between the touristicity of a conference venue and quantitative indicators of the impact of research papers presented at conferences.

We also point out that this study provides a comprehensive view on the topic, being the size of our dataset much larger (i.e. at least two orders of magnitude greater, both in terms of conferences series and papers considered) than the ones used in similar analyses, as discussed in Section 2.

This paper is organized as follows: at first, we present some related work in Section 2 to show the novelty of our research; we then moved to our methodology, detailing how we analyzed the data in Section 3; we show the results of our analysis in Section 4, and we eventually conclude our work in Section 5.

[1] https://www.weforum.org/

2 Related Work

In the Computer Science field, conferences play a very important role. They are channels of paramount importance used by scientists to share their researches, as discussed in Glänzel et al. (2006); Lisée et al. (2008); Vrettas and Sanderson (2015); Singh et al. (2016).

The importance of conferences in Computer Science is testified by quantitative indicators. For instance, Scopus 2020 Content Coverage Guide Scopus (2020) reports that 62.3% of articles in Computing and Information Science were conference papers, while only about one-third (i.e. 32.8%) were journal papers. In other disciplines like Chemical Sciences, Biological Sciences and Medical & Health Sciences only 1.9%, 2.7% and 2.9% were from conferences, respectively. In Laender et al. (2008) the authors compare the publications from 30 Computer Science graduate programs in North-American, Europe and Brazil and observe that the ratios of conference papers to journal articles were 2.9, 2.5 and 2.1, respectively. The empirical evidence supporting the impact of conferences in Computer Science is also discussed in Bar-Ilan (2010).

A lack of metrics supported by the citation databases is one of the main reasons that make hard to persuade non-computer scientists that conference papers have a value. Measures such as the acceptance rate of conferences are sometimes used as a proxy of quality, but cannot be compared with the Impact Factor of journals. Although Impact Factor is commonly calculated for journals only, recent studies developed proposals for a widely adoptable quality evaluation and ranking system for conferences, which could be equivalent to the Impact Factor Li et al. (2018); Meho (2019); Loizides and Koutsakis (2017). Besides, He and Han observe that conference papers, whose main objective is communication, have a stronger correlation with altmetrics than journal papers. Thus, altmetrics can be suitable to assess the impact of conference papers in a fast-moving research environment such that of Computer Science, as discussed in He and Han (2017). A recent study by Yang and Qi (2020) also reports that altmetrics indexes of conference papers in science fields are closely related to citations, reflecting the citation impact of proceedings papers.

One of the key aspect to explain the importance of conferences in the Computer Science domain is that, given the evolution speed of this discipline, conference papers are perceived as one of the most timely means for disseminating new ideas (Vardi (2010); Eckmann et al. (2012)). In fact, unlike the cycle of publication of articles in journals, the time between the submission of an article in a conference and the editorial decision and publication is very short.

Moreover, by offering various formats of presentations (e.g., discussions, presentations, posters, industry and position papers), conference open up opportunities to present cutting-edge ideas, even if less mature and sophisticated, bringing scientists together. As a consequence, the attendees can disseminate and get instant feedbacks about their researches and ideas, and establish new collaborations triggered by on-site interactions, as described in Freyne et al. (2010).

Many studies investigated what are the main factors that influence con-

ference participation decision-making. In Terzi et al. (2013) the authors analyze the answers of 123 academics and students to questionnaires based on the five-point Likert Scale, and discover that the destination, with its cultural and infrastructural potential, is one of the most important factors that determine the decision-making by researchers regarding their participation in conferences and exhibitions. The paper focuses on five evaluation criteria of conference location that have been derived from a literature review, and conclude that the most important ones are *means of transport* (87.81% of positive answers), *safety/hygiene* (86.99%) and *use of international language* (83.74%), followed by *security* (77.24%) and *infrastructure* (73.98%).

Similar findings are described in Borghans et al. (2010). Based on a sample of European labour economists, preferences are measured using the vignette approach where participants are asked to choose between hypothetical European Association of Labour Economists (EALE) conferences. The results show that *Keynote speakers* and *conference location* are the most important attributes to decide for a conference, while the remaining attributes *type of social event*, *time of the year*, and *conference venue* are less important for the decision. The paper also shows that varying characteristics of a conference can influence both the overall attractiveness of a conference, but also influences the type of researchers interested in participation.

In Yoo and Chon (2008) the authors developed a reliable and valid five-factor scale for convention participation decision-making based on 558 survey responses. The final measurement scale consists of five interrelated but unique dimensions of convention participation decision-making: *destination stimuli*, *professional and social networking opportunities*, *educational opportunities*, *safety and health situation*, and *travelability*.

All these studies agree that, besides scientific motivations supported by bibliometric and quantitative indicators, most of the factors considered by the authors when deciding whether to attend a conference, thus determining its success, concern the destination, with its cultural and infrastructural potential. However, only a few studies analyzed the relation between the touristicity of a conference venue and quantitative indicators of the impact of research papers presented at conferences.

A recent work by Lee (2019) examines 43,463 papers from 81 conference series in the Information Science and Computer Science fields and analyzes the contributions of conference related factors to the citation rates of the conference papers. The results of the study reports that two factors related to the attractiveness of the conference venue are able to predict the quality of presented papers, measured in terms of their citation rates: the *seasonal accessibility of conferences* (e.g. when a conference is held during the middle of a school semester, faculty and student researchers may feel burdensome to cancel or skip classes, and therefore, if there is another alternative option that is held during the vacation season, the researchers may select that alternative) and the *size of the conferences* in terms of the number of presented papers, which is a direct consequence of the venue attractiveness, especially for non top-rated conferences. The regression results also illustrate that other aspects,

such as longevity and names of the conference series, their acceptance rates, 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 are significantly predictive to the future citations of the conference papers.

In Goel et al. (2015) the authors analyze 63 networking conferences and about 39,000 conference papers published in these conferences during the years 2008 through 2012. An empirical study investigate the impact on the number of citations received by papers of three factors: conference acceptance rate, paper title and year of publication. The results show a correlation between the analyzed factors and citation count. However the authors acknowledge that there are several other factors (not considered in their empirical study) that may influence the citation count of conference papers by as much as 46%.

Another study by Lee (2020) analyze 12,237 papers published in 28 conference series in the Computer Science field. The paper analyzes the predictive power of 21 factors related to the first author and to all authors on citation counts of conference papers. In particular, the work considers authors' academic performance, degree of collaboration and topological properties of their citation networks. The results show that the author-related factors have a significantly higher power for explaining conference paper impact in terms of citation counts than the first author-related factors. Among all author-related indicators, the average degree centrality of all authors is the most relevant factor for predicting future citations, followed by the average of all authors' betweenness centrality and average citation rates of prior publications.

The works discussed in this section confirm the importance of conferences for specific research fields such as Computer Science. However, as shown in Table I, previous studies considered only a limited subset of conferences in the Computer Science field, both in terms of conference series/number of papers and time span. Moreover, most of them acknowledge that several other factors besides those traditionally considered by the bibliometric community should be considered to fully understand the impact of conferences, such as the conference destination with its cultural and infrastructural potential. However, at the best of our knowledge, no study has yet investigated in depth the relation between the touristicity of a conference location and quantitative indicators of the impact of research papers presented at conferences.

[Table 1 about here.]

3 Methods and Materials

For the analyses described in this paper, we worked on publication citation data (Section 3.1), quantifying citations for scientific conference papers, and integrating it with tourism information data (Section 3.2), characterizing the cities and countries that hosted the conferences each year. The data gathered and the software developed for our work are available in BLIND (2021) and in

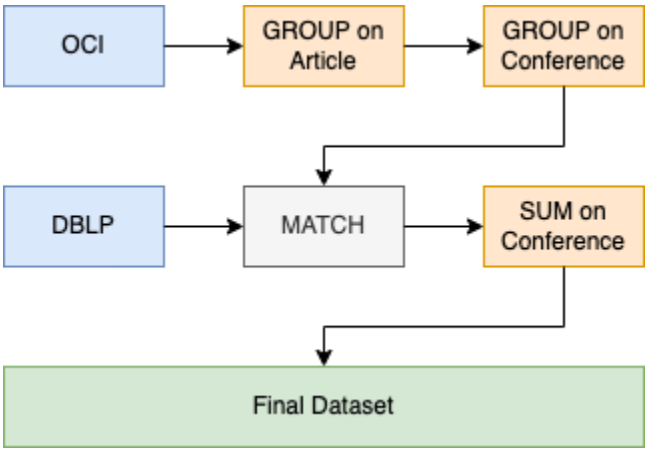


Figure 1: Data extraction flow

the GitHub repository[2], respectively.

[Table 2 about here.]

3.1 Citation data - conference impact.

To quantify the impact of a conference, we started from the raw data extracted from the OpenCitations[3] website. Opencitations is an open and promising dataset, as also stated by Visser et al. (2021). First publishing open bibliographic data in 2010, the Initiative for Open Citations[4] (I4OC) was born with the idea of promoting the release of open citation data, as discussed in Shotton (2013). As a result, now we have several millions of citation data openly available on the Web, important stakeholders supporting the movement, and several projects and studies (e.g. Di Iorio et al. (2019); Zhu et al. (2020); Bologna et al. (2021)) leveraging the open citation data available online.

In this work we used OpenCitations as the main datasource of bibliometric information for our analyses. The data is provided with records which identify each citing article and each cited article. Since we are focusing on the number of citation of a conference, we summed all the citations received by any article, so that now we are able to asses its impact. At this point we are left with a large dataset, which however contains contributions from different fields. We then merged the data with that available from DBLP[5], a free and publicly available Computer Science bibliography repository started in 1993 at the University of Trier, which describes Computer Science conferences. We matched the two datasets on the conference at which an article was presented, in order

[2](BLIND) https://github.com/.../conference-touristicity
[3]https://opencitations.net/
[4]https://i4oc.org
[5]https://dblp.org/

to keep only Computer Science articles, and then summed all the citations of articles presented in any given conference. The resulting dataset contains complete citation data for scientific publications in the field of Computer Science ranging from 1960 to the present day. The overview of the described process is pictured in Figure 1. Several variables are available, including: authors, title, booktitle, crossref, url (DBLP page link), editor, pages, publisher, series, year, cit (number of received citations). First of all, we projected the data on the variables that are relevant to our analysis: authors, title, crossref (containing a unique key for each edition of a conference), url, year, cit (see Table II for a sample publication entry). Then, we proceeded with data filtering/cleaning and we explicitly excluded from our working set:

- publications not appearing in conference proceedings or without conference information (e.g., journals, technical reports, etc.);
- publications without citation information;
- recent publications from the last 2-3 years (staring from 2018), since they are too recent for citation analysis;
- publications from too early years (we grouped the data per year and noticed a crescent number of publications trend, we required at least 2500 publications for each analyzed year and thus we selected publications from 1985 onwards).

[Table 3 about here.]

[Table 4 about here.]

We consequently obtained a working dataset of 2,581,564 entries. Starting from this data, we processed it in order to obtain different "views" and be able to later analyze it from different points of view (also by crossing such views with the tourism data which will be described in the next section):

- **Conferences high-level aggregate statistics**: grouping by conference key (e.g., "eurocrypt"), we computed total ("totCit") and average ("avgCit") citation number, total number of publications ("nPubs") and editions ("nEditions"), see Table III for sample data. In particular, the total and average citation number are the two main indicators we will use to quantify the conference impact;
- **Conferences editions high-level aggregate statistics**: same as last point, but with data grouped on conference edition (conference name and year, e.g., "eurocrypt/2010");
- **Temporal evolution data**: for each conference, the evolution through its different editions/years of the total number of citations (see Table IV), average number of citations, total number of publications;

- **Relative temporal evolution data (percentages)**: same as last point, but with data for each year expressed in percentage w.r.t. all editions/years (e.g., average citations for conference "ismb" for year XY divided by the average citations computed on all editions data, and so on for all years). This view enables a better glance at the relative variations of each monitored citation metric for each conference over time.

3.2 Tourism data - conference venue "touristicity".

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

To quantify the "touristicity" of a venue, we first of all extracted location information for the conferences of our dataset. In particular, we extracted the exact city, country and state for each edition of each conference (see Table V for an example). Then, we gathered and linked several statistics from external resources characterizing the extracted locations from different tourism-related points of view.

For each country we extracted the following indicators from the latest Travel and Tourism competitiveness report[6], which quantify the "venue factors" mentioned above:

- **TTCI** (Travel & Tourism Competitiveness Index), the most general index, quantifying the whole set of factors and policies that enable the sustainable development of the Travel & Tourism sector in a given country (range 1-7). In particular, the index is composed of 14 pillars organized into four categories: enabling environment (business environment, safety and security, health and hygiene, human resources and labour market, ICT Readiness), T&T policy and enabling conditions (prioritization of travel & tourism, international openness, price competitiveness, environmental sustainability), infrastructures (air transport, ground and port, tourist service), natural and cultural resources (natural resources, cultural resources and business travel);

- **TSI** (Tourist Service Infrastructure), quantifying the availability of sufficient quality accommodation, resorts and entertainment facilities in a given country (range 1-7). The index is measured through the number of hotel rooms complemented by the extent of access to services such as car rentals and ATMs;

- **CRBT** (Cultural Resources and Business Travel), measuring a country's cultural resources (range 1-7). Among the considered factors are the number of UNESCO cultural World Heritage sites, the number of large stadiums that can host significant sport or entertainment events, the number

[6] http://reports.weforum.org/travel-and-tourism-competitiveness-report-2019/

of online searches related to a country's cultural resources and the number of international association meetings;

- **GPI** (Ground and Port Infrastructure), measuring a country's availability of efficient and accessible transportation to key business centres and tourist attractions (range 1-7). More specifically, the index considers the presence of extensive road and railroad network, proxied by road and railroad densities, as wells as roads, railroads, and ports infrastructure that meet international standards of comfort, security and modal efficiency.

We further extended our analysis, by exploiting the possibility to derive the impact of a conference from the city in which it took place. Clearly, this is an even more complex and unexplored field, as there are additional factors to be taken into account, such as the closeness to popular attractions and local events. This is a first effort towards this novel direction of research, as we investigate this possibility by analyzing the citations of a conference against two city-wide indicators, defined as:

- **TA** (Tourist Arrivals): this is the number of tourist arrivals (in millions) to each city as monitored in 2018 by Euromonitor[7], for which we obtained all the relevant data and merged it with the conference data;
- **SWP** (Size of Wikipedia Page): this indicator represents our proposal to easily quantify the importance/attractiveness of the city and expresses the dimension (in bytes) of the associated Wikipedia page. The rationale is that the higher is the importance of a city the more information we expect to be available in its description.

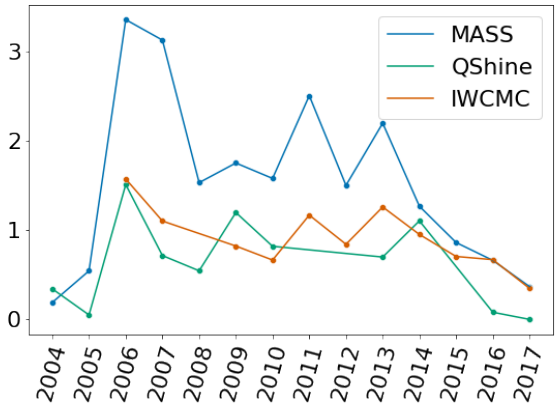
Finally, by joining the above city and country tourism data with the citation data (total and average citation and number of editions, aggregated per city), we were able to obtain new combined views of our data (see Tables VI and VII for city and country, respectively).

4 Results

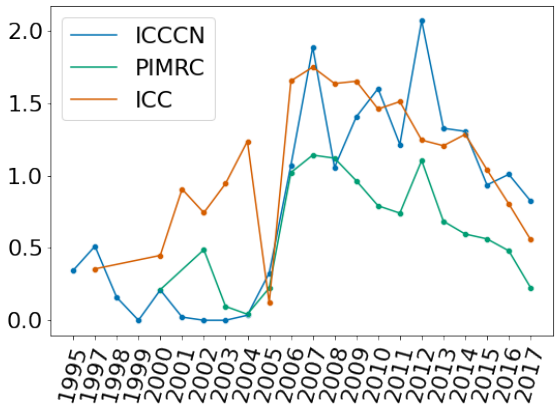
In this section we discuss the results we have obtained, with respect to the variation in the citation number among the same conference in different years and in different places.

At first, we show some examples which highlight the differences in the number of citations, for the same conference series over different years. Figure 2 and Figure 3 show the average citations for different conferences, for any given year. Each plot refers to different classes of conferences, according to the Italian Conference Ranking System[8], which comprises several other international conference evaluations, such as the CORE conference ranking, Microsoft Academic,

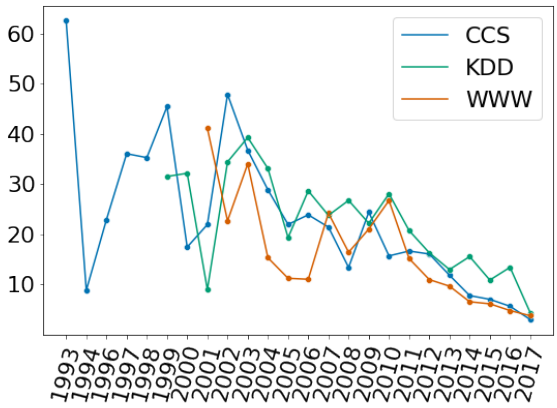
[7] https://go.euromonitor.com/white-paper-travel-2018-100-cities
[8] http://valutazione.unibas.it/gii-grin-scie-rating/



(a) Class 3

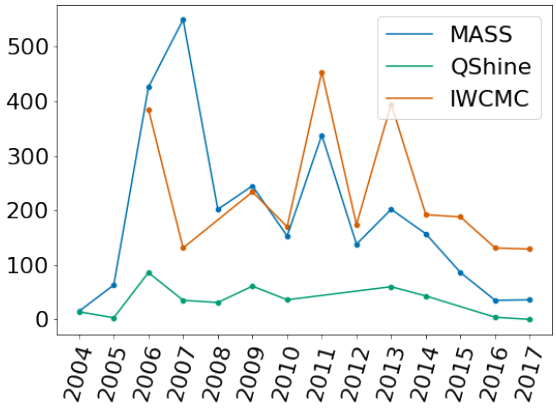


(b) Class 2

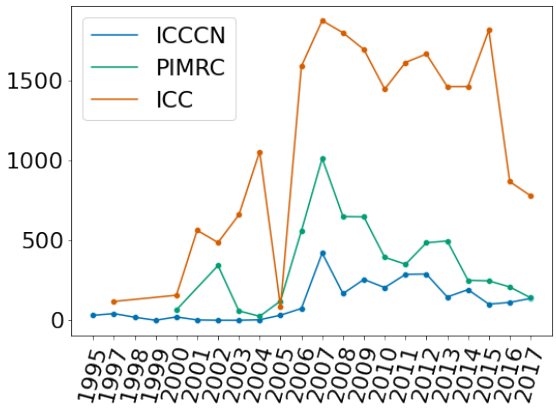


(c) Class 1

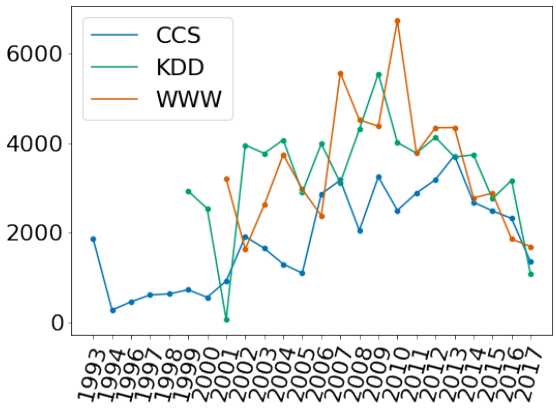
Figure 2: Examples of different conference with respect to the Average Citation number of each year.



(a) Class 3



(b) Class 2



(c) Class 1

Figure 3: Examples of different conference with respect to the Total Citation number of each year.

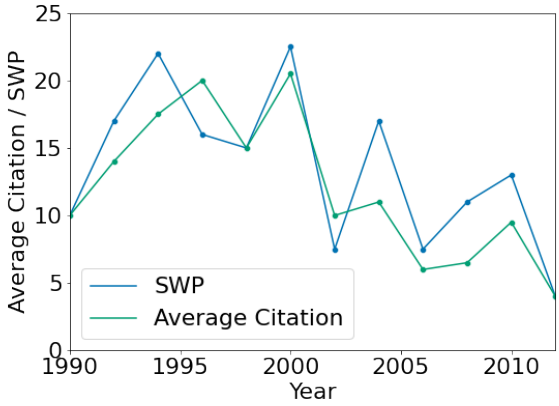


Figure 4: PPSN conference over the years

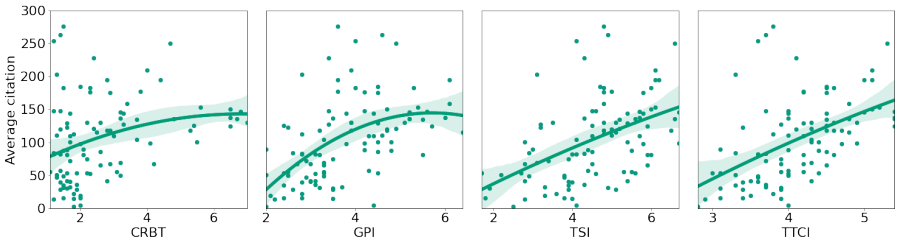


Figure 5: Average citations versus tourism indices

and LiveSHINE. We can see that, regardless of the class of the conference, both the average citation number and the total citation number vary a lot during different years, although for Class 1 conferences it is less evident. There is a global drop starting from around 2015, which is given to the fact that articles are new and have yet to collect a number of citations comparable to older articles. Although a slight variation between any year is normal, there are evident cases in which from one year to another the number of citations, and the average of it, vary even by 400%.

We now considered whether indicators such as those presented in Section 3.2 are correlated with the average citation number, meaning that better indicators should reflect a higher number of citations. In particular, we started by visually analyzing the trends focusing on one conference at a time. As an example, we show in Figure 4 the average citation number of the PPSN conference over the years, and the SWP indicator of the venue in which the conference took place on any given year. As it can be seen, the two lines follow a rather similar trend, which may indicate that indeed a correlation exists.

Then, in order to generalize our findings, we considered all the conferences

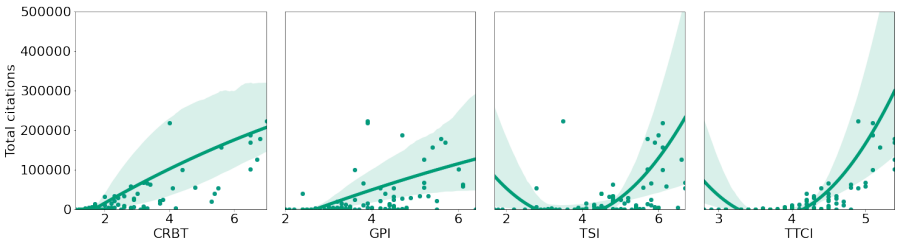
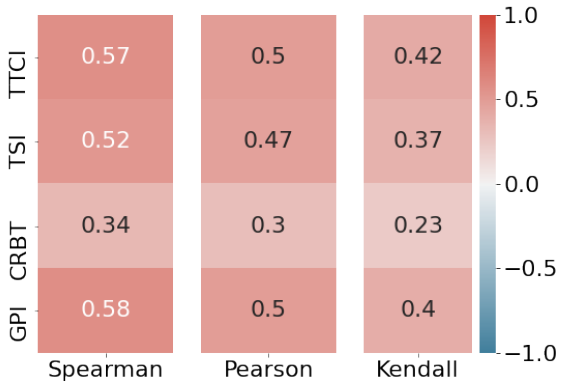


Figure 6: Total citations versus tourism indices

of our dataset and all the 4 country wide indicators presented in Section 3.2 (results are shown in Figure 5). Here, each point refers to a specific venue, and the average citation is the average of the citation number of any conference in that specific venue. On the x axis we show the different metric values for the touristicity metrics considered, and as it can be easily seen, the trend is linear for all four indicators, as also shown by the fitting curves. We also report the same evaluation with respect to the total citation in Figure 6. Also in this case the trend grows with the increase of the touristicity index, although for GPI and TSI it starts to increase more convincingly around a value of 4.

Figure 7 shows instead the results of the correlation between the Average Citation number (Figure 7a) and the touristicity metrics, and the Total Citation number (Figure 7b) and the same metrics. In these two figures, all the touristicity indicators are related to the country in which the conference took place. We started our analysis from the data related to all the conferences divided by year, and correlated all of them together with the touristicity indicators. For sake of readability, we do not report the p-values for each correlation values, which however are all below 1^{-5}, therefore representing a trusted correlation measure. We have performed the tests using three different correlation measures, namely Spearman, Pearson and Kendall. What it can be seen is that according to Spearman there is a rather strong correlation with both the Average and the Total Citation number and various indices which represent the touristic attractiveness of a city in which a conference took place. Pearson is the one which shows less correlation, though for some indices it highlights a correlation of about 0.5. Kendall shows instead good correlation for a higher number of metrics, both for the Average and the Total Citation number. In any case, it is evident that it exists a correlation between the citation number of a conference and the city in which it took place.

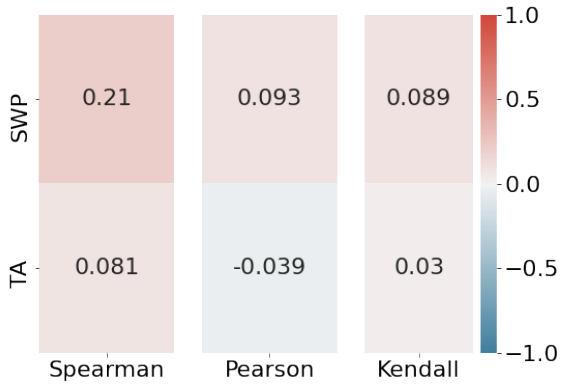
Figure 8 shows instead indices related to the attractiveness of a city instead of a country. Here we can see that although in some cases there are quite interesting correlations, values are much lower compared to Figure 7. Therefore, we can say that according to our analysis, it exists a stronger correlation when considering the country in which the conference is organized, rather than the specific city. We also note that this particular aspect is one of the main future work on this topic.



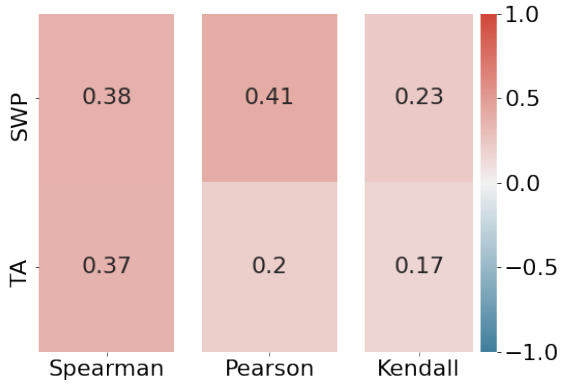
(a) Average Citations

(b) Total Citations

Figure 7: Country wide indices: Correlation between the different indices and the Average Citation number (Figure 7a) and the Total Citation number (Figure 7b).



(a) Average Citations



(b) Total Citations

Figure 8: City wide indices: Correlation between the different indices and the Average Citation number (Figure 8a) and the Total Citation number (Figure 8b).

5 Discussion and concluding remarks

In this paper we have presented a study related to the correlations between the attractiveness of the scientific conferences and the "touristicity" of their venues.

According to our original research questions, we have shown that a correlation exists between the number of the citations of the conference publications and some touristic indicators of the conference venue. So we can conclude the paper answering the two research questions.

- **RA1**: Yes, there is a correlation between the impact of a scientific conference and the venue where it takes place, and the correlation is stronger when considering country-wide indicators; we may also argue that the proximity of the participants current residence with the country of the conference could play a major role, and we plan to explore this aspect as a future work. We also acknowledge that, at this stage of research, country-wide indicators may be more robust than the proposed city-wide counterparts.

- **RA2**: The two touristic indexes that are more involved in the correlation are the TTCI, Travel & Tourism Competitiveness Index, and the TSI, which is the Tourist Service Infrastructure. The high correlation of the TTCI, being it the most comprehensive index among the ones considered, corroborates our general findings, while the particular presence of the TSI attests the specific importance of the presence of tourist/accommodation facilities in a given country.

The above indicators present an almost linear correlation with the number of citations, and this also shows that country-wide indexes are more correlated with the impact than city-wide indexes; this can be explained because the chance of visiting a touristic region goes beyond the visit of a single city.

Starting from our results concerning the correlation between different touristicity indicators and the outcome of a conference in terms of citations, it would be possible to support conference organizers in their decisions. For instance, they could plan in advance conference venues considering the same touristicity indicators, comparing different options and selecting cities which have high scores. This will allow for rapid planning of a conference venue, encompassing the easiness of travel and the attractivity of a venue, hence increasing the potential outcomes of the conference.

Regarding the social implications, our study will enable the possibility for municipalities and conference organizers to understand what it can be improved in a specific venue to make it more attractive. This may include better transport connections or selecting cities which show a high potential regarding the touristicity index. Regarding the willingness of a researcher to submit a paper to a specific conference, it would be unaltered, meaning that what our results show is that there is already a mental process, before submitting a paper to a conference, which considers these indicators.

5.1 Limitations and Future Work

We are aware of the two main limitations of our work: (i) the use of citations to evaluate the attractiveness of the conferences and (ii) the difficulty to formally define the touristic attractiveness of a venue. With regard to the attractiveness of the conferences, the most accurate parameter would be the number of the submissions, because we can define the *attractiveness of a conference* as the ability to attract submissions from the authors. The use of the number of citations is an indirect parameter: we suppose that the more the number the submissions, the more the chance of selecting good papers, the higher the quality of the conferences, the higher the number of citations. This is reasonable, but not proved yet. With regard to a definition of the *touristic attractiveness of a venue*, we are not aware of a formal and accepted one. Maybe it is difficult to formalize, because different venues can attract for different reasons (art, history, culture, food, ...) and the attractiveness can be considered at different levels (e.g., city, region, country, as shown before). The existence of different indexes is an evidence of this.

In our future work we aim at overcoming the former limitation by analyzing the number of the submissions instead of the number of citations; this should better represent the interest in submitting a paper to any conference, depending on the year hence on the organization location. This will also be useful to validate the assumption that a large number of submissions implies more citations. Moreover, we will consider: a) the definition of other touristicity indicators by extracting data from further sources, trying to quantify the impact of specific factors to the overall venue attractiveness; b) the possibility of further expanding the dataset. We also remark that the results open up new possibilities, such as allowing both conference organizers and authors to estimate in advance the impact of conferences, thus supporting them in their decisions.

Declarations

Funding: no funding was received for conducting this study.
Conflicts of interest/Competing interests: the authors have no relevant financial or non-financial interests to disclose.
Availability of data and material: https://doi.org/10.5281/zenodo.4734182.
Code availability: https://github.com/lbedogni/conference-touristicity
Authors' contributions: all authors contributed equally to this research.

References

Bar-Ilan, J. (2010). Web of science with the conference proceedings citation indexes: The case of computer science. *Scientometrics*, 83(3):809–824. cited By 42.

BLIND (2019a). Do altmetrics work for assessing research quality? *Scientometrics*, 118(2):539–562. cited By 18.

BLIND (2019b). Predicting the results of evaluation procedures of academics. *PeerJ Computer Science*, 5. cited By 3.

BLIND (2020). The practice of self-citations: a longitudinal study. *Scientometrics*, 123(1):253–282. cited By 7.

BLIND (2021). Dataset for the Article "Does the Venue of Scientific Conferences Leverage their Impact? A Large Scale study on Computer Science Conferences".

Bologna, F., Di Iorio, A., Peroni, S., and Poggi, F. (2021). Do open citations inform the qualitative peer-review evaluation in research assessments? An analysis of the Italian National Scientific Qualification. *Submitted for publication in Scientometrics, preprint available in ArXiV at* `https://arxiv.org/abs/2103.07942`.

Borghans, L., Romans, M., and Sauermann, J. (2010). What makes a good conference? analysing the preferences of labour economists. *Labour Economics*, 17(5):868–874. cited By 14.

Di Iorio, A., Poggi, F., and Peroni, S. (2019). Open data to evaluate academic researchers: An experiment with the italian scientific habilitation. In *17th International Conference on Scientometrics and Informetrics, ISSI 2019 - Proceedings*, volume 2, pages 2133–2144.

Eckmann, M., Rocha, A., and Wainer, J. (2012). Relationship between high-quality journals and conferences in computer vision. *Scientometrics*, 90(2):617–630. cited By 18.

Freyne, J., Coyle, L., Smyth, B., and Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*, 53(11):124–132. cited By 77.

Glänzel, W., Schlemmer, B., Schubert, A., and Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, 68(3):457–473. cited By 51.

Goel, U., Izurieta, C., and Wittie, M. P. (2015). Understanding factors influencing the citation count of networking conference papers. In *24th International Conference on Software Engineering and Data Engineering, SEDE*, pages 12–14.

He, L. and Han, Z. (2017). Do usage counts of scientific data make sense? an investigation of the dryad repository. *Library Hi Tech*, 35(2):332–342.

Laender, A. H., de Lucena, C. J., Maldonado, J. C., de Souza e Silva, E., and Ziviani, N. (2008). Assessing the research and education quality of the top brazilian computer science graduate programs. *ACM SIGCSE Bulletin*, 40(2):135–145.

Lee, D. (2019). Predictive power of conference-related factors on citation rates of conference papers. *Scientometrics*, 118(1):281–304. cited By 9.

Lee, D. (2020). Author-related factors predicting citation counts of conference papers: focusing on computer and information science. *Electronic Library*, 38(3):463–476. cited By 0.

Li, X., Rong, W., Shi, H., Tang, J., and Xiong, Z. (2018). The impact of conference ranking systems in computer science: a comparative regression analysis. *Scientometrics*, 116(2):879–907. cited By 7.

Lisée, C., Larivière, V., and 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. cited By 91.

Loizides, O.-S. and Koutsakis, P. (2017). On evaluating the quality of a computer science/computer engineering conference. *Journal of Informetrics*, 11(2):541–552. cited By 4.

Meho, L. I. (2019). Using scopus's citescore for assessing the quality of computer science conferences. *Journal of Informetrics*, 13(1):419–433.

Scopus (2020). Content Coverage Guide. Available at `https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf`.

Shotton, D. (2013). Publishing: open citations. *Nature News*, 502(7471):295.

Singh, M., Chakraborty, T., Mukherjee, A., and Goyal, P. (2016). Is this conference a top-tier? confassist: An assistive conflict resolution framework for conference categorization. *Journal of Informetrics*, 10(4):1005–1022.

Terzi, M. C., Sakas, D. P., and Seimenis, I. (2013). International events: the impact of the conference location. *Procedia-Social and Behavioral Sciences*, 73:363–372.

Vardi, M. (2010). Revisiting the publication culture in computing research. *Communications of the ACM*, 53(3):5. cited By 17.

Visser, M., van Eck, N. J., and Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1):20–41.

Vrettas, G. and Sanderson, M. (2015). Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology*, 66(12):2674–2684. cited By 33.

Yang, S. and Qi, F. (2020). Do proceedings papers in science fields have higher impacts than those in the field of social science and humanities? *Library Hi Tech*, 39(1):284–307. cited By 1.

Yoo, J. J.-E. and Chon, K. (2008). Factors affecting convention participation decision-making: Developing a measurement scale. *Journal of Travel Research*, 47(1):113–122.

Zhu, Y., Yan, E., Peroni, S., and Che, C. (2020). Nine million book items and eleven million citations: a study of book-based scholarly communication using opencitations. *Scientometrics*, 122(2):1097–1112. cited By 2.

Work	N. conf. series	N. papers	Time Span
Goel et al. (SEDE, 2015)	63	38,895	5 years (2008-2012)
Lee (Scientometrics, 2019)	81	43,463	4 years (2009-2012)
Lee (Electronic Library, 2020)	28	12,337	4 years (2009-2012)
This work	3,838	2,581,564	33 years (1985-2017)

Table I: Comparison of the related work with our study.

authors	**title**	**crossref**	**url**	**year**	**cit**
Mark Harman	Searching for better configurations	conf/sigsoft/2013	db/conf/cdc/cdc2016	2013	25
...	...	...	...	...	...

Table II: Example of raw data fields

conf	**totCit**	**avgCit**	**nPubs**	**nEditions**
eurocrypt	52855	40.347	1310	36
ismb	10875	35.656	305	5
...	...	...	...	...

Table III: Citation data: high-level conference aggregate statistics

conf	**...**	**2014**	**2015**	**2016**	**2017**
ACISicis	...	53	79	148	89
ACMace	...	205	94	55	35
...	...	...	...	...	...

Table IV: Citation data: temporal evolution of conference total citations per year

conf/year	**city**	**country**	**state**
cscw/2000	Philadelphia	USA	PA
mobicom/2000	Boston	USA	MA
eurocrypt/2005	Aarhus	Denmark	-
...	...	...	...

Table V: Tourism data: conference location information

city	totCit	avgCit	nEditions	SWP	TA
Beijing	61913	168.70	368	179383	4
Paris	62430	190.34	328	218583	17.56
San Francisco	82004	274.26	299	177857	2.9
...	...	...	...	...	...

Table VI: Citation/tourism data: aggregate citation and tourism information for each city

country	totCit	avgCit	nEditions	TTCI	TSI	CRBT	GPI	SI
USA	1539133	249.78	6167	5.3	6.6	4.7	4.9	21.5
China	223458	130.45	1714	4.9	3.5	7	3.9	19.3
Germany	168601	124.61	1356	5.4	5.9	6.5	5.7	23.5
...	...	...	**...**	...	...	...	...	...

Table VII: Citation/tourism data: aggregate citation and tourism information for each country