Network Analysis of the Science of Science: A Case Study in SOFSEM Conference

Antonia Gogoglou[1], Theodora Tsikrika[2], and Yannis Manolopoulos[1]

[1] Department of Informatics, Aristotle University, 54124 Thessaloniki, Greece
agogoglou,manolopo@csd.auth.gr
[2] Centre for Research and Technology Hellas, Thessaloniki, 57001 Greece
theodora.tsikrika@iti.gr

Abstract. A rising issue in the scientific community entails the identification of temporal patterns in the evolution of the scientific enterprise and the emergence of trends that influence scholarly impact. In this direction, this paper investigates the mechanism with which citation accumulation occurs over time and how this affects the overall impact of scientific output. Utilizing data regarding the SOFSEM Conference (International Conference on Current Trends in Theory and Practice of Computer Science), we study a corpus of 1006 publications with their associated authors and affiliations to uncover the effects of collaboration network on the conference output. We proceed to group publications into clusters based on the trajectories they follow in their citation acquisition. Representative patterns are identified to characterize dominant trends of the conference, while exploring phenomena of early and late recognition by the scientific community and their correlation with impact.

Keywords: scientometrics, bibliographic data, time series clustering, trends

1 Introduction

With the extensive recording of scientific endeavors in large scale online databases and a rising interest in assessing scientific impact, the "science of science" [1] has attracted significant attention. However, the age old question in the quantification and evaluation of scientific impact still remains: Does a pattern for success exist and what can cause a publication or scholar to stand out? First, it is necessary to quantify success effectively and then investigate the process that leads to high performance levels. Since the seminal work of Eugene Garfield [10], the acknowledgment received by peers in the form of citations serves as the most straightforward measure for representing visibility and recognition by one's cohorts; therefore it is the most widely used metric for popularity. Even though many different approaches exist for measuring citations and correlating them with impact [20], the timing of each received citation is also of high importance. How do citations accumulate? Is the process unique for each individual or are there identifiable trends and, if so, how do they relate to impact? What is the role of collaboration in citation acquisition?

Across different scientific disciplines, countries and performance levels, the process of accumulating citation varies widely. Efforts have focused on profiling scholars and their publications to compartmentalize their growth and identify similarities amongst seemingly unrelated scientific entities (e.g. publications from different authors or time periods). In [4] an extensive study of Computer Science publications revealed six dominant categories based on their citation attraction process and associated this categorization with year of publication, publishing venue and topological features of the citation network. At author level, in [11], five scholarly profiles were distinguished for Computer Scientists in terms of temporal evolution and their overall impact was correlated with frequency of publications ("publish or perish effect" [8]). Grouping of scientific entities in profiles proves to be of assistance to the estimation of future impact evolution [3, 7], since past behavior may not only determine current but also future status. Given the diversity of observed profiles, building a specialized prediction model for each profile can produce more accurate predictions.

Modeling citation trajectories as spatio-temporal objects can shed light into the process that leads to success. More specifically, the citation time series of a paper reveals whether the number of citations increases steadily, or it saturates after some time, or whether the paper seems to receive a belated citation explosion. Citation time series have been utilized for identifying scientific breakthroughs [17, 21], while entire citation networks have been studied accounting for temporal degeneracy [5]. Focusing specifically on the timing of citation shifts [9], citation cascades have been associated with paradigm shifting in scientific discoveries [15]. These cascading events were also found to reveal unique patterns, such as the "sleeping beauty" effect [13] where a publication exhibits a long hibernation period before receiving recognition or early discoveries, where a citation boost occurs soon after publication [6]. It turns out that these distinguishing citation patterns do not constitute an isolated scarce phenomenon, but occur often in science highly affecting careers, future visibility and even award giving or fund allocation.

The real challenge in these efforts is to determine a trend given the diversity of publishing behaviors that arise in science. Essentially, fair comparisons need to be computed amongst publishing and citing patterns of scientists of different age and background in different time periods. In this work, we attempt to tackle this challenge and contemplate the following research questions:

- What is the dominant trend in the temporal evolution of publications in e.g. a particular conference? Are they steady, rising or decaying over time?
- When does the peak of citations occur for most publications and does increased output mean deviation from trends?
- And finally, is there a correlation between these temporal patterns and the total output or other academic features (e.g. affiliations)?

To address the aforementioned questions, we contemplate the temporal popularity dynamics of the citation curves for individual publications associated with the SOFSEM conference[3]. We conduct a bibliometric analysis of the con-

[3] https://link.springer.com/conference/sofsem

ference records and identify prominent participants, frequent contributors and associated communities. Next, we extract citation curves from the historical data of the conference; a *citation curve* is defined as the set of points that represent citations acquired at given time steps (e.g. yearly). We attempt to fit these curves into representative profiles, while characterizing the members in each profile according to their collective output, set of authors, and associated affiliations. Indeed, our goal is not to perfectly model the popularity evolution of all possible trajectories, but rather capture the most prevalent tendencies based on shape similarity, regardless of differences in amplitude and phase. We build upon similar efforts that address online content growth as a time series pattern mining problem studying how different pieces of user generated content compete for attention in mircoblogs (e.g. Twitter) [22]. Apart from the shape of the total curve, we additionally micro-analyze the timing of shifts in the time series of citations to comprehend the mechanism causing citation boosts and how it relates to the total impact.

The rest of the paper is organized as follows: Section 2 describes the process for collecting the data, while Section 3 provides an overview of our bibliometric analysis for the SOFSEM conference. Section 4 focuses on the temporal evaluation of publications and Section 5 concludes the article.

2 Data Acquisition

SOFSEM (SOFtware SEMinar) was first held in 1974 as a local Czechoslovakian event to bring together theorists and practitioners of computing. Since 1995 it has been steadily evolving into a fully-fledged annual multidisciplinary international conference on Current Trends in Theory and Practice of Informatics with participants from multiple European countries including UK, France, Germany, and Spain. For the next 21 years, the conference location has alternated between Czech Republic and Slovakia, while in 2017 it was first held in a different location (Limerick, Ireland). Since 1995, the conference proceedings have featured in the Lecture Notes in Computer Science (LNCS) series by Springer.

To obtain the SOFSEM 1995-2017 publication data, the DBLP XML dump [14] (as downloaded on June 23, 2017) was processed using appropriate XQuery queries that featured the "conf/sofsem" keyword. This led to the collection of 1027 publication titles in the *main* SOFSEM proceedings (i.e., excluding papers/posters published in the SOFSEM Student Research Forum proceedings) over 22 years[4], resulting in an average of 47 publications per year. The next step was to gather metadata and citation records on these publications.

Regarding the citation data, many online data sources are available, either proprietary, such as the Web of Science[5] by Clarivate Analytics and Scopus[6] by

[4] The SOFSEM 2003 proceedings are not listed in DBLP and thus omitted from this study.

[5] `https://apps.webofknowledge.com/`

[6] `https://www.scopus.com/`

Elsevier, or open source ones, such as Google Scholar[7] and Microsoft Academic[8]. Each follows a different data collection policy that affects both the publications covered and the number of citations found, while differences in their coverage may affect the assessment of scholarly impact metrics [12]. For the purposes of our analysis, we focused on freely available databases that do not require subscriptions and we opted for the newly introduced Beta 2.0 version of Microsoft Academic. Even though Google Scholar also offers wide coverage of citation records, Microsoft includes a more structured collection of scientific entities (conferences, journals, author and institutional profiles). Therefore, we queried its database for the publication titles collected from DBLP adding the keyword SOFSEM, since publications with the same title are often published later on in other venues (e.g. journals) as well. Additionally, Microsoft Academic offers author profiles which alleviates author name disambiguation issues that often arise in other citation databases. Out of the original publication set, 1006 publication titles (98%) were identified in Microsoft Academic and their publication year, authors with related affiliations, as well as yearly citation records were obtained.

3 Bibliometric Analysis

First, we conduct a bibliometric analysis of the records collected to identify the most prominent participants in the conference over the years, the diversity of participants and their institutions and explore the dynamics of collaboration amongst them. Table 1 illustrates the highest ranking publication based on various citation rates (total, average and peak), while Tables 2 and 3 illustrate, respectively, authors and institutions with the highest participation rates in the conference and the biggest impact, as measured by the total citations acquired by their publications in SOFSEM. An interesting observation in the selected publication titles is that the ones with the highest total citation count are not necessarily the ones that received the biggest boost in citations or the ones with a steady average citation rate over the years. This leads us to the realization that different citation patterns can lead to increased overall impact. Another intriguing finding regarding authors and their affiliated institutions is that high productivity, meaning a high participation rate, does not guarantee a higher impact level. Therefore the question rises, what makes an author stand out in this conference?

To explore the presence of each author amongst their collaborators and the effects of it on their impact, we created the co-authorship network that is represented as an undirected graph where nodes correspond to authors and edges correspond to a co-authored publication. The resulting graph is depicted in Figure 1 filtered by size and color based on two centrality metrics: *degree* and *betweeness* centrality. *Degree centrality* is computed by counting the neighbors of each node, whereas *betweeness centrality* is equal to the number of shortest

[7] `https://scholar.google.com`

[8] `http://academic.research.microsoft.com/`

Table 1: Top rated publications in the SOFSEM proceedings.

# total citations		
Chevaleyre et al. : A Short Introduction to Computational Social Choice	2007	155
Appelt: WWW Based Collaboration with the BSCW System	1999	143
Rahman & Iliopoulos: Indexing Factors with Gaps	2007	136
Allauzen et al.: Factor Oracle: A New Structure for Pattern Matching	1999	115
Bodlaender: Discovering treewidth	2005	115
# average citations		
Chevaleyre et al.: A Short Introduction to Computational Social Choice	2007	14
Lee et al.: Efficient Group Key Agreement for Dynamic TETRA Networks	2007	13
Rahman & Iliopoulos: Indexing Factors with Gaps	2007	12
Navigli: A quick tour of word sense disambiguation	2012	11
Dolog: Designing Adaptive Web Applications	2008	8
# citations peak		
Lee et al.: Efficient Group Key Agreement for Dynamic TETRA Networks	2007	34
Chevaleyre et al.: A Short Introduction to Computational Social Choice	2007	28
Rahman & Iliopoulos: Indexing Factors with Gaps	2007	22
Appelt: WWW Based Collaboration with the BSCW System	1999	20
Navigli: A quick tour of word sense disambiguation	2012	16

Table 2: Most prolific and cited authors

# publications		# citations	
Mária Bieliková	11	Yann Chevaleyre	155
Costas S. Iliopoulos	7	Ulle Endriss	155
Shunsuke Inenaga	7	Nicolas Maudet	155
Friedrich Otto	7	M. Sohel Rahman	152
Henning Fernau	6	Costas S. Iliopoulos	146

Table 3: Most prolific and cited institutions

# publications		# citations	
Charles University in Prague	64	ETH Zurich	662
Slovak Univ. of Technology in Bratislava	50	King's College London	449
University of Latvia	49	University of Amsterdam	392
Masaryk University	47	University of Latvia	318
ETH Zurich	40	Lamsade (Univ. Paris-Dauphine)	316

paths from all nodes to all others that pass through that specic node (i.e. author). *Closeness centrality* was also calculated for the participating authors, which is the mean distance from a node to others. For our co-authorship graph, as it contained a number of disconnected nodes, we utilized the harmonic mean to calculate representative values for the closeness centrality [18]. Essentially a high degree centrality indicates a scientist with a large number of co-authors, while betweeness centrality gives highest values to individuals through whom information is more likely to pass, i.e. they bridge different groups of collaborators.

Fig. 1: Visualization of the SOFSEM co-authorship network, with darker colored nodes representing high number of authored publications and bigger sized nodes representing higher betweeness centrality values.

Closeness centrality, in turn, highlights the actors who will be able to contact easily all other members of the network, meaning they share many common collaborators with other participants. As seen in Figure 1, a large number of small author communities appear that are seemingly disconnected from the rest of the network. We observe though some densely connected groups formed around nodes with high betweeness centrality further indicating these nodes' level of influence.

We also performed a similar analysis on a higher level of granularity by considering the co-authorship network where the nodes correspond to the countries of the authors' affiliations and edges to the co-authored publications. The analyzed SOFSEM publications were collaboratively produced by authors affiliated with institutions in 55 countries. Figure 2 shows the collaborations between the different countries in the SOFSEM community and depicts 51 countries and 158 edges. The most prolific country in terms of publications (Germany) is also the most extrovert with the most collaborations. On the other hand, the second

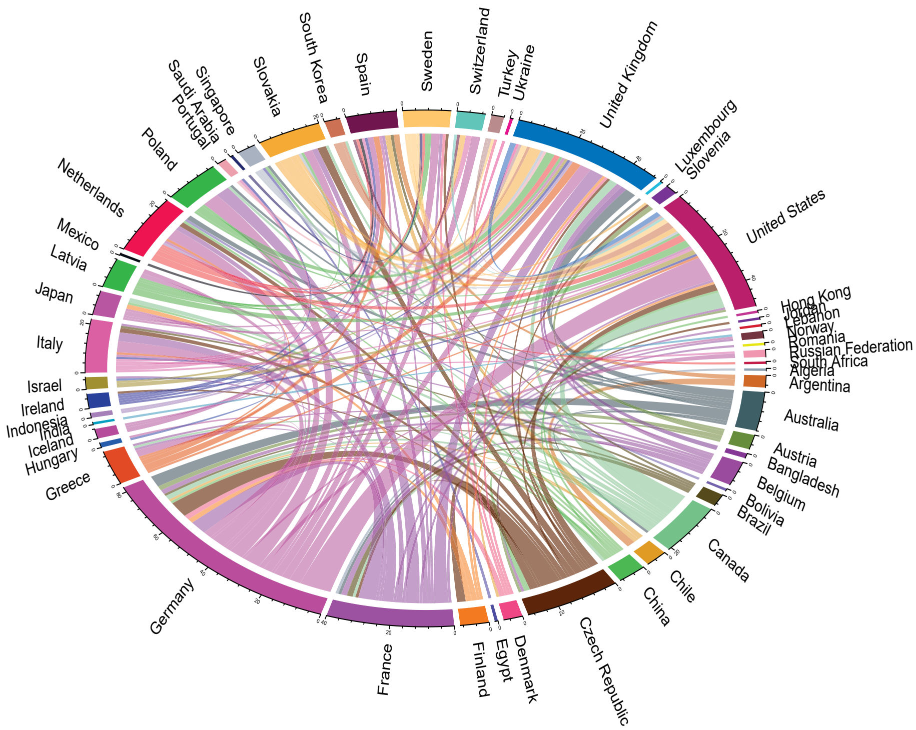


Fig. 2: Co-authorship network based on the countries of the authors' affiliations.

most prolific country in terms of publications authored (Czech Republic) is fifth in terms of collaborations, indicating a more conservative approach.

Apart from collaboration relationships, we identify the set of authors that have consistently participated in SOFSEM and received high recognition to distinguish the patterns that led to their increased status. The citation time series of the selected authors are included in Table 4 along with their closeness and betweeness centrality values. The selected scientists are ranked in descending order of publication number (size of citation vector) in SOFSEM conference. As we observe, they appear to follow very different citation patterns, with some achieving high boosts in citations (e.g. Keith G.Jeffrey) while others displaying a moderate but steady rate (e.g. Michal Barla). However, the majority of the selected prominent scientists share high values in betweeness centrality indicating that obtaining strategic collaborations with scientists from diverse co-authorship groups and bridging them together is the most effective pattern for overall increased visibility and popularity. On the other hand, establishing multiple co-authorship relationships (higher closeness centrality) appears to have little effect on impact.

Table 4: Citation records, closeness and betweeness centrality values for authors with more than 2 SOFSEM publications and more than 10 citations overall.

Author Name	Citation vector	Closeness Centrality	Betweeness Centrality
Costas S. Iliopoulos	[136, 1, 11, 1, 0, 0, 16]	1.25	22
Keith G. Jeffrey	[143, 1, 0, 2, 0]	1.00	12
Hans L. Bodlaender	[115, 1, 0, 0, 4]	1.00	17
Juraj Hromkovi	[0, 11, 56, 0, 5]	1.25	45
Petr Jancar	[12, 0, 13, 30]	1.30	3
Michal Barla	[6, 7, 13, 20]	2.00	74.3
Hans-Joachim Bckenhauer	[1, 11, 56, 1]	1.00	12.5
Nieves R. Brisaboa	[0, 17, 39]	1.80	9
Oscar Pedreira	[0, 17, 39]	1.00	22
Michal Tvaroek	[6, 13, 20]	2.60	6
Maxime Crochemore	[115, 11, 1]	1.00	9
Wojciech Rytter	[0, 11, 27]	1.42	1
Johannes Uhlmann	[17, 5, 12]	1.40	1
Ngoc Thanh Nguyen	[27, 5, 1]	1.00	3

Next, we will explore the patterns that lead to high impact at publication level and how they correspond to author impact.

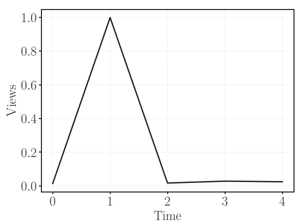
4 Temporal Dynamics of Scholarly Impact

Time-series sequences, such as citation curves, advance with respect to two axis, time and scale (or magnitude). We propose two different approaches to study a set of such sequences and identify temporal patterns: one is macroscopic focusing on the *shape* of the resulting curves regardless of citation scale or timing of shifts, while the other one is microscopic contemplating the *relationship* between magnitude of citations and the timing of occurrence. The result of the first approach is a set of profiles of publications going through similar stages of impact. The second approach provides a different categorization of publications with respect to the timing of their recognition and their aging process.

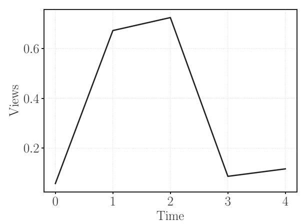
4.1 Publication Profiles

The need for clustering time series with scale- and shift-invariant methods has emerged in multiple fields, such as business, social media, medicine, biology, etc. [11, 16], with the goal to identify and summarize interesting patterns and correlations in the underlying data. In this work, we employ a recently proposed time series clustering algorithm called K-spectral clustering (KSC) [22] that has been utilized to discover common trends in the spread of online content. The KSC algorithm groups times series based on the shape of the curve and thus respects invariants of scale in the popularity axis and shifts in the time axis. That is, two entities that have their popularity evolving according to similar processes (e.g.

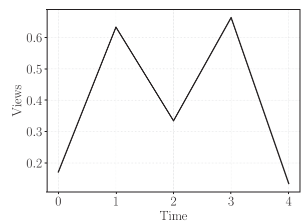
linear growth) will be assigned to the same cluster by KSC, regardless of the popularity values. KSC requires that all time series are comprised of the same number of points.



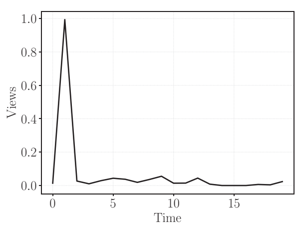
(a) Cluster 0 for TS=5



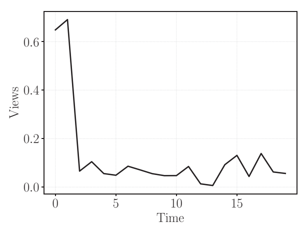
(b) Cluster 1 for TS=5



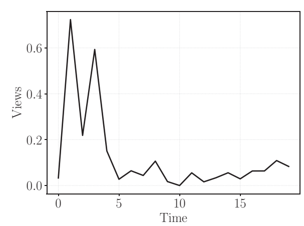
(c) Cluster 2 for TS=5



(d) Cluster 0 for TS=20



(e) Cluster 1 for TS=20



(f) Cluster 2 for TS=20

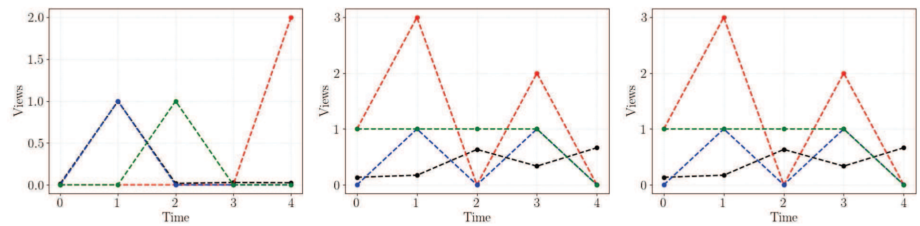
Fig. 3: Citation patterns for the centroids of each of the three clusters for two different time spans: short-term $TS = 5$ (top) and long-term $TS = 20$ (bottom).

Regarding the citation vectors, we represent each publication with a series of t points each corresponding to the citations this particular publication acquired in one particular year, starting from its publication year. Because publication ages vary from 1 to 22 for our given time period (1995-2017), we define four time spans ($t = TS$) that correspond to the minimum age of the publications included in each span and consider only the first t years of a publication's life. We consider $TS = 5, 10, 15$ *and* 20 years so that patterns for both long- and short-term impact can be studied. A predefined number of clusters k also needs to be determined and in our case we opted for $k = 3$ based on optimal inter- and intra-cluster distance amongst publications.

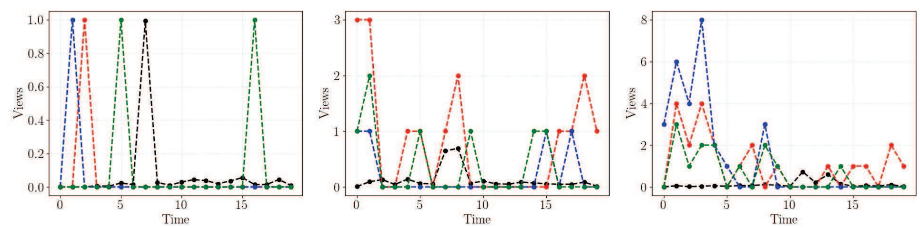
The implementation of KSC we adopted[9] closely resembles the classic k-means but with a different definition for the distance metric. The similarity between two vectors x and y (in our case of citations) is calculated as follows:

$$d(x, y) = \min_{a,q} \frac{||x - \alpha y_{(q)}||}{||x||} \tag{1}$$

[9] http://github.com/aviovdf/pyksc



(a) Cluster 0 for TS=5 (b) Cluster 1 for TS=5 (c) Cluster 2 for TS=5



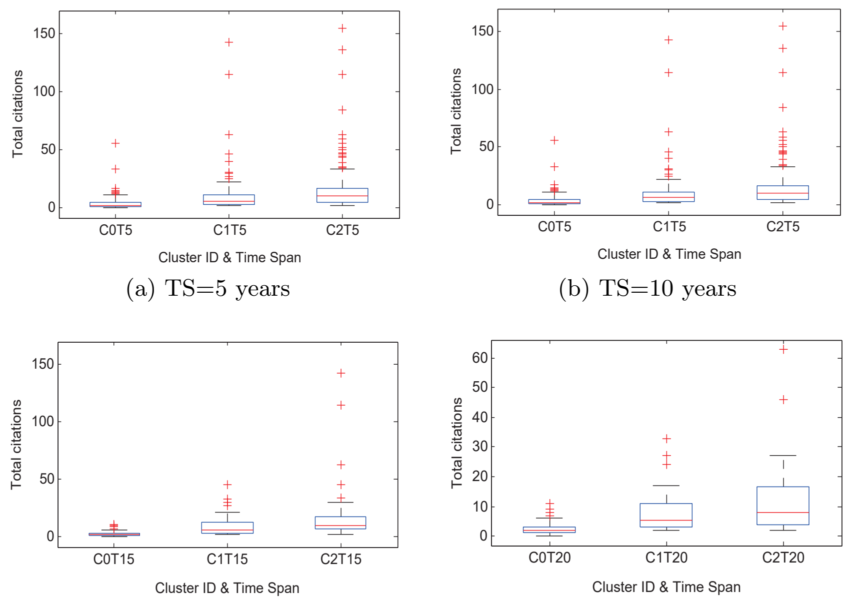
(d) Cluster 0 for TS=20 (e) Cluster 1 for TS=20 (f) Cluster 2 for TS=20

Fig. 4: Four examples of members from each of the three clusters for two different time spans: short-term $TS = 5$ (top) and long-term $TS = 20$ (bottom).

where $y_{(q)}$ represents the shift of vector y by q units and $||.||$ the l^2-norm [2]. In the above dual minimization problem there is no straightforward way to compute q; therefore, we follow a heuristic proposed in the original paper [22] that includes searching for the optimal value of q in the range of all integers $(-t, t)$, where t is the size of the time series, as mentioned above. Given a fixed q, the exact solution for α can be obtained by computing the minimum distance d from Equation 1.

By shifting citation vectors to find optimal values for the distance metric, we were able to match publications to three prevalent patterns. The interesting finding here is that these patterns, as represented by the cluster centroids, appear to be similar over time, meaning that analogous patterns are identified when contemplating either the first 5 or 20 years of a publication's history. As can be seen in Figure 3, the three patterns can be summarized as one with a steep peak (referred to as cluster 0), another one with a peak followed by a more smooth decay (cluster 1) and, finally, a curve with two prominent peaks and a relatively steady acquisition rate (cluster 2). Figure 4 displays four examples of citation trajectories from each cluster for two selected time spans ($TS = 5$ and 20 years).

How do these patterns relate to impact? Figure 5 depicts the distribution of total citation count for each cluster over all time spans. A clear pattern here is that cluster 2 is associated with higher citation counts, whereas cluster 0 that includes single peak publications leads to lower overall impact. Therefore, one can assume that a single boost of citations does not relate to actual impact, whereas a pattern of multiple peaks amongst a steady rate of citations indicates an influential publication over time. But does the timing of the peak/s matter?



(a) TS=5 years

(b) TS=10 years

(c) TS=15 years

(d) TS=20 years

Fig. 5: Boxplots of total citation counts for all three clusters for each of the four time spans; e.g. C0T5 represents cluster 0 at time span equal to 5 years.

4.2 Publication Recognition: Timing and Aging

In this subsection, we explore the timing of citation shifts and the aging process of publications. Studies examining citation patterns have identified different behaviors of early recognition or long hibernation periods for publications. As introduced in [19], a metric to calculate the obsolescence of publications, without examining each citation curve individually to identify shifts, is defined as:

$$G_s = 1 - \frac{2 \times [n \times C_1 + (n-1) \times C_2 + ... + C_n] - C}{C \times n} \tag{2}$$

where n is the age of a publication, C is the total number of citations, and C_i corresponds to the citations until the i^{th} year. We refer to G_s as the *aging coefficient* and dependent on its calculated values, we can assign publications to groups related to the timing of their recognition.

For the purposes of our study and given the citation rates observed in our dataset, we employ the following thresholds to define three distinct *timing* categories for publications with "extra-ordinary" citation trajectories:

- $0.1 < G_s < 1$ and $C > 10$ indicates a *sleeping beauty*, meaning a publication that received recognition after a long period of time;
- $G_s < 0$ and $C > 10$ indicates a *flash in a pan*, meaning a publication that received a citation boost soon after its release; and
- $0 < G_s < 0.05$ and $C > 10$ indicates an *aging gracefully* publication, meaning it maintains a steady citation rate for longer periods.

Table 5 contains information on publications categorized in one of the above groups based on their aging coefficient. We observe highly prestigious institutions and authors in all three categories indicating that the timing of impact does not directly relate to the size of impact. Moreover, one of the most seminal publications of the conference, "A Short Introduction to Computational Social Choice", managed to acquire citations steadily leading to a graceful aging period, while another highly popular publication, "Automatic Testing of Object-Oriented Software"", appears to have acquired 59 citations in total with the majority of them occurring soon after publication. On the other hand, a comprehensive survey by A. Goldberg, "Point-to-Point Shortest Path Algorithms with Preprocessing", did not rise in popularity until several years after publication. Looking into the citation ranges and the categories that mostly populate them in Figure 6, we further realize that publications from all categories can obtain high citation counts, with a slight competitive edge attributed to the flashes in a pan category.

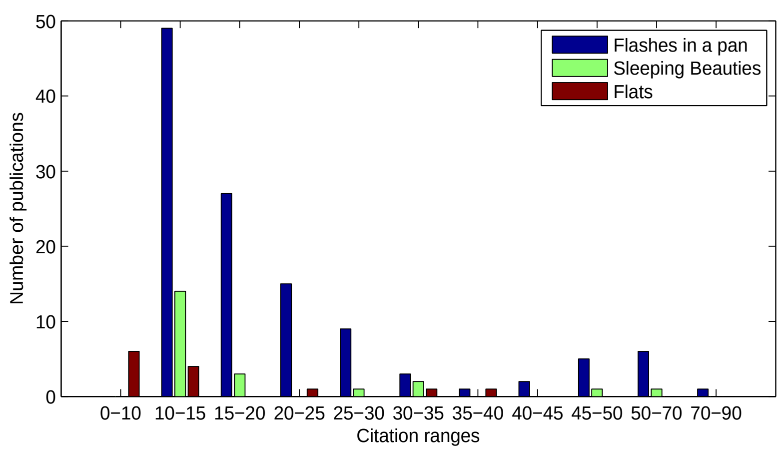


Fig. 6: Number of publications from each timing category that belong to various citation ranges.

Table 5: Examples of publications belonging to each timing category based on the timing of their recognition including title, authors and affiliations.

Category	Titles and # of citations	Authors
flashes in a pan	Automatic Testing of Object-Oriented Software (59)	Bertrand Meyer Ilinca Ciupa Andreas Leitner Lisa Ling Liu (ETH Zurich)
	Sample Method for Minimization of OBDDs (27)	Anna Slobodova (Comenius University in Bratislava), Christoph Meinel (Universitt Potsdam)
	Improving watermark resistance against removal attacks using orthogonal wavelet adaptation (40)	Jan Stolarek (University of Edinburh), Piotr Lipiski (University of Edinburh)
	Explicit Connectors in Component Based Software Engineering for Distributed Embedded Systems (16)	Dietmar Schreiner (Vienna University of Technology), Karl M. Gschka (Vienna University of Technology)
sleeping beauties	On the NP-Completeness of some graph cluster measures (50)	Jiri Sima (Academy of Sciences Czech Republic), Satu Elisa Schaeffer (Helsinki University of Technology)
	Domain Engineering: A Software Engineering Discipline in Need of Research (11)	Dines Bjrner (Technical University of Denmark)
	Fuzzy Set Theory and Medical Expert Systems: Survey and Model (14)	Nguyen Hoang Phuong (Academy of Sciences Czech Republic)
	Point-to-Point Shortest Path Algorithms with Preprocessing (25)	Andrew V. Goldberg (Microsoft)
aging gracefully	A Short Introduction to Computational Social Choice (155)	Yann Chevaleyre (Lamsade), Ulle Endriss (University of Amsterdam), Jrme Lang (Centre national de la recherche scientifique), Nicolas Maudet (Lamsade)
	Complexity of model checking for modal dependence logic (21)	Johannes Ebbing (Leibniz University of Hanover), Peter Lohmann (Leibniz University of Hanover)
	Spatial Selection of Sparse Pivots for Similarity Search in Metric Spaces (39)	Oscar Pedreira (University of A Corua), Nieves R. Brisaboa (University of A Corua)
	Recent challenges and ideas in temporal synthesis (13)	Orna Kupferman (Hebrew University of Jerusalem)

5 Conclusions

In this work, we conducted a bibliometric analysis of publication and citation records of the SOFSEM conference to determine the mechanism that leads to high impact scientific output. Exploring the effects of affiliations and co-authorship we realized that scientists bridging together different communities through collaboration are more likely to produce popular publications. We then focused on identifying citation patterns over the years and an interesting finding was that there exist three distinct trajectory patterns in citation acquisition for both long- and short-term impact irrespective of timing and magnitude of popularity. Going one step further, we revealed publications with different timing in receiving recognition and concluded that the timing of citation boosts does not correlate to impact in the same degree as the overall shape of the citation time series. Therefore, increased popularity is mostly achieved by publications that obtain multiple citation sprees and manage to age gracefully over time.

References

1. Börner, K., Dall'Asta, L., Ke, W., Vespignani, A.: Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. Complexity 10(4), 57–67 (2005)
2. Bourbaki, N., Eggleston, H., Madan, S.: Topological Vector Spaces. Éléments de mathématique, Springer-Verlag (1987)
3. Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., Mukherjee, A.: Towards a stratified learning approach to predict future citation counts. In: Proceedings 14th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL). pp. 351–360 (2014)
4. Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., Mukherjee, A.: On the categorization of scientific citation profiles in computer science. Communications of the ACM 58(9), 82–90 (2015)
5. Clough, J.R., Evans, T.S.: Time and citation networks. In: Proceedings 16th Conference of the International Society of Scientometrics & Informetrics (ISSI) (2015)
6. Costas, R., van Leeuwen, T.N., van Raan, A.F.: Is scientific literature subject to a sell-by-date? a general methodology to analyze the durability of scientific documents. Journal of the American Society for Information Science and Technology 61(2), 329–339 (2010)
7. Davletov, F., Aydin, A.S., Cakmak, A.: High impact academic paper prediction using temporal and topological features. In: Proceedings 23rd ACM International Conference on Conference on Information & Knowledge Management (CIKM). pp. 491–498 (2014)
8. Editorial: Publish or perish. Nature 467, 252–252 (2010)
9. Egghe, L., Bornmann, L., Guns, R.: A proposal for a first-citation-speed-index. Journal of Informetrics 5(1), 181 – 186 (2011)
10. Garfield, E.: The application of citation indexing to journals management. Current Contents 33, 3–5 (1994)
11. Gonçalves, G.D., Figueiredo, F., Almeida, J.M., Gonçalves, M.A.: Characterizing scholar popularity: A case study in the computer science research community. In: Proceedings IEEE/ACM Joint Conference on Digital Libraries (JCDL). pp. 57–66 (2014)

12. Harzing, A., Alakangas, S.: Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison. Scientometrics 106(2), 787–804 (Feb 2016)
13. Ke, Q., Ferrara, E., Radicchi, F., Flammini, A.: Defining and identifying sleeping beauties in science. Proceedings National Academy of Sciences 112(24), 7426–7431 (2015)
14. Ley, M.: Dblp: Some lessons learned. Proceedings VLDB Endowment 2(2), 1493–1500 (2009)
15. Mazloumian, A., Eom, Y., Helbing, D., Lozano, S., Fortunato, S.: How citation boosts promote scientific paradigm shifts and nobel prizes. PLOS ONE 6(5), 1–6 (2011)
16. Paparrizos, J., Gravano, L.: *k*-shape: Efficient and accurate clustering of time series. In: Proceedings ACM International Conference on Management of Data (SIGMOD). pp. 1855–1870 (2015)
17. Revesz, P.Z.: A method for predicting citations to the scientific publications of individual researchers. In: Proceedings 18th International Database Engineering & Applications Symposium (IDEAS). pp. 9–18 (2014)
18. Rochat, Y.: Closeness centrality extended to unconnected graphs: The harmonic centrality index. In: ASNA. No. EPFL-CONF-200525 (2009)
19. Sun, J., Min, C., Li, J.: A vector for measuring obsolescence of scientific articles. Scientometrics 107(2), 745–757 (2016)
20. Wildgaard, L., Schneider, J.W., Larsen, B.: A review of the characteristics of 108 author-level bibliometric indicators. Scientometrics 101(1), 125–158 (2014)
21. Wolcott, H.N., Fouch, M.J., Hsu, E.R., DiJoseph, L.G., Bernaciak, C.A., Corrigan, J.G., Williams, D.E.: Modeling time-dependent and-independent indicators to facilitate identification of breakthrough research papers. Scientometrics 107(2), 807–817 (2016)
22. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: Proceedings 4th ACM International Conference on Web Search and Data Mining (WSDM). pp. 177–186 (2011)