A new dynamic growth model reveals how citation networks evolve over time, pointing the way toward reformulated scientometrics.

BY TANMOY CHAKRABORTY, SUHANSANU KUMAR, PAWAN GOYAL, NILOY GANGULY, AND ANIMESH MUKHERJEE

On the Categorization of Scientific Citation Profiles in Computer Science

A CONSENSUS IN the literature is that the citation profiles of published articles follow a universal pattern—initial growth in number of citations the first two to three years following publication with a steady peak of one to two years and then decline over the rest of the lifetime of the article. This observation has long been the underlying heuristic in determining major bibliometric factors (such as quality of publication,

growth of scientific communities, and impact factor of publication venue). Here, we analyze a dataset of 1.5 million computer science papers maintained by Microsoft Academic Search, finding the citation count of the articles over the years follows a remarkably diverse set of patterns—a profile with an initial peak (PeakInit), with distinct multiple peaks (PeakMul) exhibiting a peak later in time (PeakLate) that is monotonically decreasing (MonDec), monotonically increasing (MonIncr), and cannot be categorized into any other category (Oth). We conducted a thorough experiment to investigate several important characteristics of the categories, including how individual categories attract citations, how categorization is influenced by year and publication venue, how each category is affected by self-citations, the stability of the categories over time, and how much each of the categories contribute to the core of the network. Further, we show the traditional preferential-attachment models fail to explain these citation profiles. We thus propose a novel dynamic growth model that accounts for both preferential attachment and the aging factor in order to replicate the real-world behavior of various citation profiles. This article widens the scope for a serious reinvestigation into the existing bibliometric indices for scientific research, not just for computer science.

» key insights

▪ **Analyzing a massive dataset of scholarly papers revealed six distinctive citation profiles for papers, ranging from a single peak to multiple peaks to peaks that increase monotonically or decrease over time.**

▪ **Following characterization of the profiles, we found major modifications of the existing bibliographic indices could better reflect real-world citation history.**

▪ **Unlike existing network-growth models, these profiles can be reproduced but only if they account for "preferential attachment" and "aging."**
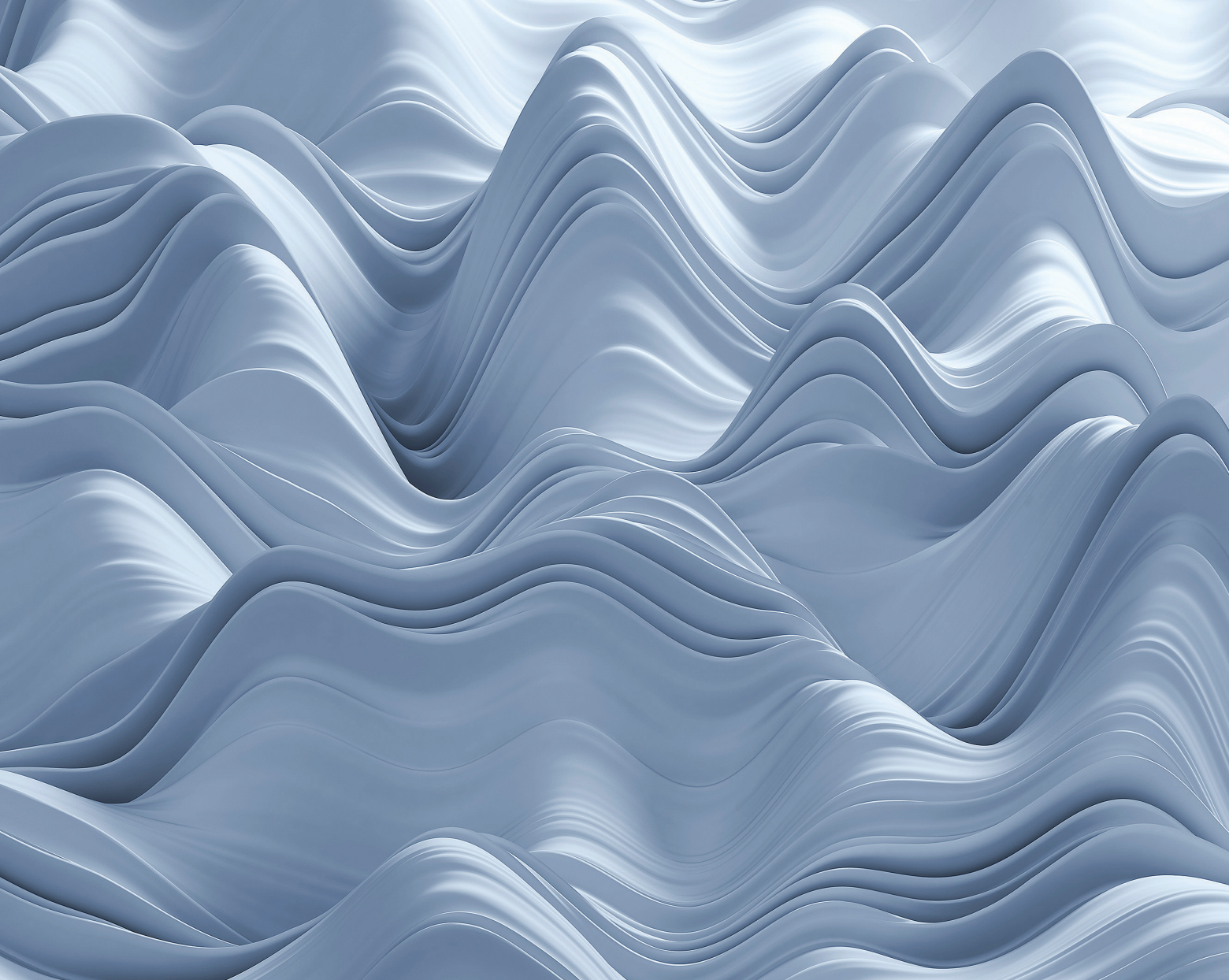
Quantitative analysis in terms of counting, measuring, comparing quantities, and analyzing measurements is perhaps the main tool for understanding the impact of science on society. Over time, scientific research itself (by recording and communicating research results through scientific publications) has become enormous and complex. This complexity is today so specialized that individual researchers' understanding and experience are no longer sufficient to identify trends or make crucial decisions. An exhaustive analysis of research output in terms of scientific publications is of great interest to scientific communities that aim to be selective, highlighting significant or promising areas of research and better managing scientific investigation.[5,24,25,27] Bibliometrics, or "scientometrics,"[3,22] or application of quantitative analysis and statistics to publications (such as research articles and accompanying citation counts), turns out to be the main tool for such investigation. Following pioneering research by Eugene Garfield,[14] citation analysis in bibliographic research serves as the fundamental quantifier for evaluating the contribution of researchers and research outcomes. Garfield pointed out a citation is no more than a way to pay homage to pioneers, give credit for related work (homage to peers), identify methodology and equipment, provide background reading, and correct one's own work or the work of others.[14]

A citation network represents the knowledge graph of science, in which individual papers are knowledge sources, and their interconnectedness in terms of citation represents the relatedness among various kinds of knowledge; for instance, a citation network is considered an effective proxy for studying disciplinary knowledge flow, is used to discover the knowledge backbone of a particular research area, and helps group similar kinds of knowledge and ideas. Many studies have been conducted on citation networks and their evolution over time. There is already a well-accepted belief among computer science scholars about the dynamics of citations a scientific article receives following publication: initial growth (growing phase) in number of citations within the first two to three years following publication, followed by a steady peak of one to two years (saturation phase), and then a final decline over the rest of the lifetime of the article (decline and obsolete phases) (see Figure 1).[15–17] In most cases, this observation is drawn from analysis of

Figure 1. Hypothetical example showing the traditional view by computer science scholars of the citation profiles of scientific papers following publication.
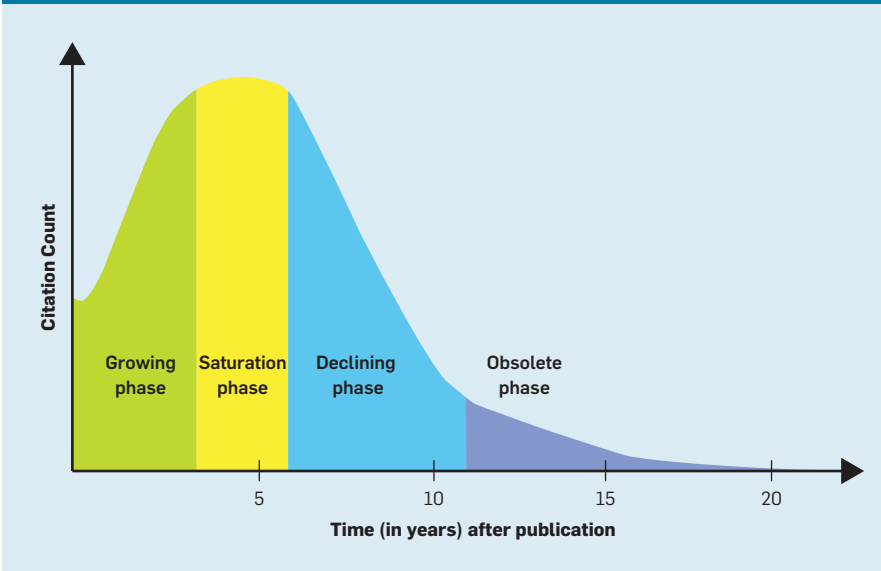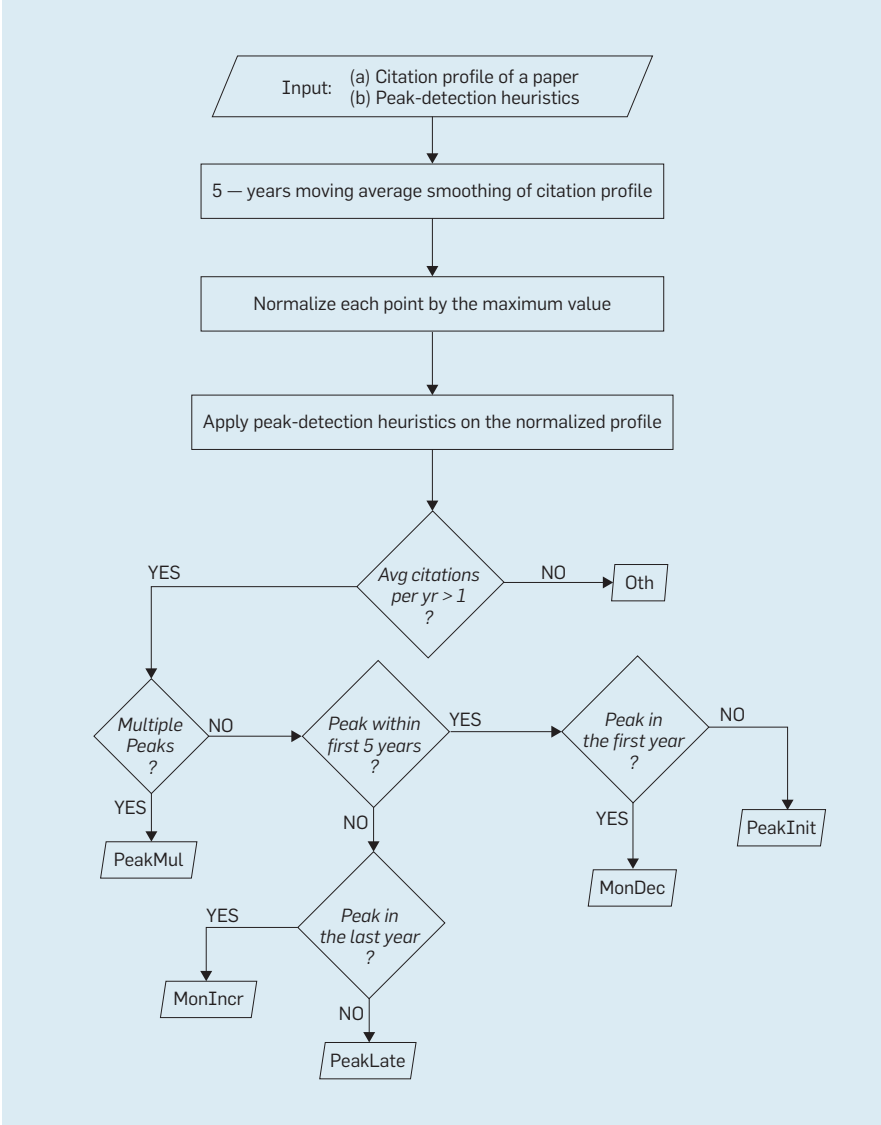

Figure 2. Systematic flowchart of the rules for classifying training samples.

a limited set of publication data,[7,13] possibly obfuscating some true characteristics. Here, we conduct our experiment on a massive bibliographic dataset in the computer science domain involving more than 1.5 million papers published worldwide by multiple journals and proceedings from 1970 to 2010 as collected by Microsoft Academic Search. Unlike earlier observations about paper citation profiles, we were able to define six different types of citation profiles prevalent in the dataset we label PeakInit, PeakMul, PeakLate, MonDec, MonIncr, and Oth. We exhaustively analyzed these profiles to exploit the microdynamics of how people actually read and cite the papers, controlling the growth of the underlying citation network unexplored in the literature. This categorization allows us to propose a holistic view of the growth of the citation network through a dynamic model that accounts for the accepted concept of preferential attachment,[1,2,26] along with the aging factor[20] in order to reproduce different citation profiles observed in the Microsoft Academic Search dataset. To the best of our knowledge, ours is the first attempt to consider these two factors together in synthesizing the dynamic growth process of citation profiles.

Our observations not only help reformulate the existing bibliographic indices (such as "journal impact factor") but enhance general bibliometric research (such as "citation link prediction," "information retrieval," and "self-citation characterization"), reflecting several characteristics:

Citation trajectory. In earlier studies, an article's citation trajectory was assumed by the research community to increase initially, then follow a downward trajectory;

Six trajectories. Analyzing the massive dataset of computer science papers, we identified six distinct citation trajectories; and

Revisit. Since citation profiles can be categorized into at least six different types, all measures of scientific impact (such as impact factor) should be revisited and updated.

Massive Publication Dataset
Most experiments in the literature

on analyzing citation profiles have worked with small datasets. In our experiment, we gathered and analyzed a massive dataset to validate our hypothesis. We crawled one of the largest publicly available datasets,[a] including, as of 2010, more than 4.1 million publications and 2.7 million authors, with updates added each week.[9] We collected all the papers published in the computer science domain and indexed by Microsoft Academic Search[b] from 1970 to 2010. The crawled dataset included more than two million distinct papers that were further distributed over 24 fields of computer science, as categorized by Microsoft Academic Search. Moreover, each paper included such bibliographic information as title, unique index, author(s), author(s) affiliation(s), year of publication, publication venue, related field(s), abstract, and keyword(s). In order to remove the anomalies that crept in due to crawling, we passed the dataset through a series of initial prepro-

cessing stages. The filtered dataset included more than 1.5 million papers, with 8.68% of them belonging to multiple fields, or "interdisciplinary" papers; the dataset is available at http:// cnerg.org (see "Resources" tab).

Categorization of Citation Profiles
Since our primary focus was analyzing a paper's citation growth following publication, we needed an in-depth understanding of how citation numbers vary over time. We conducted an exhaustive analysis of citation patterns of different papers in our dataset. Some previous experimental results[9,14] showed the trend followed by citations received by a paper following publication date is not linear in general; rather, there is a fast growth of citations within the first few years, followed by exponential decay. This conclusion is drawn mostly from analysis of a small dataset of the archive. Here, we first took all papers with at least 10 years and a maximum of 20 years of citation history, then followed a series of data-processing steps. First, to smooth the time-series data points in a paper's citation profile, we used five-year-moving-average filtering; we then

scaled the data points by normalizing them with the maximum value present in the time series, or maximum number of citations received by a paper in a particular year; finally, we ran a local-peak-detection algorithm[c] to detect peaks in the citation profile. We also applied two heuristics to specify peaks: the height of a peak should be at least 75% of the maximum peak-height, and two consecutive peaks should be separated by more than two years. Otherwise we treated them as a single peak (see Figure 2).

We found most papers did not follow the traditional citation profile mentioned in the earlier studies, as in Figure 1; rather, we identified the six different types of citation profiles based on the count and position of peaks in a profile. We defined six types of citation profiles, along with individual proportions of the entire dataset:

*PeakInit.* Papers with citation-count peaks in the first five years following

---

a http://academic.research.microsoft.com
b The crawling process took us approximately six weeks, concluding in August 2013.

c The peak detection algorithm is available as a Matlab Spectral Analysis http://www.mathworks.in/help/signal/ref/findpeaks.html; we used MINPEAKDISTANCE=2 and MINPEAKHEIGHT=0.75 and the default values for the other parameters.

Figure 3. Citation profiles for the first five categories we obtained from analyzing the Microsoft Academic Search citation dataset (top panel) and how it compared with the results obtained from the model (bottom panel).

Each frame corresponds to each category; Oth does not follow a consistent pattern and is not shown. In each frame, a citation belt is formed by the lines Q1 (green line) and Q3 (blue line) representing the first quartile (10% points are below the line) and third quartile (10% points are above the line) of data points, respectively, or effectively 80% points are within the citation belt. The red line within the citation belt represents the average behavior of all profiles corresponding to that category. In the top panel, for each category, one representative citation profile is shown at the middle of the belt (broken black line). The color coding in the bottom panel is similar to the top panel, though the broken lines are the results we obtained from our model.
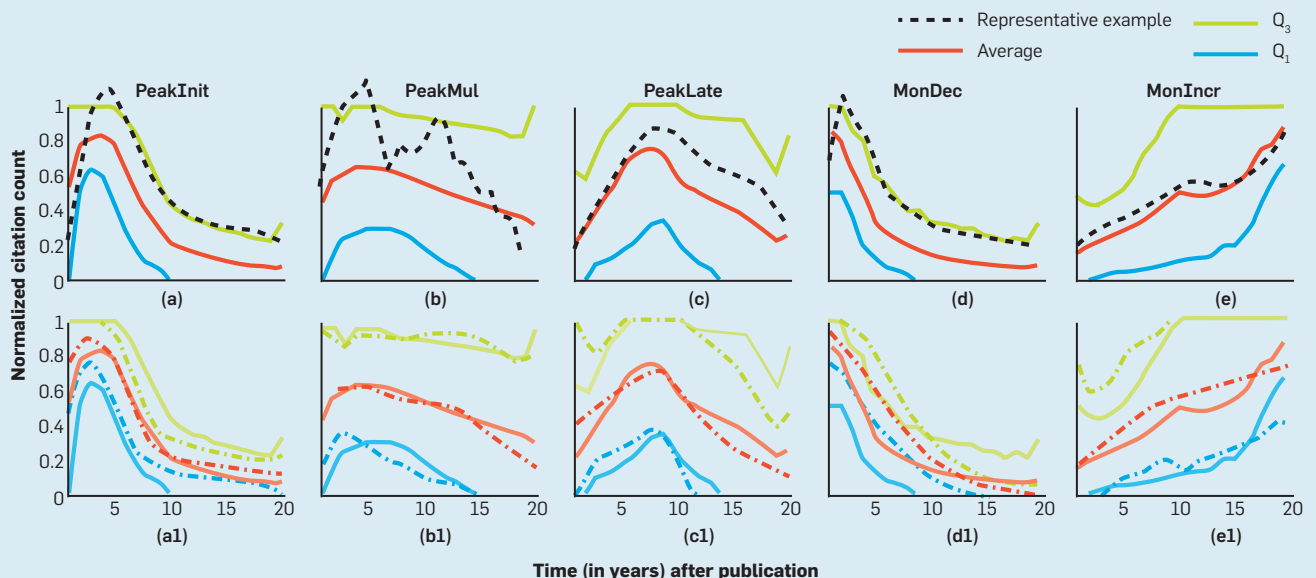
publication (but not the first year) followed by an exponential decay (proportion = 25.2%) (see Figure 3a);

PeakMul. Papers with multiple peaks at different time points in their citation profiles (proportion = 23.5%) (see Figure 3b);

PeakLate. Papers with few citations at the beginning, then a single peak after at least five years after publication, followed by an exponential decay in citation count (proportion = 3.7%) (see Figure 3c);

MonDec. Papers with citation-count peaks in the immediate next year of publication followed by monotonic decrease in the number of citations (proportion = 1.6%) (see Figure 3d);

MonIncr. Papers with monotonic increase in number of citations from the beginning of the year of publication until the date of observation or after 20 years of publication (proportion = 1.2%) (see Figure 3e); and

Oth. Apart from the first five types, a large number of papers on average receive fewer than one citation per year; for them, the evidence is not significant enough to assign them to one of the first five categories, so they remain as a separate group (proportion = 44.8%).

The rich metadata in the dataset further allowed us to conduct a second-level analysis of the categories for multiple research fields in computer science. We thus measured the percentage of papers in different categories for each of the 24 research fields after filtering out all papers in the Oth category. We noticed that while for all other fields, the largest fraction of papers belong to the PeakMul category, for the Web this fraction is maximum in the PeakInit category (see Figure 4). A possible reason could be since the Web is mostly a conference-based research field, the papers in PeakInit generally dominate the field, as discussed later, in light of three observations:

Web. Most Web-related papers fall into the PeakInit category;

MonDec. Simulation and computer education have the largest proportion of papers in the MonDec category, and bioinformatics and machine learning have the smallest; and

PeakLate. Security and privacy, as well as bioinformatics, have the largest proportion of papers in the PeakLate category, and simulation and the Web have the smallest.

Categories in Citation Ranges

One aspect of analyzing scientific publications is determining how acceptable they are to the research community. A paper's acceptability is often measured by raw citation count; the more citations an article receives from other publications, the more it is assumed to be admired by researchers and hence the greater its scientific impact.[6] In our context, we ask, which among the six categories includes papers admired most in terms of citations? To answer, we conducted a study in which we divided total citation range into four buckets (ranges 11–12, 13–15, 16–19, 20–11,408) such that each citation bucket included an almost equal number of papers. For a deeper analysis of the highest citation range, we further divided the last bucket (20–11,408) into four more ranges, obtaining seven buckets altogether. We then measured the proportion of papers contributed by a particular category to a citation bucket (see Figure 5). Note in each citation bucket, we normalized the number of papers contributed by a category by total number of papers belonging to that category. The figure is a histogram of conditional probability distribution, the probability a randomly selected paper falls in citation bucket i given that it belongs to category j. Normalization was required to avoid population bias across different categories. Note the higher citation range is occupied mostly by the papers in the PeakLate and MonIncr categories, followed by PeakMul and PeakInit. Also note the MonDec category, which has the smallest proportion in the last citation bucket, shows a monotonic decline in the fraction of papers as citation range increases. This initial evidence suggests a general and nonintuitive interpretation of citation profiles; if a paper does not attract a large number

Figure 4. Percentage of papers in six categories for various research fields in computer science; the pattern is generally consistent, except for World Wide Web.
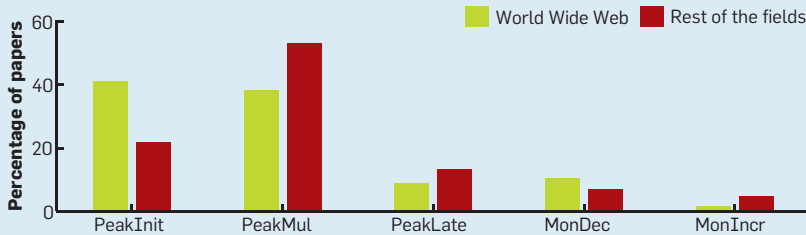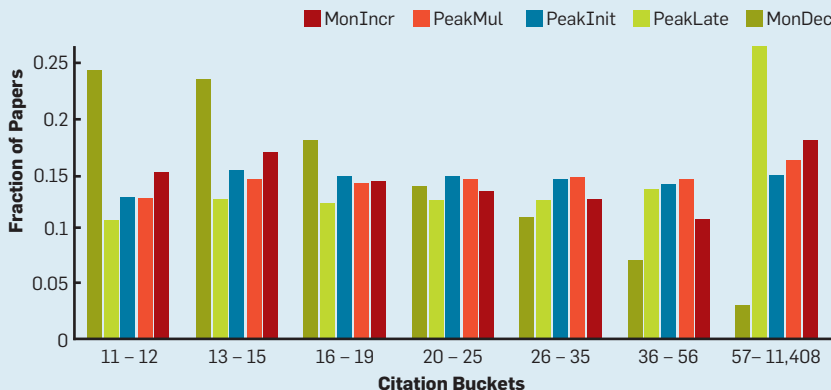

Figure 5. Contribution of papers from each category in different citation buckets. The entire range of citation value in the dataset is divided into seven buckets in which the contribution of papers from a particular category is normalized by the total number of papers in that category.

of citations within the immediate few years following its publication, it does not necessarily mean it will continue to be low impact through its lifetime; rather its citation growth rate might accelerate and could indeed turn out to be well accepted in the literature of computer science. We further explain this behavior in the next section.

Characterizing Different Citation Profiles

The rich metadata information in the publication dataset further allowed us to understand the characteristic features of each of the six categories at finer levels of detail.

Influence of publication year and venue on categorization. One might question whether this categorization could be influenced by the time (year) a paper is published; that is, papers published earlier might be following the well-known behavior, whereas papers published more recently might indicate a different behavior. To verify categorization is not biased by publication date, we measured the average year of publication of the papers in each category. Table 1, second column, suggests the citation pattern is not biased by year of publication, since average years correspond roughly to the same time period. On the other hand, the mode of publication in conferences differs significantly from that of journals, and the citation profiles of papers published in these two venues are expected to differ. To analyze venue effect on categorization, we measured the percentage of papers published in journals vis-à-vis in conferences for each category, as in Table 1, third and fourth columns, respectively. We observed while most of the papers in the PeakInit (64.35%) and MonDec (60.73%) categories were published in conferences, papers in Peak-Late (60.11%) and MonIncr (74.74%) were published mostly in journals. If a publication starts receiving more attention or citations later in its lifetime, it is more likely to have been published in a journal and vice versa, reflecting two trends:

Conferences. Due to increasing popularity of conferences in applied domains like computer science, conference papers get quick publicity within a few years of publication and likewise quick decay of that popularity; and

Journals. Though journal papers usually take time to be published and gain popularity, most journal papers are consistent in attracting citations, even many years after publication.

Another interesting point from these results is that although the existing formulation of journal impact factor[14] is defined in light of the citation profile, as in Figure 1, most journal papers in PeakLate or MonIncr do not follow such a profile at all; at least for papers in PeakLate, the metric does not focus on the most relevant timeframe of the citation profile (mostly the first five years after publication). In light of our results, the appropriateness of the formulation of bibliographic metrics (such as journal impact factor) are doubtful; for example, a journal's impact factor[15] at any given time is the average number of citations received per paper published during the two preceding years.

Effect of self-citation on categorization. Another factor often affecting citation rate is "self-citation,"[12] which can inflate the perception of an article's or a scientist's scientific impact, particularly when an article has many authors,

Table 1. Mean publication year Y (its standard deviation (σ(Y)) and the percentage of papers in conferences and journals for each category of a citation profile.

Category	Mean publication year (σ(Y))	% of conference papers	% of journal papers
PeakInit	1994 (5.19)	64.35	35.65
PeakMul	1991 (6.68)	39.03	60.97
PeakLate	1992 (6.54)	39.89	60.11
MonDec	1994 (5.44)	60.73	39.27
MonIncr	1993 (7.36)	25.26	74.74

Table 2. Confusion matrix representing the transition of categories due to the removal of self-citations.

A value x in the cell (i, j) represents x fraction of papers in category i would have been placed in category j if self-citations were absent from the entire dataset. Note no row is specified for the Oth category, as papers in this category never move to other categories through deletion of citations.

Category	PeakInit	PeakMul	PeakLate	MonDec	MonIncr	Oth
PeakInit	0.72	0.10	0.03	0.01	0	0.15
PeakMul	0.02	0.81	0.04	0	0.10	0.11
PeakLate	0.01	0.06	0.86	0	0.01	0.06
MonDec	0.05	0.14	0	0.41	0	0.35
MonIncr	0	0.02	0.01	0.01	0.88	0.09

Figure 6. Fraction of self-citations per paper in different categories over different time periods after publication and fraction of papers in each category migrating to the Oth category due to removal of self-citations, assuming different category thresholds.
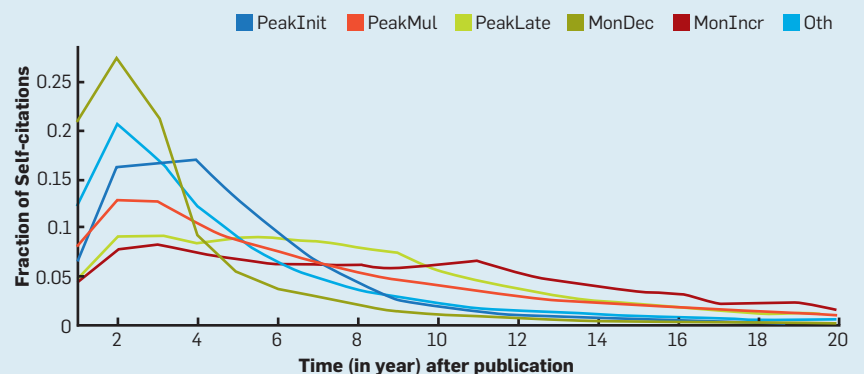
increasing the possible number of self-citations;[11,29] there have thus been calls to remove self-citations from citation-rate calculations.[29] We conducted a similar experiment to identify the effect of self-citation on the categorization of citation profiles. We first removed a citation from the dataset if the citing and the cited papers had at least one author in common, then measured the fraction of papers in each category migrating to some other category due to this removal. Table 2 is a confusion matrix, where labels in the rows and the columns represent the categories before and after removing self-citations, respectively.

Note papers in MonDec are strongly affected by the self-citation phenomenon. Around 35% of papers in MonDec would have been in the Oth category if not for self-citations. However, this percentage might be the result of the thresholding we impose, as discussed earlier, when categorizing papers; papers with fewer than or as many as 10 citations in the first 10 years following publication are considered to be in Oth category. Looking to verify the effect of thresholding on inter-category migration following removal of self-citations, we varied the category threshold from 10 to 14 and plotted

the fraction of papers in each category migrating to Oth due to our removal of self-citations (see Figure 6). The result agreed with the Table 2 observation that the MonDec category is most affected by self-citations, followed by PeakInit, PeakMul, and PeakLate. This result indicates the effect of self-citations is due to the inherent characteristics of each category, rather than to the predefined threshold setting of the category boundary, following three trends:

Authors. Authors tend to cite their own papers within two to three years of publication to increase visibility;

Conference papers. The MonDec and PeakInit categories, or mostly conference papers, are strongly affected by self-citation; and

Visibility. Self-citation is usually seen soon after publication in an attempt to increase publication visibility.

Figure 6 reflects how self-citations are distributed across different time periods for individual categories; we aggregated all self-citations and plotted the fraction of self-citations following publication. As expected, for the MonDec category we found most self-citations are "farmed" within two to three years of publication. A similar observation holds for both the PeakInit and Oth categories. Note, PeakInit and MonDec are composed mostly of conference papers. We conclude conference papers are the most affected by self-citations. However, the characteristics of the highly cited categories (such as MonIncr and PeakLate) are mostly consistent through the years, showing these categories are less dependent on self-citation.

Stability of Different Categories

The number of citations for a paper changes over time depending on the paper's effect on the scientific community that might change the shape of the citation profile. Studying the temporal evolution of each citation profile can help researchers understand the stability of the categories individually. Since we know the category of papers with at least 20 years of citation history, we further analyzed how the shape of the profile evolves over those 20 years. Following publication of a paper at time T, we identified its category at time T + 10, T + 15 and T + 20 based on the heu-

Figure 7. Alluvial diagram representing evolution of papers in different categories and the flows between categories at time T + 10, T + 15, and T + 20.

The colored blocks correspond to different categories. Block size indicates number of papers in a category, and the shaded waves joining the regions represent flow of papers between the regions, such that the width of the flow corresponds to the fraction of papers. The total width of incoming flows is equal to the width of the corresponding region.
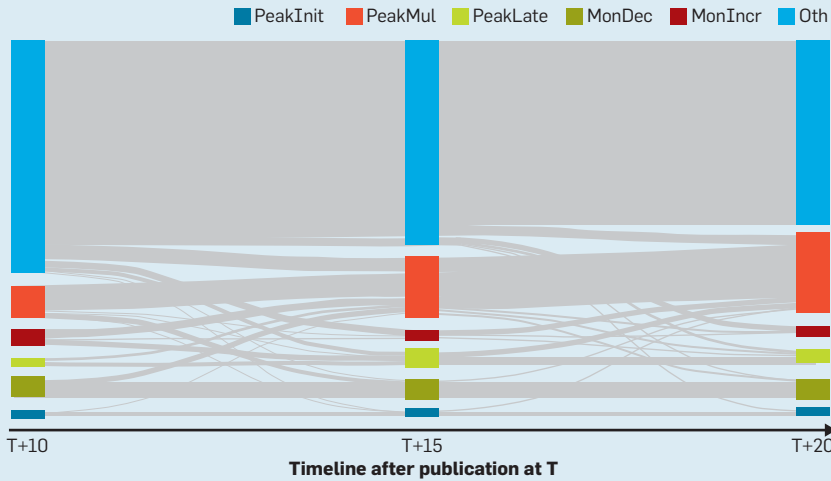

Figure 8. Multi-level pie chart for years 2000, 2004, 2007, and 2010, showing the composition of each category in different k^s-shell regions, where colors represent different categories and the area covered by each colored region in each k^s-shell represents the proportion of papers in the corresponding category in that shell.

The innermost shell is the core region, and the outermost shell is the periphery region. For better visualization of the different regions, we divided the total number of shells identified from the citation network in each year into six broad shells; the core-periphery structure in each year thus has six concentric layers.
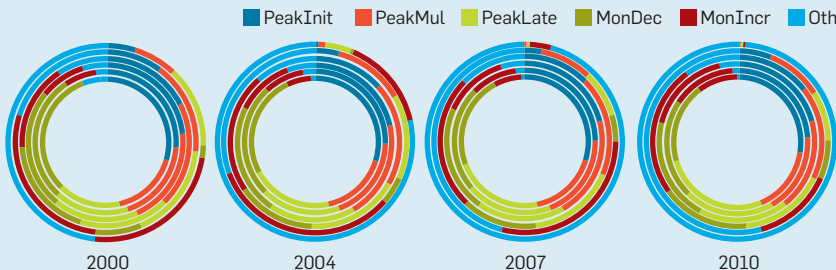
ristics discussed earlier. We hypothesize a stable citation category tends to maintain its shape over a paper's entire published timeline. The colored blocks of the alluvial diagram[28] in Figure 7 correspond to the different categories for three different timestamps. We observed that apart from the Oth category, which indeed includes a major proportion of all the papers in our dataset, MonDec seemed the most stable, followed by PeakInit. However, papers we assumed to belong in the Oth category quite often turned out to be MonIncr papers at a later time. This analysis demonstrates a systematic approach to explaining the transition from one category to another with increased numbers of citations.

Core-Periphery Analysis

Although Figure 5 indicates the influence of different categories in terms of raw citation count, it neither explains the significance of the papers in each category forming the core of the network nor gives any information regarding the temporal evolution of the structure. For a better and more detailed understanding of that evolution, we performed k-core analysis[8,21] on the evolving citation network by decomposing the network for each year into its k^s-shells such that an inner shell index of a paper reflects a central position in the core of the network.

We constructed a number of aggregated citation networks in different years—2000, 2004, 2007, and 2010—such that a citation network for year Y included the induced subgraph of all papers published at or before Y. For each such network, we then ran several methods, starting by recursively removing nodes with a single inward link until no such nodes remained in the network. These nodes form the 1-shell of the network, or k^s-shell index $k^s = 1$. Similarly, by recursively removing all nodes with degree 2, we derived the 2-shell. We continued increasing k until all nodes in the network were assigned to one of the shells. The union of all shells with index greater than or equal to k^s is the k^s-core of the network, and the union of all shells with index smaller or equal to k^s is the k^s-crust of the network. The idea is to show how the papers in each category (identified in 2000) migrate from one shell to an-

Since our primary focus was analyzing a paper's citation growth following publication, we needed an in-depth understanding of how citation numbers vary over time.

other after attracting citations over the next 10 years. It also allowed us to observe the persistence of a category in a particular shell.

In Figure 8, most papers in the Oth category are in the periphery and their proportion in the periphery increases over time, indicating they are increasingly less popular over time. The Peak-Mul category gradually leaves the peripheral region over time and mostly occupies the two innermost shells. PeakInit and MonDec show similar behavior, with the largest proportion of papers in inner cores in the initial year but gradually shifting toward peripheral regions. On the other hand, MonIncr and PeakLate showed the expected behavior, with their proportions increasing in the inner shells over time, indicating rising relevance over time. This helped us identify the temporal evolution of the importance of different categories in terms of how each of them contributes to the central position of a citation network.

Dynamic Growth Model

Extensive research has gone toward developing growth models to explain evolution of citation networks;[19,30] for example, models like those from Barabási-Albert[1,2] and Price[26] attempt to generate scale-free networks through a preferential-attachment mechanism. Most such work seeks to explain the emergence of a network's degree distribution. Here, we propose a novel dynamic growth model to synthesize the citation network, aiming to reproduce the citation categories seen in the Microsoft Academic Search dataset. To the best of our knowledge, this model is the first of its kind to take into account preferential attachment[1] and aging[18,20] to mimic real-world citation profiles.

As input to the model for comparing against the real-world citation profiles, we used the following distributions: number of papers over the years (to determine the number and type of papers entering the system at each time step) and reference distribution (to determine outward citations from an incoming node). At each time step (corresponding to a particular year), we selected a number of nodes (papers) with outdegree (references) for each, as determined preferentially from the refer-

ence distribution. We then assigned the vertex preferentially to a particular category based on the size of the categories (number of papers in each) at that time step. To determine the other end point of each edge associated with the incoming node, we first selected a category preferentially based on the in-citation information of the category, then selected within the category a node (paper) preferentially based on its attractiveness. We determined attractiveness by time elapsed since publication (aging) and number of citations accumulated till that time (preferential attachment). Note the formulation of the attractiveness in our model also varies for different categories.

We found remarkable resemblance between real-world citation profiles and those obtained from the model in Figure 3, bottom panels. Each frame of the figure includes three lines depicting first quartile (10% points below this line), third quartile (10% points above this line), and mean behavior. We also compared the in-degree distributions obtained from the model and from the real dataset for different categories, observing a significant resemblance. Our model thus reflects a holistic view of the evolution of a citation network over time, along with the intra- and inter-category interactions that account for the observable properties of the real-world system.

Conclusion

Access to a massive computer science bibliographic dataset from Microsoft Academic Search made it possible for us to conduct an exhaustive analysis of citation profiles of individual papers and derive six main categories not previously reported in the literature. At the micro-level, we provide a set of new approaches to characterize each individual category, as well as the dynamics of its evolution over time. Leveraging these behavioral signatures, we were able to design a novel dynamic model to synthesize the network evolving over time. The model in turn revealed citation patterns of different categories, showing significant resemblance to what we obtained from the real data.

This article thus offers a first step toward reformulating the existing quantifiers available in scientomet-

rics to leverage different citation patterns and formulate robust measures. Moreover, a systematic machine-learning model of the behavior of different citation patterns has the potential to enhance standard research methodology, including discovering missing links in citation networks,[10] predicting citations of scientific articles,[31] predicting high-impact and seminal papers,[23] and recommending scientific articles.[4]

In future work, we plan to extend our study to the datasets of other domains, possibly physics and biology, to verify the universality of our categorizations. We are also keen to understand the micro-level dynamics controlling the behavior of the PeakMul category, which is significantly different from the other four. One initial observation in this direction is that PeakMul behaves like an intermediary between PeakInit and PeakLate. Also in future work, we would like to understand different characteristic features, particularly for PeakMul. **C**

References
1. Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics 74*, 1 (Jan. 2002), 47–97.
2. Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (Oct. 1999), 509–512.
3. Bensman, S.J. The evaluation of research by scientometric indicators. *Journal of the Association for Information Science and Technology 62*, 1 (Jan. 2011), 208–210.
4. Bogers, T. and van den Bosch, A. Recommending scientific articles using CiteULike. In *Proceedings of the Second ACM Conference on Recommender Systems* (Lausanne, Switzerland, Oct. 23–25). ACM Press, New York, 2008, 287–290.
5. Boonchom, S., Nuchwana, L., and Amorn, M. The development of standards, factors, and indicators for evaluating the quality of classroom action research. *Procedia - Social and Behavioral Sciences 69* (Dec. 2012), 220–226.
6. Bornmann, L. and Daniel, H.-D. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation 64*, 1 (Jan.-Feb. 2008), 45–80.
7. Callaham, M., Wears, R.L., and Weber, E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association 287*, 21 (June 2002), 2847–2850.
8. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., and Shir, E. A model of Internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences 104*, 27 (July 2007), 11150–11154.
9. Chakraborty, T., Sikdar, S., Tammana, V., Ganguly, N., and Mukherjee, A. Computer science fields as ground-truth communities: Their impact, rise, and fall. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (Niagara Falls, Canada, Aug. 25–28). ACM Press, New York, 2013, 426–433.
10. Clauset, A., Moore, C., and Newman, M.E.J. Hierarchical structure and the prediction of missing links in networks. *Nature 453*, 7191 (May 2008), 98–101.
11. Della Sala, S. and Brooks, J. Multi-authors' self-citation: A further impact factor bias? *Cortex 44*, 9 (Oct. 2008), 1139–1145.
12. Fowler, J. and Aksnes, D. Does self-citation pay?

Scientometrics 72, 3 (Sept. 2007), 427–437.
13. Garfield, E. Which medical journals have the greatest impact? *Annals of Internal Medicine 105*, 2 (Aug. 1986), 313–20.
14. Garfield, E. The impact factor and using it correctly. *Der Unfallchirurg 6*, 101 (June 1998), 413–414.
15. Garfield, E. Journal impact factor: A brief review. *Canadian Medical Association Journal 161*, 8 (Oct. 1999), 979–980.
16. Garfield, E. Impact factors, and why they won't go away. *Nature 411*, 6837 (May 2001), 522.
17. Garfield, E. The history and meaning of the journal impact factor. *Journal of the American Medical Association 295*, 1 (Jan. 2006), 90–93.
18. Gingras, Y., Larivier, V., Macaluso, B., and Robitaille, J.-P. The effects of aging on researchers' publication and citation patterns. *PLoS ONE 3*, 12 (Dec. 2008), e4048.
19. Golosovsky, M. and Solomon, S. Stochastic dynamical model of a growing citation network based on a self-exciting point process. *Physical Review Letters 109*, 098701 (Aug. 2012).
20. Hajra, K.B. and Sen, P. Aging in citation networks. *Physica A 346*, 1 (Feb. 2005), 44–48.
21. Harris, J., Hirst, J., and Mossinghoff, M. *Combinatorics and Graph Theory. Springer Undergraduate Texts in Mathematics and Technology.* Springer-Verlag, New York, 2008.
22. Lariviere, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C.R. Bibliometrics: Global gender disparities in science. *Nature 504*, 7479 (Dec. 2013), 211–213.
23. McNamara, D., Wong, P., Christen, P., and Ng, K.S. Predicting high impact academic papers using citation network features. In *Proceedings of the Pacific-Asia Conference in Knowledge Discovery and Data Mining Workshop* (Gold Coast, Australia, Apr. 14–17). Springer, Berlin, Heidelberg, 2013, 14–25.
24. Meho, L.I. The rise and rise of citation analysis. *Physics World 1*, 20 (Jan. 2007), 32–36.
25. Peritz, B.C. On the objectives of citation analysis: Problems of theory and method. *Journal of the Association for Information Science and Technology 43*, 6 (July 1992), 448–451.
26. Price, D.D.S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology* (Sept. 1976), 292–306.
27. Radicchi, F. and Fortunato, S., Castellano, C. Universality of citation distributions: Towards an objective measure of scientific impact. *Proceedings of the National Academy of Sciences 105*, 45 (Nov. 2008), 17268–17272.
28. Rosvall, M. and Bergstrom, C.T. Mapping change in large networks. *PLoS ONE 5*, e8694 (Jan. 2010).
29. Schreiber, M. Self-citation corrections for the Hirsch index. *Europhysics Letters 78*, 3 (May 2007), 1–6.
30. Sen, P. Directed accelerated growth: Application in citation network. *Physica A 346*, 1–2 (Feb. 2005), 139–146.
31. Yan, R., Huang, C., Tang, J., Zhang, Y., and Li, X. To better stand on the shoulder of giants. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (Washington, D.C., June 10–14). ACM Press, New York, 2012, 51–60.

Tanmoy Chakraborty (its_tanmoy@cse.iitkgp.ernet. in) is a Google India Ph.D. fellow in the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India.

Suhansanu Kumar (suhansanu.kumar@cse.iitkgp.ernet.in) is a Ph.D. student in the Department of Computer Science at the University of Illinois, Urbana-Champaign, IL.

Pawan Goyal (pawang@cse.iitkgp.ernet.in) is an assistant professor in the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India.

Niloy Ganguly (niloy@cse.iitkgp.ernet.in) is a professor in the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India.

Animesh Mukherjee (animeshm@cse.iitkgp.ernet.in) is an assistant professor in the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India, and a Simons Associate in the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy.