Understanding Factors Influencing the Citation Count of Networking Conference Papers

Utkarsh Goel, Clemente Izurieta, and Mike P. Wittie
Department of Computer Science, Montana State University, Bozeman MT USA 59717
{utkarsh.goel, clemente.izurieta, mwittie}@cs.montana.edu

Abstract

The impact of a journal on its community is usually judged based on its impact factor. Researchers often aim to publish their research findings at journals with high impact factors. Conversely, researchers tend to judge the quality of a conference based on its acceptance rate. However, it remains unknown if the low acceptance rate of a conference reflects a high impact on its community compared to a conference with high acceptance rate. We performed an empirical study to analyze the influence of conference rankings on the number of citations a paper may receive over time. Our results show that papers published at highly reputed conferences tend to have a higher number of citations on average, than low ranked conferences. Additionally, we found that the title of the paper also influences the citation count. We believe our work would bring forth the opportunities for authors of working papers to increase the number of citations of their papers as well increase the impact of their research towards the growth of their respective communities.

keywords: Citation count, conference rankings, paper title, publication year, impact factor.

1 Introduction

The networking research community has been publishing a growing number of articles in computer networking journals and conferences over the last several decades. To ensure that the research gets published in high quality journals and that the research has a significant influence on the growth of the society, researchers judge the quality of different journals based on their impact factors and tend to publish articles in journals with either high impact factors [12, 18, 22] or in journals where acceptance of an article is challenging, relative to other journals in the same field [10]. However, conferences do not have impact factors and therefore researchers tend to judge the quality of a conference based on the conference's acceptance rate. We argue that the acceptance rate of a conference may not reflect the impact that papers have on the growth of their communities, similar to journals' impact factors. Therefore, it is important for researchers to understand whether publishing research papers in conferences with low acceptance rates would result in higher impact on their communities than publishing in conferences with low acceptance rates.

We performed an empirical study on a large data set to explore potential relationships between several factors (for example, the reputation of the conference where the paper is accepted, the title of the paper, and the publication year of the paper) and the number of citations that a paper may receive over time. Our current data set consists of information from 63 networking conferences and about 39000 conference papers published in these conferences during the years 2008 through 2012. We summarize the major findings of this study as follows:

- The reputation of the conference in which a paper is published may statistically influence the citation count of that paper by as much as 13%
- The choice of keywords in the paper title may statistically influence the number of citations that the paper may receive over time by as much as 9.3%
- The publication year may statistically influence the citation count of the paper by as much as 9.3%, and
- The interactions between conference reputation, paper title, and publication year may statistically influence the citation count by as much as 45.7%

Our conclusions may assist authors in increasing the number of citations their papers receive over time. Specifically, we hope that by using attractive keywords in the paper title, submitting to conferences with high reputation, and publishing research contributions at the appropriate time would potentially increase the citation count, which may also reflect an increase in the growth of the research. For experimental replication purposes, we make our data set and code publicly available at `https://github.com/ugoel/ESOF522`.

This paper is organized as follows. In Section 2, we describe our experimental setup to collect data from different conferences and papers published from online conference ranking databases and Google Scholar. Section 3 describes our evaluation results. In Section 4, we offer a discussion of our approach for evaluating the impact of different factors on citation counts, as well as, list a few possible threats to validity of our results. In Section 5, we outline related work, and we conclude in Section 6, with research directions for our future work.

2 Experimental Setup

Throughout our study, we consider the networking conferences listed at Prof. Kevin Almeroth's blog as the conferences that contribute the most in the field of computer networking [14]. For every conference listed in the blog, we collect the average number of papers submitted, average number of papers accepted, and the average acceptance rate for conferences conducted from 2008 to 2012. Our communication with Prof. Kevin Almeroth at the University of California Santa Barbara has revealed that the statistics of different networking conferences maintained by Prof. Kevin Almeroth are the most accurate, since general chairs of different networking conferences provide the conference statistics to Prof. Kevin Almeroth after the conference concludes. Documentation of various statistics of the networking conferences collected for our study are readily available [19].

2.1 Classifying Conference Reputation

The goal for this empirical study is to identify relevant factors that may affect the citation count of conference papers published in networking conferences. We argue that the reputation of a conference to which the paper is submitted could be one of the major factors that impact the citation count of a paper. Agreement about the reputations of different conferences is the first challenge. Several online conference ranking portals may vary in the way they classify reputations of different conferences [8, 9, 15, 16]. Further, the research community's opinion about a conference reputation generally differs from what is available in online conference portals [21]. Therefore, it is important to characterize the rankings given to different conferences by different online conference portals, as well as, by the research community.

To understand the classifications used by different conference ranking portals (Rank A, B, or C), we collected conference rankings, for each conference in our dataset, from the CORE Conference Portal (CCP), the ERA Conference Portal (ECP), and from the research community [1, 3, 21]. We collected conference rankings for years 2008, 2013, and 2014 from CCP, and for year 2010 from ECP.[1] To understand the research community's opinion on conference reputation, we conducted a survey to ask researchers and professors at different industries and institutions for their opinion on conference reputations in terms of ranks such as A, B, C, or no ranking [20]. We distributed the survey Web URL to 755 members (as of March 2015) of the Technical Committee on Computer Communications (TCCC) research community [13]. We configured the Google survey to randomize the order of conference names to prevent responders from guessing the conference rankings based on their position on the survey. We received a total of 47 responses from TCCC members indicating their opinions on reputations of different conferences. We document the responses of the TCCC community members and their positions and affiliations, in the order in which they were received [21]. Further, most of the responses we received were from researchers representing different positions at different industry research labs or academic institutions. Although we received responses from 47 people, only 34 opted to provide their titles and affiliations.

Table 1: Breakdown of different conferences into rankings

Acceptance Rate Scale (%)	Conference Reputation	No. of Conf.	No. of Papers
11 to 20.9	High (A)	22	8224
21 to 30.9	Medium (B)	22	7134
31 to 40.9	Low (C)	19	23537
	Total	**63**	**38895**

Based on the data collected from CCP, ECP, and the survey, we classify (only for the purpose of this study) the conferences in different rankings based on their acceptance rates. We argue that deciding the reputation of a conference by its acceptance rate is ideal, since to maintain reputation, a highly reputed conference would accept lesser number of papers than a medium reputed conference. Therefore, we classify and show the conference rankings based on their acceptance rates in Table 1 and consider such classification as the `Oracle` of rankings throughout our study.

2.2 Collecting Citation Counts

We posit that factors such as conference rank, publication year, and the paper title may influence the the number of citations that the paper receives over time. Specifically, we argue that authors of working

[1]Neither CCP or ECP provides public access to conference ranking database for years 2009, 2011, and 2012.

papers may consider the quality of a conference where the reference is published, among the factors that could increase the strength of their publication. Authors may also consider the publication year of their reference as it may reflect the timeliness of the facts illustrated by the reference. Finally, authors may also consider finding references on Web Search tools (such as Bing [2], Google [5], and Yahoo! [6]) relevant to their manuscript. As a result, the choice of keywords used in the paper title may affect whether the paper is discovered by a search tool, which may eventually affect whether the paper gets cited.

Therefore, to understand the significance of factors affecting the citation count of networking conference papers, we used Google Scholar as a publicly available online repository to collect information about different research articles published in various computer networking conferences [4]. Specifically, for each paper published at a given networking conference, we collected the paper title, the year when the paper was published, number of times the paper was cited (as of February 2015), and the conference name to which the paper was accepted.

2.3 Challenges in Collecting Data from Google Scholar

Although Google Scholar is one of the most popular sources of getting the most up-to-date data for research articles, as of March 2015 the Google Scholar portal lacked a developer API to assist researchers in collecting data. Therefore, we used an early implementation of Christian Kreibich's open source code as a template to collect publication data from the Google Scholar portal [7]. However, we encountered several challenges during our data collection process from Google Scholar.

Hard limit on the number of Search queries: In order to protect server resources and to ensure high availability of Scholar content, Google prevents automated queries from being processed by configuring a limit on the number of queries that can be sent from a client device in a given day. Further, the maximum number of search results that Google Scholar provides for every query is limited to 20. We sent a total of 1998 successful HTTP GET requests to Google servers to collect data for papers published in 63 conferences held over a period of five years.

The data collection process took approximately two weeks, since the requests sent by our script were frequently identified as a potential abuse to Google servers; which led to zero search results returned by Google servers. Therefore, to speed up the process of data collection and get appropriate responses, we distributed our data collection script on several computers in different IP networks. Our immediate goal was to minimize the number of automated requests sent by a single IP address. In some cases, we spoofed HTTP headers in our script to include a valid cookie generated by Google Scholar servers on a Mozilla Firefox browser. Our immediate goal here was to have the script pretend as if the requests sent by it were generated by the regular use of a Web browser.

Choice of search keyword: Although we filtered the Google Scholar search results based on the conference name, Google Scholar oftentimes either did not provide information for all the papers published at the queried conference or included information for papers published at conferences similar to the queried conference name. Further, for a few of the conferences considered in our study, Google Scholar included the information for workshop and poster papers published at the conference in the search results, which increased the total number of queries going to Google servers and alleviated the frequency of our search requests being denied by Google.

To eliminate the risk of Google Scholar returning search results for different conferences and/or papers published as workshop/poster papers, we manually refined the choice of our search keywords until we received information for papers published at the queried conference. Further, for each conference, we also manually validated the paper information given by Google Scholar with the paper information available at the conference official website.

3 Results

We show that the conference name, publication year, and the choice of keywords in the paper title have a significant effect on the number of citations that a conference paper receives over time. However, before reporting how much impact these factors have on citation count, we discuss different online conference ranking databases and community opinion.

3.1 Variation in Conference Rankings among Community Opinion and Online Databases

We compare the conference reputations (rankings) as classified by our `Oracle` [2] , CCP, ECP, and the community opinion survey in Figure 1. Specifically, we show the variation in conference rankings between `Oracle` and CCP, `Oracle` and ECP, and `Oracle` and

[2]We refer Oracle as a conference-to-rank mapping database that ranks conferences as per Table 1.

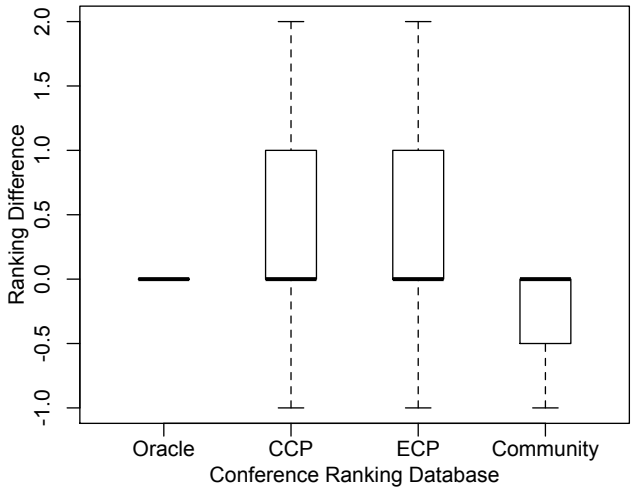

Figure 1: Difference in the conference rankings among the `Oracle` and online databases/community opinion survey.

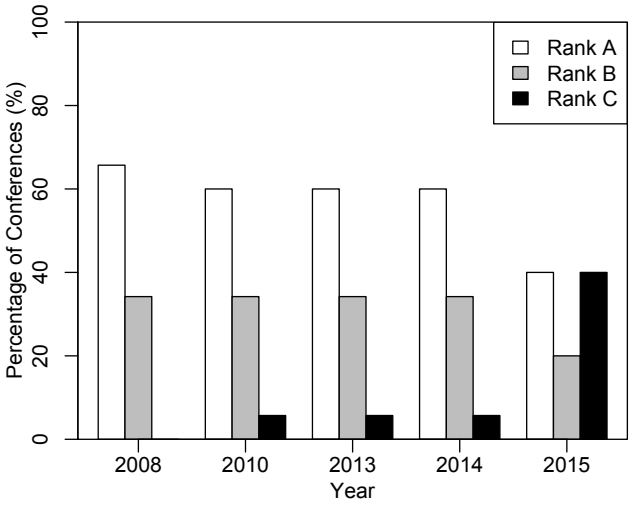

Figure 2: Distribution of different networking conference rankings over a five year period.

community opinion survey. For representing the variation, we translate the reputation for each conference from ranks such as A, B, or C into numerical values such as 1.0, 2.0, and 3.0 respectively. The y-axis in Figure 1 shows the distribution of differences in conference rankings between `Oracle` and other tools. A positive value on the y-axis indicates that a given tool considered a conference to have a higher reputation than the `Oracle`. A negative value on the y-axis indicates that a given tool considered a conference to have a lower reputation than the `Oracle`. The x-axis shows different tools that provide conference rankings. We show that the rankings provided by both CCP and community opinion differ from the `Oracle`. Specifically, for 75 percent of conferences, both CCP and ECP classifies conferences to have higher reputation than the `Oracle`. However, for the same 75 percent of conferences, the community opinion is the same as that of `Oracle`. We therefore argue that CCP and ECP's approach to classifying conferences may not be accurate and may not indicate the true reputation of different networking conferences. We further argue that the research community has a good understanding of the reputations of different networking conferences, most likely because the research community is actively involved in the conference program/steering committee and also publishes research ideas at different conferences.

Next, we investigate whether or not the online databases update their rankings for different conferences over time. In Figure 2, we show the number of conferences (in percentage) that were classified as rank A, B, or C over the period of 5 years. For the years 2008, 2013, and 2014, we collected conference rankings from the CCP tool [3]. For the year 2010, we collected the conference rankings from the ECP tool, available at the CCP website [1, 3]. Finally, for the year 2015, we consider the conference rankings provided by our community opinion survey [21]. We observe that the conference rankings over the year 2008 and 2010 did not change significantly. Specifically, classifications for only about 2% of the conferences that were classified as rank A and rank C changed from 2008 to 2010. Further, the rankings for different conferences in years 2010 to 2014 did not change. And finally, we show that in 2015, the number of conferences that were classified as rank A and rank C changed by about 20% and that the conferences classified as rank B changed by about 10%.

3.2 Impact of Conference Reputation on Citation Count

We investigate the impact that publishing papers at conferences of different reputation have on the citation count. In Figure 3, we show the average citations received (as of March 2015) by papers published at conferences with different acceptance rates.[3] In general, we can observe that the number of citations that a conference paper may receive over time decreases as the acceptance rate of the conference increases. Specifically, the average drop in number of citations for papers published in conferences ranked as A and B is about 40. However, we see an exception to that conclusion, in that, the number of citations received by papers published in conferences with acceptance rates between

[3]Refer Table 1 for translating conference acceptance rates in to rankings.

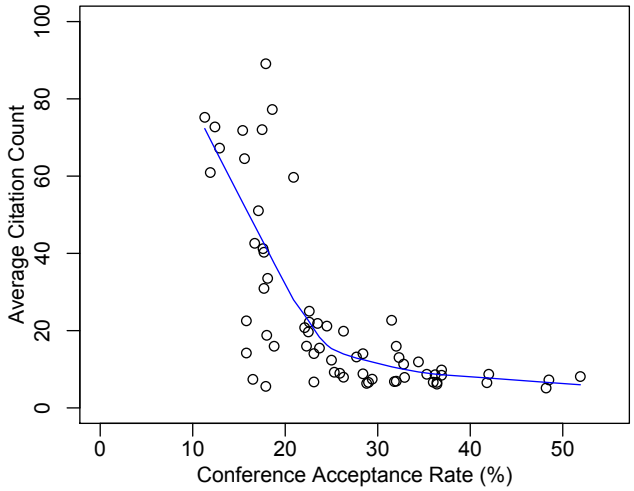

Figure 3: Distribution of number of average citations with respect to conference acceptance rate.

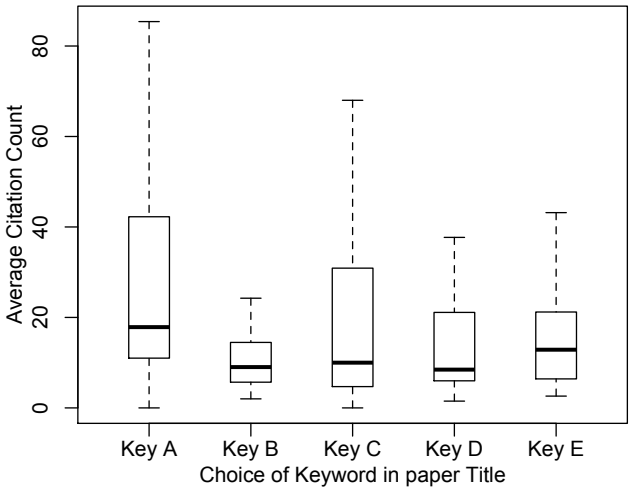

Figure 4: Distribution of citation count with respect to keywords used in the paper title.

16% to 25% are very close. Therefore, we posit that some papers published at conferences that we classify as rank A (such as ACM SenSys, ACM Sigmetrics, IEEE ICNP, ACM ACNS, IEEE ICDCS, IEEE HiPC, and EWSN) may receive the same number of citations as the conferences that we classify as rank B (such as P2P, IEEE SECON, CloudCom, ACM DEBS, IEEE/ACM IWQoS, ICPP, ANCS, MSWIM, WoWMoM, IEEE SASO, WiMOB, IEEE ICCCN, IPCCC, ACM Mobiquitous, and IEEE SAINT). Finally, we show that papers published in conferences with rank B and C receive relatively same number of citations over time.

3.3 Impact of Paper Title on Citation Count

We investigate the impact of paper titles on the number of citations that a paper may receive over time. For this part of our study, we consider five different keywords (also shown in Table 2) popular in the field of computer networking. Our selection of these five keywords is based on the popularity of these keywords over 2008 until 2012, according to Google Trends service [11]. While we ensured that the keywords used in our study had significantly high popularity during 2008 to 2012, we also ensured that the popularity of the keywords we selected did not increase or decrease significantly over the period of 2008 to 2012. Our immediate goal here was to mitigate the effect of keyword popularity rise and fall on the paper citation count.

Figure 4 shows the distribution of the number of average citations that conference papers receive with different keywords in their titles. On average, the choice of keywords in the paper title has a significant

Table 2: List of keywords used in our study

Keyword Code	Keyword in Paper Title
Key A	Cloud
Key B	Distributed
Key C	Security
Key D	Mobile
Key E	Performance

impact on the citation count. Specifically, for 75 percent of the papers with the keyword `Cloud` (Key A)' in their paper titles, the number of average citations received are much higher than received by papers with keyword `Distributed` (Key B), though both the keywords reflect similar technologies to some extent. Further, keywords such as `Mobile` (Key D) and `Performance` (Key E) mostly represent technologies which may be independent of each other. Therefore, for 75 percent of papers, we do not see much difference between the number of citations received by papers with either of those keywords in the title.

3.4 Impact of Publication Year on Citation Count

We investigate the impact of the publication year on the number of citations that a conference paper may receive over time. For this part of the study, we compare the number of citations received by papers published in years 2008, 2009, 2010, 2011, and 2012. In Figure 5, we show the distribution of the number of average citations received by papers published in different conference years. In general, we see that for 75 percent

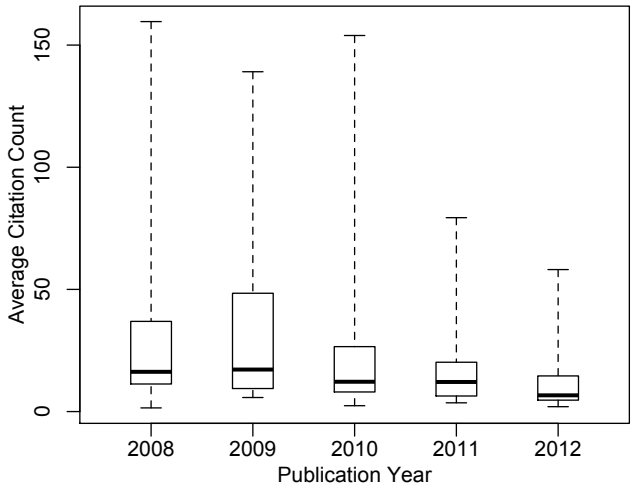

Figure 5: Distribution of citation count with respect to publication year.

of papers, more recent published papers have lower number of citations than papers published in older years. Specifically, the number of citations received by papers published in 2008 are higher than papers published in 2010, 2011, and 2012. Similarly, the number of citations received by papers published in 2010 are higher than papers published in 2011, and 2012. However, there is one exception in our data, that is, the average number of citations for papers published in year 2009 is higher than papers published in year 2008. We believe that a number of papers published in 2009 made significant research contributions, because those papers received higher citations than papers published in 2008.

3.5 Statistical Analysis

We showed that conference reputation, paper title, and the publication year impact the citation counts of conference papers. However, it is still unclear whether the impact of these factors on citation count is statistically significant. Therefore, in Table 3, 4, 5, we represent average citation counts as response variable, and publication year and keywords (in paper titles) as three factors for conferences with ranks A, B, and C, respectively. Our ANOVA analysis from the data represented in these tables shows that the conference ranking, choice of keyword in the paper title, and publication year impact the citation counts of conference papers by as much as 13%, 9%, and 9% respectively. We also observe that interactions among the three factors impact the citation count by as much as 23%.

Table 3: A distribution of citation counts (response variable) for conferences with Rank A, where publication year (2008 through 2012) and keywords (A, B, C, D, E) are considered as factors.

	Publication Year				
Keyword	**2008**	**2009**	**2010**	**2011**	**2012**
Key A	494.2	297.2	155.5	48.8	27.6
Key B	57.3	43.3	50.9	21.5	17.6
Key C	31	29.3	39.7	15.6	15.5
Key D	42.8	61.6	67.3	38.3	15.7
Key E	67.4	75	46.9	25.6	19.4

Table 4: A distribution of citation counts (response variable) for conferences with Rank B, where publication year (2008 through 2012) and keywords (A, B, C, D, E) are considered as factors.

	Publication Year				
Keyword	**2008**	**2009**	**2010**	**2011**	**2012**
Key A	95.5	106.3	30.1	18.4	9.5
Key B	15.4	16.4	19.8	15.9	5.8
Key C	12.1	14.2	9.1	6.8	5.6
Key D	4.7	45.9	27.3	6.3	4.7
Key E	12.8	10.2	12.8	6	5.7

Table 5: A distribution of citation counts (response variable) for conferences with Rank C, where publication year (2008 through 2012) and keywords (A, B, C, D, E) are considered as factors.

	Publication Year				
Keyword	**2008**	**2009**	**2010**	**2011**	**2012**
Key A	75.2	24.7	26.5	16.4	7.8
Key B	10.5	12.1	7.8	7	4.5
Key C	14.5	9.8	7.2	5.8	3.7
Key D	15.2	9.8	9.5	6	3.5
Key E	9.6	7.7	8.2	5.5	4.3

4 Discussion and Threats to Validity

Although, our study did not explore the impact of all possible factors on the citation count of conference papers, we argue that conference reputation, paper title, and publication year are the most important factors deciding citation counts for conference papers. However, we acknowledge the fact that there may be other factors for which we could not collect data that may also have an impact on the citation count. For example, the popularity of the author of the published paper, awareness that a relevant paper was published, relevance of the paper to the published articles, timeliness of the published paper, the citation source of the published paper, and public access to published papers through websites such as IEEE Xplore or ACM Digital Libraries.

We also acknowledge possible internal threats to validity. There may be a few factors that may have influenced the ranking of a conference as represented by the survey we conducted. Specifically, the awareness of anonymous survey responder about different conference fields and reputation. For example, a responder with expertise in Internet measurements may not have expertise and not be aware of conferences and their reputations in the field of computer security. Such a responder may not have provided a ranking or may have selected the 'No Ranking' option in our survey for conferences outside their field of expertise. To deal with such cases, for each conference in our survey, we use the ranking that received the highest votes from different responders, as the community opinion for that conference. In conferences where we observed a tie between one or more rankings, we ensured that we excluded such conferences from our analysis (Section 3.1).

This study is limited to Networking conferences, and results cannot be generalized to other Computer Science fields. This poses an external threat to the validity of this work.

5 Related Work

A number of empirical studies have been conducted to understand the factors that may affect the impact of journals in different fields [10, 12, 18]. However, to the best of our knowledge, our work is the first large scale study to investigate the factors affecting citation counts of papers accepted to Networking conference papers. Closest to our study work performed by Beverly *et al.* The authors discuss the factors that influence the acceptance of a paper submitted to a conference and the number of citations that a paper may receive [17]. However, this work is limited to papers submitted and accepted to only one networking conference (that we classify as Rank A) in the year 2010. Our work, improves on their work and provides insights for a large number of conferences in the field computer networking.

6 Conclusions and Future Work

The research community has been publishing a number of research contributions in computer network related conferences for several years. However, the impact those papers have on the research community has not been well understood. We conducted a large scale empirical study to find potential relationships between several factors (the reputation of conferences where the papers are accepted, the paper title, and the publication year of the paper) and the number of citations that a paper may receive over time. We showed that the reputation of the conference, the paper title, and the publication year may impact the citation count of papers by as much as 13%, 9%, 9%, respectively. We also show that the interactions among these factors may influence the citation count by as much as 23%. Finally, we acknowledge that there are several unknown factors (not considered as a part of our empirical study) that may influence the citation count of conference papers by as much as 46%.

As a part of our future work, we would like to develop predictive mechanisms that authors may use to identify potential conferences and paper titles that may lead to high citation counts.

Acknowledgements

The authors would like to thank Kevin Almeroth for sharing statistics on network conferences, Christian Kreibich for developing a sample code for fetching data from Google Scholar website, and the anonymous TCCC members for completing our survey on conference ranking. The authors would also like to thank Robert Beverly, Ajay Miyapuram, Anburaj Muthumani, Kanika Shah, and Sandhya Thangavel for suggested improvements to our data analysis techniques and to an early version of this manuscript.

References

[1] ERA Conference Rankings Access. `http://lamp.infosys.deakin.edu.au/era/?page=cmain`, Feb. 2010.

[2] Bing Web Search. https://www.bing.com/, Feb. 2015.

[3] CORE Conference Rankings. http://www.core.edu.au/index.php/conference-rankings, Mar. 2015.

[4] Google Scholar. https://scholar.google.com/, Apr. 2015.

[5] Google Web Search. https://www.google.com/, Feb. 2015.

[6] Yahoo Search - Web Search. https://search.yahoo.com/, Feb. 2015.

[7] Christian Kreibich. A parser for Google Scholar, written in Python. https://github.com/ckreibich/scholar.py, Feb. 2015.

[8] Core Conference Ranking Portal. Computer Science Conference Rankings Descriptions. http://www.core.edu.au/documents/RankingDescriptions2014.pdf, Apr. 2015.

[9] Database Systems Lab. Computer Science Conference Rankings. http://dsl.serc.iisc.ernet.in/publications/CS_ConfRank.htm, Apr. 2015.

[10] Eugene Garfield. Journal impact factor: a brief review. *Canadian Medical Association Journal*, 161(8):979–980, 1999.

[11] Google Trends. In 2014 we searched trillions of times. What do these searches say about us? https://www.google.com/trends/, Mar. 2010.

[12] Paul J Haensly, Paul E Hodges, and Shirley A Davenport. Acceptance rates and journal quality: An analysis of journals in economics and finance. *Journal of Business & Finance Librarianship*, 14(1):2–31, 2008.

[13] IEEE Communications Society. Technical Committee on Computer Communications (TCCC). http://committees.comsoc.org/tccc/, Feb. 2015.

[14] Kevin Almeroth. Networking Conferences Statistics. https://www.cs.ucsb.edu/~almeroth/conf/stats/, Mar. 2015.

[15] Nanyang Technological University. Computer Science Conference Rankings. http://www3.ntu.edu.sg/home/assourav/crank.htm, Apr. 2015.

[16] Osmar R. Zaane. Computer Science Conference Rankings. https://webdocs.cs.ualberta.ca/~zaiane/htmldocs/ConfRanking.html, Oct. 2011.

[17] Robert Beverly and Mark Allman. Findings and Implications from Data Mining the IMC Review Process. *ACM SIGCOMM Computer Communication Review*, 43(1), January 2013.

[18] Somnath Saha, Sanjay Saint, and Dimitri A Christakis. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*, 91(1):42, 2003.

[19] Utkarsh Goel. Networking Conference Rankings : (2008 to 2012). http://goo.gl/hXVEQo, Mar. 2015.

[20] Utkarsh Goel and Mike P. Wittie. Network Conference Ranking Survey Questionnaire. https://docs.google.com/forms/d/1ZxfFYl8NAFgCtOqx5VCTqD-l_GerzWmbg8LDEnP3cLw/viewform?c=0&w=1, Mar. 2015.

[21] Utkarsh Goel and Mike P. Wittie. Network Conference Ranking Survey Responses. https://docs.google.com/forms/d/1ZxfFYl8NAFgCtOqx5VCTqD-l_GerzWmbg8LDEnP3cLw/viewanalytics, Mar. 2015.

[22] Web of Science. The Thomson Reuters Impact Factor. http://wokinfo.com/essays/impact-factor/, Mar. 2015.