Gatekeeper: Quantifying the Impacts of Service to the Scientific Community

Spyke Krepshaw and Dongwon Lee

The Pennsylvania State University, PA, USA
{spyke,dongwon}@psu.edu

Abstract. Academic scholars have several duties, including teaching, research, and service to the community and society. While a scholar's research impacts can be reasonably measured and tracked via citation analysis in existing digital libraries, to our best knowledge, there has been no system that systematically collects and quantifies a scholar's impacts of *service* to the scientific community. In particular, we are interested in measuring scholars' impacts as "gatekeepers," who play a key role in the spread of research findings and new knowledge via the accept/reject decisions of research articles. In this work, toward this goal, we present a prototype digital library, Gatekeeper, that crawls, extracts, and quantifies the impacts of service based on one's roles in the technical program committees of Computer Science conferences.

Keywords: Service Impact · Citation Analysis · h-index

1 Introduction

Being able to model and quantify the impacts of a scholars service to the society has many utilities in applications–e.g., hiring and promoting scholars, or finding experts for service based committees. Yet, it is inherently subjective and ambiguous to quantify the impacts of one's service. Unlike quantifying one's research impacts that has been well studied and implemented by means of citation analysis, the term "service" itself has broad interpretations with discipline-specific definitions and examples. For instance, a scholar's service may include diverse activities such as participating in conference organization/technical committees, serving in editorial boards of journals, delivering talks/keynotes in events, reviewing books, serving in funding related panels, or interviewing with press and media. To our best knowledge, there is currently no digital library that collects scholars' service related activities and quantifies the impacts of service. In this work, therefore, we aim to address this gap and present the prototype digital library, named as Gatekeeper.

As the initial attempt, we first focus on one type of scholars' service activity–i.e., serving in the technical program committees (TPC) of Computer Science (CS) conferences for several reasons: (1) As well noted, CS is a unique discipline where conferences play a major role in disseminating significant findings. There is a well maintained digital library such as DBLP that shows a comprehensive list of CS conferences and their past websites; (2) Often, CS conference websites list detailed membership information of organization and program committees, along with their full names and identifying information (e.g., email or homepage) often available for easy extraction; and (3) As TPC members critically review research articles and contribute to the decision of accepting/rejecting the articles, thereby acting as a "gatekeeper" of new knowledge into the scientific community, collecting and quantifying their impacts of service is critically important.

Even with the focus on the service activities only in TPC, there are several challenges to address, including: (1) how to extract and differentiate different roles in TPC (e.g., program chair vs. area chair vs. senior program committee vs. program committee); (2) how to contrast the impacts of service across conferences or sub-disciplines (e.g., TPC in TPDL vs. TPC in AAAI); (3) how to factor in the size of TPC (e.g., TPC of 20 vs. TPC of 200); and (4) how to deal with the quality of conference in quantifying the impacts of service (e.g., TPC in a top CS venue vs. TPC in an obscure venue). Note that answers to all these challenges may vary as they are subjective in nature. In this paper, therefore, we present an approach that we took and a prototype that demonstrates the proof-of-concept.

2 Related Work

There are few works on quantifying the impacts of service, but abundant works on evaluating the impacts of research or scholars thereof.

2.1 Quantifying the Impacts of Research

Although journal articles are normally viewed more significant than conference articles in many disciplines, a culture in CS is radically different such that majority of major findings are reported in the form of conference articles. With such an abundance of articles being submitted to conferences, it has naturally become increasingly important to be able to discern the quality of conferences [11]. However, being able to understand the importance of a particular conference is not always immediately available. In the following, we review several representative methods to rate the quality of venues.

The (Journal) Impact Factor (IF) [6] was originally intended to assist in the selection of journals. It is calculated by taking the number of citations received in a specific year, adding the number of articles published in that journal during the two preceding years, and dividing by the total number of articles published in that journal during the two preceding years [7]. It is important to note that comparing IF scores between journals across disciplines is in general not useful. The idea of IF can be equally applied to conferences. In general, despite some pitfalls, IF scores are still one of the most popular indicators to assess the qualities of venues [10]. Applying the idea of PageRank [2] to a citation graph, one can infer impactful articles such that articles with more incoming citations, especially those from other impactful articles are viewed more impactful [3].

Similarly, other popular website ranking algorithms such as HITS [9] can be also applied to a citation graph, yielding hubs and authorities of research articles. When applied to a set of conferences and TPC list of those conferences, methods such as HITS can naturally identify important conferences and TPC members.

Other lesser-known research into ranking conferences includes ranking by way of TPC characteristics [13]. The idea of measuring TPC characteristics looks at the number of members, number of publications by members, average number of authors and so forth. Using this method, results have shown a high accuracy rate of classifying conferences. Unlike the previously mentioned methods, this model of ranking can be achieved without citation-based analysis. Conceptually related to PageRank, [12] takes an approach of *browser-based measure*, which takes the reader's behavior into consideration. This new method determines quality based on how a reader might "jump" from paper to paper. Such jumps could occur by looking for other papers with the same author(s), or finding a paper which was cited int he original.

2.2 Quantifying the Impacts of Scholars

One of the most popular methods to quantify the research impacts of a scholar is the h-index [8] that measures both the productivity (i.e., how many articles) and impact (i.e., how many citations) of one's research articles. In order words, a scholar receives an h-index score of h if she has published at least h articles have been cited at least h times. While capturing both the productivity and impact of one's research well, the h-index method fails to recognize scholars who have made seminar findings with a small number of publications as they will have a low number of h. To improve on this shortcoming, the g-index method [4] further modifies the h-index such that a scholar receives a g-index score of g if she has published at least g articles that have been cited "collectively" at least g^2 times. This change has the effect of allowing highly-cited articles to effectively assist the low-cited articles in the calculation.

3 Quantifying the Impacts of Service

To quantify one's impacts of service, we apply both h-index and g-index ideas to our context and propose several "Gatekeeper"-index methods. Just like both h-index and g-index attempted to capture both productivity and impact of research concurrently, we aim to capture both productivity (i.e., how many TPC a scholar has served) and impact (i.e., how good a conference is) of service. Therefore, a scholar A whose has served in 100 TPCs is deemed to have a higher service impact than a scholar B who has served in 10 TPCs. Similarly, a scholar C who has served in 10 TPCs of top CS venues may be deemed to have a higher service impact than a scholar D who has served in 15 TPCs of obscure venues.

Further, to capture how good a conference is, we leverage on the well-studied citation analysis. Figure 1 illustrates a power-law like distribution (in a log-log plot) between the number of citations and the rank of conferences. That is, a

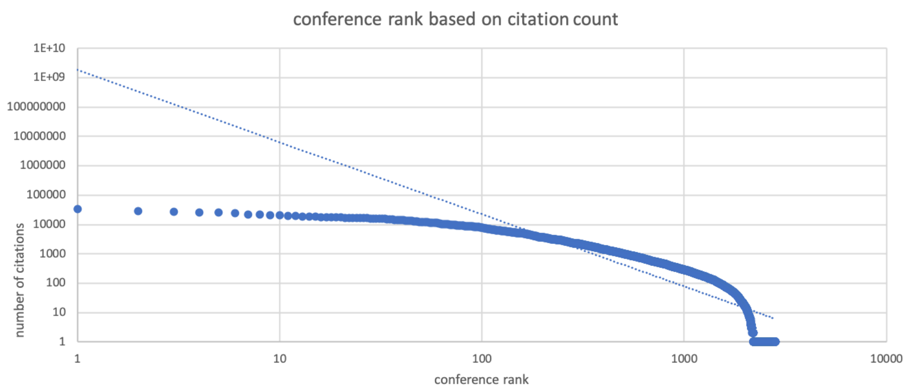

Fig. 1: A power-law like distribution of conference citations.

small fraction of conferences accrue a large number of citations while a lot of conferences receive only a small number of citations. Using a total number of citations per conference as a measure of the goodness of a conference, then, we propose our first Gatekeeper-index as follows:

Definition 1 (G_1-index). *A scholar has the G_1-index score of N if she has served in N conferences as TPC and each conference has accrued at least a total of $f(N)$ citations, where $f()$ is a normalization function.*

For $f()$, for instance, one may simply use $\sqrt{}$ or $\sqrt[3]{}$ to oppress the influence of the large number of the total citation of a conference. In practice, due to a wide range of citation count per conference and delay of data collection, one may have to use a customized normalization function to have more realistic G_1-index scores. Table 1 shows a list of conferences where Bruce Worthman has served in TPC and their corresponding citation counts using the Microsoft Academic Graph (MAG)[1] digital library. Note that citation counts of many recent conferences have not been collected yet and there is a huge number of differences in citations between conferences.

For a scholar to have the G_1-index score of 10, for instance, she has to serve in at least 10 TPCs of conferences, where each conference has received at least a total of $\sqrt{100}$ or $\sqrt[3]{1000}$ citations. The G_1-index is intuitive and relatively easy to compute. However, as it uses the collective number of citations of a conference, it would disproportionately favor a large conference with many articles (thus larger total citations). To address this shortcoming, next, we use an average number of citations per article in a conference, instead of a total number of citations of a conference, to capture the quality of a conference.

[1] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/

Table 1: A listing of conferences served by Bruce Worthman.

Conference Title	Citation Count	Conference Title	Citation Count
WCNC 2019	0	HPSR 2015	70
IM 2019	0	IM 2015	0
VNC 2018	10	INFOCOM 2015	4,150
CNS 2018	0	WCNC 2015	1,811
ICC 2018	268	GLOBECOM 2014	9,725
MobiSec 2018	0	ICC 2014	8,921
GLOBECOM 2017	450	NOMS 2014	1,792
Healthcom 2017	29	ICC 2013	10,040
SECON 2017	50	ICC 2012	13,279
CISS 2017	95	VNC 2011	582
IM 2017	17	PIMRC 2011	2,749
INFOCOM 2017	722	INFOCOM 2011	14,230
WCNC 2017	482	INFOCOM 2010	17,310
GLOBECOM 2016	2,305	IM 2009	1,424
CNS 2016	0	WCNC 2008	5,706
IEEE PIMRC 2016	536	IEEE SECON 2005	0
SECON 2016	194	INFOCOM 2004	28,439
ISPLC 2016	54	INFOCOM 2003	33,133
GLOBECOM 2015	954		

Definition 2 (G_2-index). *A scholar has the G_2-index score of N if she has served in N conferences as TPC and an article of each conference has accrued on average at least N citations.*

Using the G_2-index formula, a scholar who has served in many TPCs, or who has served in impactful conferences, whose articles have high citations on average, is likely to have a high G_2-index score. Next, applying the g-index idea to our context, we propose our third Gatekeeper-index as follows:

Definition 3 (G_3-index). *A scholar has the G_3-index score of N if she has served in TPC of top-N conferences (sorted in descending order of citations) that have collectively received at least N^2 citations: i.e., $\sum_{N \geq i} C_i \geq N^2$, where C_i is the citation count of a conference among top-N conferences.*

Equivalently, G_3-index can be defined as the largest number N of highly cited conferences (whose TPCs a scholar has served) for which the average number of citations is at least N. A scholar who has served in many highly-cited impactful conferences is likely to have a higher G_3-index score.

4 Data Collection

As indicated throughout this document, the need for a model to measure the impact of a gatekeepers service is evident. To accomplish this task, various types

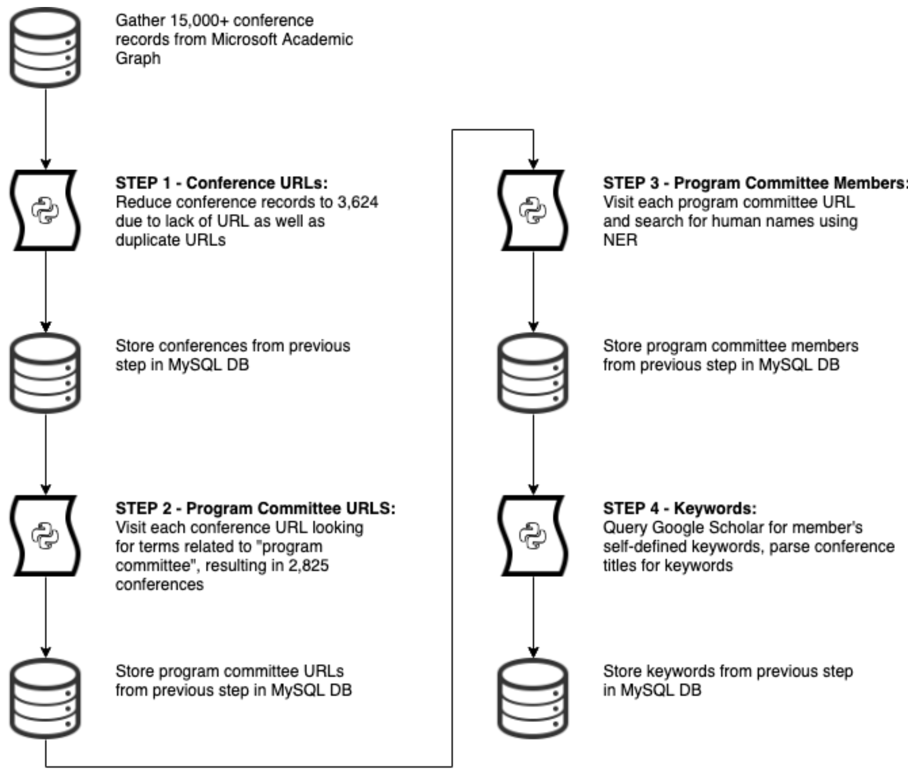

Fig. 2: Data collection workflow block diagram.

of data from several sources are needed. The data which is required is that of conferences, the scholars who serve as the TPC members at the conferences, and their metadata. Unfortunately, widely renowned data resources such as DBLP or Google Scholar do not offer this specific type of data. Therefore, a further solution needs to be sought out. A block diagram of the data collection workflow can be seen in Figure 2.

4.1 STEP 1: Finding Conference Websites

Initially, the data collection will start out by focusing on conferences in Computer Science. A valuable resource with a multitude of conference listings which is frequently updated can be viewed using the Microsoft Academic Graph (MAG). At the time of this writing, the MAG boasts over 15,000 conferences. The following attributes have been retrieved for each conference record from the MAG:

- `ConferenceInstanceId`: Unique MAG identifier for each conference
- `NormalizedName`: Full conference title
- `DisplayName`: Shorthand conference title

- `ConferenceSeriesId`: Determines if conference is part of a series
- `Location`: Geographic location of conference
- `OfficialUrl`: URL of conference website
- `StartDate/EndDate`: Start and end date of conference
- `PaperCount`: Number of papers accepted at conference
- `CitationCount`: Total number of citations for papers accepted at conference

The calculation of the aforementioned G_1-index uses the `CitationCount` while that of G_2-index uses the `PaperCount` in addition. The `OfficialUrl` attribute, if not null, readily provides the location of conference websites. When it is null, one can still attempt to locate the URLs of conferences using other means such as search engines.

4.2 STEP 2: Finding TPC Webpages

A conference website many have hundreds of webpages underneath. Among these webpages, in this step, we need to identify a set of webpages that list the TPC member information (i.e., gatekeepers). Essentially, we implemented a simple pattern-matching based detection (e.g., using a set of relevant keywords or phrases {program committee, TPC, reviewers, ...}). However, one can also build a more sophisticated machine learning model using features from contents of webpages and structures of websites.

4.3 STEP 3: Scraping Gatekeepers

Once TPC webpages are found, next, we attempt to scrape gatekeeper information from the webpages. This scraping is a non-trivial task due to varying and disagreeing formats that conference websites use. For instance, Figure 3 shows four example TPC webpages, where gatekeeper information is listed in vastly different formats. Despite their differences in formats, however, we can derive a few heuristic rules: (1) Gatekeepers' information is displayed in some type of repeating pattern such as "list," and (2) Gatekeepers have human-like names. Therefore, we wrote a script to detect human-like names in a repeating fashion in the given webpage. The script goes through the entire HTML page, tokenizing each tag, and apply Stanford NER (Name Entity Recognizer) package[2] to identify each gatekeepers name and additional accompanying information such as email, homepage URL, and affiliation (if one exists). The recent study [1] revealed that Stanford NER outperformed other NER packages such as Illinois NET, OpenCalais NER WS, and Alias-i LingPipe. All encountered unique human names are kept in an underlying database, forming many-to-many relationships between conferences and gatekeepers. At the end, we have scraped 56,187 gatekeepers serving at 2,825 computer science conferences.

[2] https://nlp.stanford.edu/software/CRF-NER.html

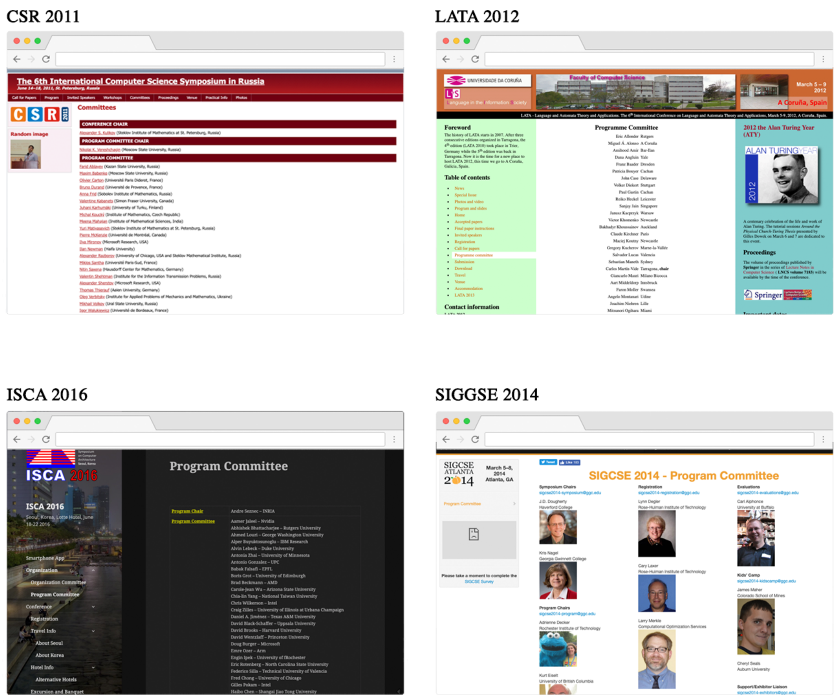

Fig. 3: Example TPC webpages with varying formats of gatekeeper information.

4.4 STEP 4: Scraping Research Keywords

Conference and gatekeeper information are the primary data to be collected, however, research keywords are also being collected. Research keywords are usually relevant words to either the topic/genre of a conference or research interests of a gatekeeper. These keywords can aid users in searching for specific gatekeepers or conferences. We first considered to run a topic model method such as LDA over the contents of conference websites or a collection of (titles of) articles authored by gatekeepers. However, by connecting gatekeepers to their corresponding pages in Google Scholar, where scholars often voluntarily pick a small number of research areas or keywords, we were able to scrape a list of self-defined research keywords of gatekeepers.

The summary of data statistics in our current prototype is listed in Table 2.

5 Prototype Gatekeeper

We have built a prototype digital library, Gatekeeper, with a limited number of conferences and gatekeepers. The backend of Gatekeeper is operating on a MySQL database. Using a combination of PHP and HTML, a frontend interface

Table 2: Summary of data statistics in the prototype.

Data Type	Count
Conferences	2,825
Gatekeepers	56,187
Edges/Connections	87,368
Years	27

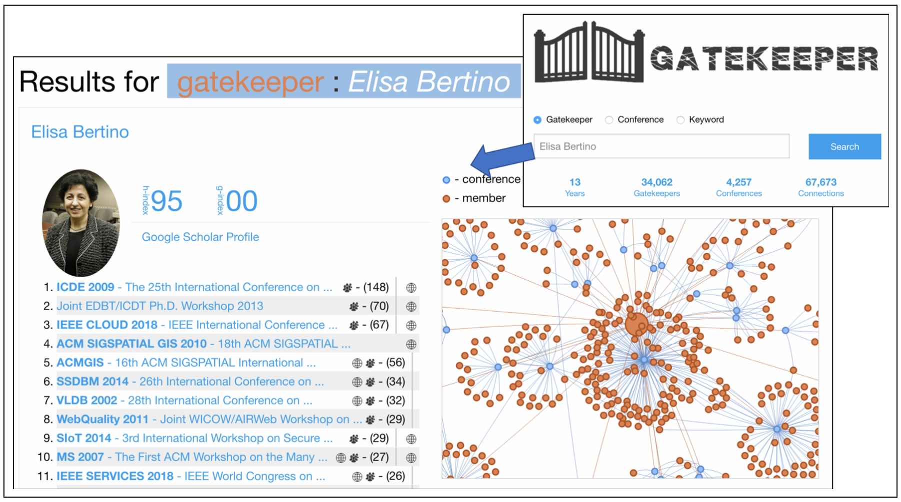

Fig. 4: Querying a gatekeeper "Elisa Bertino."

allows users to query information from Gatekeeper. The entire Gatekeeper runs on an AWS (Amazon Web Services) Windows Server-based EC2 (Elastic Cloud Computing) server. The system in its current state allows for users to browse a list of top conferences (with respect to their research impacts) or top gatekeepers (with respect to their service impacts), or search by a gatekeeper name, a conference name, or a research keyword.

5.1 Querying Gatekeepers

A query for a gatekeepers name will produce a list of conferences served by the gatekeeper, sorted in descending order with respect to one of Gatekeeper-index scores. Users can switch among Gatekeeper-index methods to see different result. For instance, Figure 4 shows an example for a gatekeeper "Elisa Bertino." In the conference list, next to each conference title will be an indicator of how many gatekeepers have served at that particular conference. Lastly, a node-based graph will represent the connections of the gatekeeper with their conferences. Viewing larger served conferences or more servicing gatekeepers results in being able to see the formation of clustering communities.

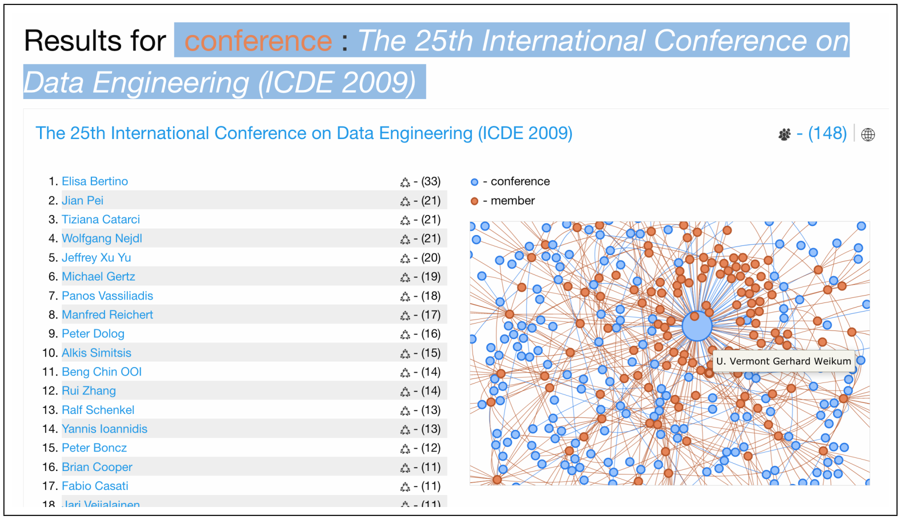

Fig. 5: Querying a conference "International Conference on Data Engineering / ICDE 2009."

5.2 Querying Conferences

Much like a query for a gatekeepers name, a query for a conference title will produce a list of all matching records in the database. The returned list is currently ordered by the number of gatekeepers serving the conferences from greatest to least. After the calculation of the one of Gatekeeper-index methods, users will be offered multiple ways to sort the returned information. There is an indicator next to each conference title returned that identifies the number of gatekeepers who have served in the conference. Upon clicking on the conference, the conference detail page will show the list of gatekeepers who have served at the conference. For instance, Figure 5 shows an example for a conference "International Conference on Data Engineering / ICDE 2009." Next to each gatekeeper name will be an indicator of how many conferences they have served. Lastly, a node-based graph will represent the connections of the conferences with their gatekeepers.

5.3 Querying Research Keywords

A query for a research keyword will produce a list of conferences (whose research themes match the keyword) or gatekeepers (whose research interests overlap the keyword). For instance, Figure 6 shows an example for the keywords "artificial intelligence."

5.4 Other Applications of Gatekeeper

While not implemented yet, there are many useful applications of Gatekeeper that we plan to work in future.

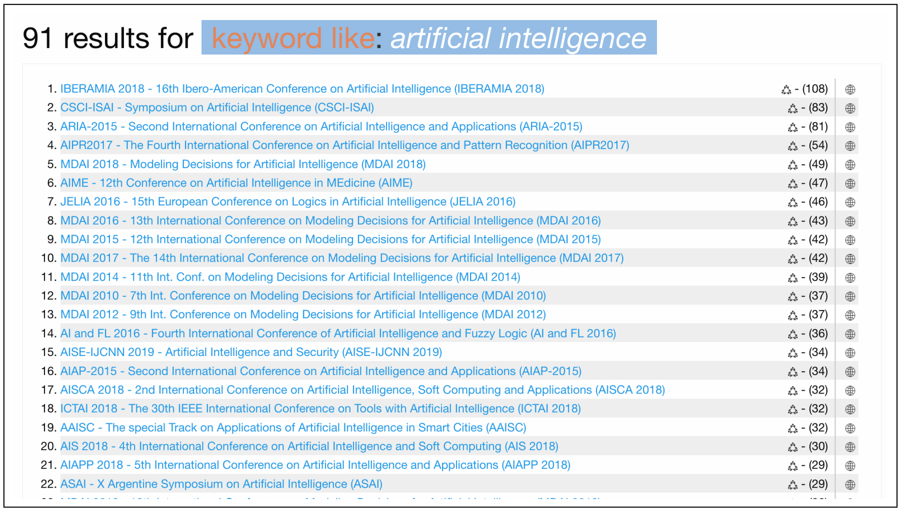

Fig. 6: Querying a keyword "artificial intelligence."

First, consider an application that requires a group of experts in various settings. For instance, a program director at a funding agency may want to identify 20 experts to convene a review panel whose expertise closely match the theme of a program X. Similarly, a program chair of a new conference on a topic Y may want to identify 30 scholars who can serve as TPC member. This is so-called the expert-finding problem [5]. By ranking gatekeepers with particular research keywords, then, one can easily find a group of experts for such settings. In addition, such tasks can be also solved by modeling it as a graph-based community detection or a recommendation problem on top of a bipartite graph between gatekeepers and conferences.

Second, by solving the link prediction problem in a graph of gatekeepers, one can recommend new service collaboration among gatekeepers. If two gatekeepers have not served in the TPC of the same conferences in past, but have neighboring gatekeepers in common, then such gatekeepers can be put into the same service collaboration in future.

5.5 Future Work

As the prototype currently covers only TPC members of Computer Science conferences, it needs to be significantly expanded to cover other types of gatekeepers (e.g., journal editors), other disciplines (e.g., Physics), and differentiate the roles of services (e.g., TPC member vs. Senior TPC member vs. Track Chair). This requires the development of a more sophisticated web crawling and entity/attribute scraping algorithm.

In addition, to have meaningful distributions of Gatekeeper-index scores, the dataset needs to be significantly expanded to cover a larger number of conferences over a more comprehensive period across multiple disciplines.

Also, we intend to use the Wayback Machine to explore conferences with identical year-to-year website URLs. Use of the Wayback Machine may be able to provide us with previous TPC committee lists.

Finally, we plan to conduct user studies to understand Gatekeeper-index methods better. For instance, we need to study the following questions: Among three variants, which method's scores are more likely to be in sync with scholars' impressions? Do the score distributions of Gatekeeper-index methods make sense?

6 Conclusion

In this work, we have presented an early attempt to design and implement a digital library to assess the impacts of service to the scientific community. To quantify the impacts of service in serving as TPC members of computer science conferences, we proposed three Gatekeeper-index methods that capture both productivity as well as impact of service.

References

1. Samet Atda and Vincent Labatut. A comparison of named entity recognition tools applied to biographical texts. In *International Conference on Systems and Computer Science*, 2013.
2. Sergey Brin and Larry Page. The anatomy of a large scale hypertextual Web search engine published version. *Computer Networks and ISDN Systems*, 1998.
3. Patrick Chen, Huafeng Xie, Serfei Maslov, and Sidney Redner. Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, 2007.
4. Leo Egghe. Theory and practise of the g-index. *Scientometrics*, 2006.
5. Hui Fang and ChengXiang Zhai. Probabilistic Models for Expert Finding. In *Proceedings of the 29th European Conference on IR Research*, pages 418–430, Berlin, Heidelberg, 2007. Springer-Verlag.
6. Eugene Garfield. Citation indexes for science. *Science*, 1955.
7. Eugene Garfield. The history and meaning of the journal impact factor, 2006.
8. Jorge E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 2005.
9. Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 1999.
10. Xiancheng Li, Wenge Rong, Haoran Shi, Jie Tang, and Zhang Xiong. The impact of conference ranking systems in computer science: a comparative regression analysis. *Scientometrics*, 116(2):879–907, 8 2018.
11. Antonis Sidiropoulos and Yannis Manolopoulos. A new perspective to automatically rank scientific conferences using digital libraries. *Information Processing and Management*, 2005.

12. Su Yan and Dongwon Lee. Toward Alternative Measures for Ranking Venues: A Case of Database Research Community. In *7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pages 235–244, New York, NY, USA, 2007. ACM.
13. Ziming Zhuang, Ergin Elmacioglu, Dongwon Lee, and C. Lee Giles. Measuring conference quality by mining program committee characteristics. In *7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, 2007.