Understanding Popularity of Academic Entities: From Papers to Authors to Venues

Mayank Singh
Dept. of Computer Science and Engg.
mayank.singh@cse.iitkgp.ernet.in

ABSTRACT

With the exponential growth of research volume in the recent decades, academic entities like articles, authors, venues, organisations, fields etc. have evolved qualitatively and quantitatively. The scientific community has always been demanding for better algorithms, metrics and features for ranking and categorization of academic entities leading to one of the interesting and well researched problem of understanding and estimating the popularity of these academic entities. We study several interesting factors that influence the popularity of research articles. Specifically, we utilize information generated immediately after the publication to estimate its long-term popularity. This generated information includes both network-based and content-based information. We also propose the first plausible network-driven models for obsolescence in the context of research paper citations, based on a natural notion of relay-linking. Our model is based on a surprising inversion or undoing of triangle completion, where an old node relays a citation to a younger follower in its immediate vicinity. We show that our proposed models remarkably better fit with real bibliographic data. We also demonstrate the development of *ConfAssist* which is a novel conflict resolution framework that can assist experts to resolve conflicts in deciding whether a conference is a top-tier or not by expressing how (dis)similar the conference is to other well accepted top-tier/ non top-tier conferences.

CCS CONCEPTS

• **Information systems → Digital libraries and archives**; **Data analytics**;

KEYWORDS

Long-term scientific impact, early citers, growth models, venue categorization

1 INTRODUCTION

Success of academic entities like research papers, authors, publication venues, organizations, etc. is estimated by their scientific impact. Quantifying scientific impact through citation counts or metrics [3, 10–12] has received much attention in the last two decades. This is primarily owing to the exponential growth in the literature volume requiring the design of efficient impact metrics for policy making concerning with recruitment, promotion and funding of faculty positions, fellowships etc. Although these approaches are quite popular, they appear to be highly debatable [13, 16]. Additionally, the previous works fail to take into account the future accomplishments of a researcher, article or venue. Prediction of future citation counts is an extremely challenging task because of the nature and dynamics of citations [6, 23, 29]. Recent advancement in prediction of future citation counts has led to the development of complex mathematical and machine learning based models. The existing supervised models have employed several paper, venue and author centric features that can be obtained at the publication time.

Similarly, with rapidly growing publication repositories, understanding the networked process of obsolescence is as important to the emerging field of *academic analytics*[1] as understanding the rise to prominence (popularity). Parolo *et al.* [9] present evidence that it is becoming "increasingly difficult for researchers to keep track of all the publications relevant to their work", which can lead to reinventions, redundancies, and missed opportunities to connect ideas. On the other hand, Verstak *et al.* [25] claim that fear of evanescence is misplaced, and that older papers account for an increasing fraction of citations as time passes. Chakraborty *et al.* [7] present a nuanced analysis that naturally clusters papers into the ephemeral and the enduring. This gives hope that not all creativity is lost in the sands of time; but neither do older papers capture all our attention.

On a similar note, publication venue (Conference or Journal) ranking adds another dimension to estimation of scientific impact. Venue ranking/categorization has always been considered highly debatable. Different organizations, researchers and forums provide different rankings[2,3,4,5,6] and categorization of venues[7,8,9,10,11]. These systems have several limitations. First, most of the existing systems provide category-based rankings, but no clear demarcation between these categories. Second, existing systems that provide category based classification fail to provide the main intuitions behind such classification. Third, ranking systems use h-index[12,13] and impact factor based metrics [14,15], which in turn are very debatable [16, 17, 19, 30]. Fourth, almost all such systems are domain dependent [5].

The current thesis attempts to understand the underlying factors of popularity of academic entities. We utilize two open source computer science datasets, both crawled from the Microsoft Academic

[1]https://en.wikipedia.org/wiki/Academic_analytics
[2]http://scholar.google.co.in.
[3]http://academic.research.microsoft.com/.
[4]http://arnetminer.org.
[5]http://www.scimagojr.com/journalrank.php.
[6]http://admin-apps.webofknowledge.com/JCR/JCR.
[7]http://www.ntu.edu.sg/home/assourav/crank.htm.
[8]http://webdocs.cs.ualberta.ca/ zaiane/htmldocs/ConfRanking.html.
[9]http://perso.crans.org/ genest/conf.html.
[10]http://portal.core.edu.au/conf-ranks/.
[11]http://dsl.serc.iisc.ernet.in/publications/CS_ConfRank.htm.
[12]http://scholar.google.co.in.
[13]http://academic.research.microsoft.com/.
[14]http://www.cs.iit.edu/ xli/CS-Conference-Journals-Impact.htm.
[15]http://www.scimagojr.com/journalsearch.php?q=conference.

Search (MAS)[16]. First dataset (bibliographic dataset) was crawled by Chakraborty et al. [6] for a similar prediction work. The dataset consists of bibliographic information of more than 2.4 million papers, such as, the title, the abstract, the keywords, its author(s), the affiliation of the author(s), the year of publication, the publication venue, and the references. Second dataset (citation context dataset) was prepared by Singh et al. [23]. This dataset consists of more than 26 million citation contexts, pre-processed and annotated with the cited and the citing paper information.

2 RESEARCH OVERVIEW AND WORK DONE

2.1 The role of citation context in predicting long-term citation profiles

The impact and significance of a scientific publication is measured mostly by the number of citations it accumulates over the years. Early prediction of the citation profile of research articles is a significant as well as a challenging problem. In this work, we argue that features gathered from the **citation contexts** of the research papers can be very relevant for citation prediction. Citation context refers to textual descriptions of a given scientific paper found in other papers in the document collection which cites it [2]. A citation context is, in principle, a set of sentences where a paper is referred to. The intuition behind using the citation context features comes from the hypothesis that citation contexts reflect the opinion of the scientific community about the particular work. We show that even using some very simplistic features extracted from the citation context can boost the performance of a citation prediction system significantly.

Towards this objective, we extract two features from the citation contexts – average countX (number of times a paper is cited within the same article, averaged over all the citing papers) and average citeWords (number of words within the citation context, averaged over all the citing papers). We investigate whether the average countX and average citeWords values over the years are correlated with the number of citations a paper receives. We reiterate that both average countX and average citeWords are normalized with respect to the number of citations received by the paper. We divide the set of papers in our dataset into six buckets based on the number of citations. For each of the citation buckets, we observe the temporal profile for the average countX values, averaged for all the papers within that bucket (see Figure 1). Interestingly, as per our hypothesis, various citation ranges show differences in terms of the average countX values. Some important observations from Figure 1 are:

(1) There is an increase in value of countX in initial years irrespective of the citation bucket, and it further decreases continuously over the years. A slight increase is observed for the 10^{th} year after publication.
(2) Highly cited papers are cited more number of times in a single paper.

We also divided the set of papers into six citation categories mentioned in [7]. Figure 2 presents the temporal profile of average countX values for each of these 6 categories. The countX values can well discriminate between the six citation categories identified.

[16]http://academic.research.microsoft.com

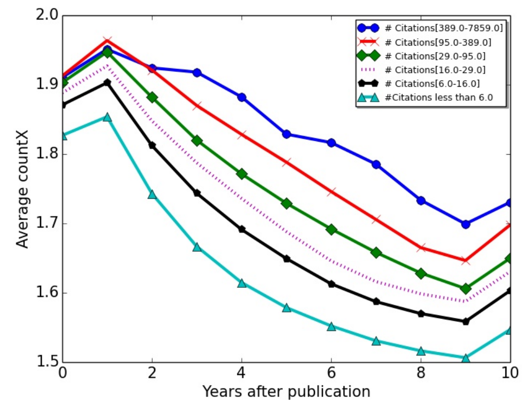

Figure 1: Average countX: temporal profiles for six citation buckets over the publication age

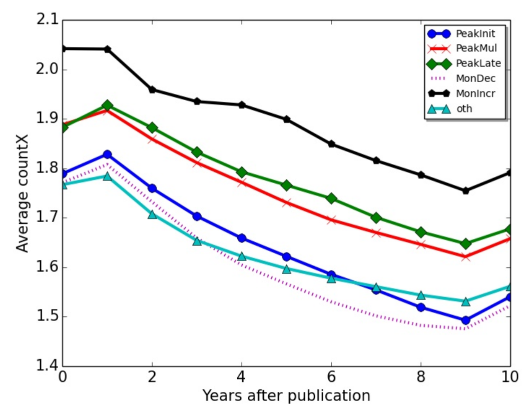

Figure 2: Average countX: temporal profiles for the six citation categories [7] over the publication age

We observe similar trends for average citeWords for both citation buckets and citation categories. Average citeWords also show quite discriminative and exhibit different trends not only for different citation ranges but also for the citation categories.

We then append these features along with various other features in an earlier framework based on *stratified learning* [6]. Experimental results show that the proposed model significantly outperforms the existing citation prediction models by a margin of 8-10% on an average under various experimental settings. Specifically, these features help in predicting the long-term scientific impact (*LTSI*) of the research papers. We would like to stress here that this study brings forth the tremendous potential of the content of a scientific article in predicting future citation counts; the huge success of only two very simple content related features proposed here makes the authors believe that deeper analysis of the content can lead to further significant improvements in the related areas of research. **This work was presented at conference on Information and Knowledge Management (CIKM 2015) [23].**

2.2 Understanding the Impact of Early Citers on Long-Term Scientific Impact

This work explores an interesting new dimension to the challenging problem of predicting long-term scientific impact (*LTSI*) usually measured by the number of citations accumulated by a paper in the long-term. It is well known that early citations (within 1–2 years after publication) acquired by a paper positively affects its *LTSI*. However, there is no work that investigates if the set of authors who bring in these early citations to a paper also affect its *LTSI*. In this paper, we demonstrate for the first time, the impact of these authors whom we call *early citers* (EC) on the *LTSI* of a paper.

We identify two distinct categories of EC – we call those authors who have high overall publication or citation count in the dataset as *influential* and the rest of the authors as *non-influential*. We investigate three characteristic properties of EC to better understand their complex nature and their influence on *LTSI*:

- **Popularity**: It is measured by the citation counts of early citers.
- **Productivity**: It is measured by the publication counts of early citers.
- **Connectivity**: It is measured by the collaboration distance between the authors of candidate papers and its early citers.

We presented an extensive analysis of how each category correlates with *LTSI* in terms of these properties. This study illustrates the fact that influential EC negatively affect the long-term citations. A plausible explanation could be that in general, researchers tend to cite works written by influential authors. Therefore, once an influential author cites an article, researchers tend to cite the influential author's paper, instead of the original paper. The attention from the original paper moves to the paper written by the influential citer toward the very beginning of the life-span of the original paper. Therefore, instead of flourishing, the long term citation count of the original paper gets negatively affected. This phenomenon of attention relaying from the less popular article to the more popular article is described as *attention stealing* [28]. In case of non-influential EC, the citation count of the candidate paper exhibits a positive correlation with PC. However, with the passage of time, this positive correlation diminishes due to ageing effect associated with paper's life span [26]. In case of influential EC, same ageing effect leads to increase in the negative correlation over the passage of time. Figure 3 represents correlation between EC *Productivity* and candidate paper's cumulative citation count at five later time periods after publication. *Productivity* and *Popularity* both follow similar trends as above.

In the same lines, *Connectivity* also shows similar trends as the other two properties (see Figure 4). The most striking observation from this property is the effect of immediate co-authors on *LTSI*. Even though, both influential or non-influential immediate co-authors maximally correlate with *LTSI*, influential immediate co-authors negatively affect the citation of the candidate paper in the long term due to intensified *attention stealing* effect. Informally stating, *Connectivity* reveals that this stealing effect is more profound if an EC is nearer to the authors of the paper being investigated.

Motivated by above empirical observations, we incorporate the EC properties in a well recognized citation prediction framework [29]. Our citation prediction framework employs a set of features that can be computed at the time of publication plus a set of features that can be extracted from the citation information generated within two years after publication We show that incorporating EC properties in the state-of-the-art supervised citation prediction models leads to high performance margins. We also built an online portal to visualize EC statistics along with the prediction results for a given query paper. The portal is accessible online at: http://www.cnergres.iitkgp.ac.in/earlyciters/. **This work was presented at JCDL 2017 [22].**

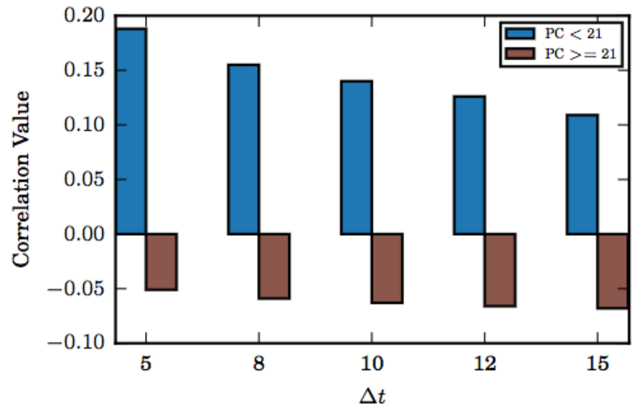

Figure 3: (Color online) Correlation between EC *Productivity* (measured by publication count) and candidate paper's cumulative citation count at five later time periods after publication, $\Delta t = 5, 8, 10, 12, 15$. Papers with lower value of $PC(< 21)$ exhibit positive correlation diminishing over the time. Papers with high value of $PC(>= 21)$ show an opposite trend. The overall separation decreases over time.

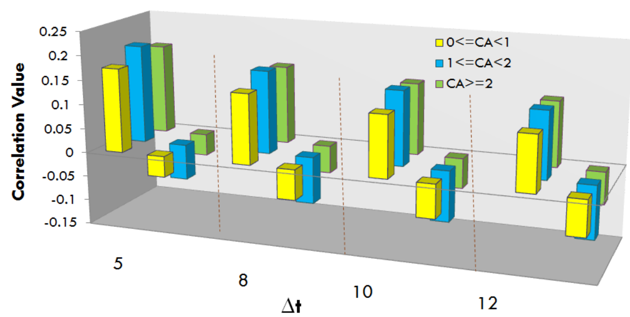

Figure 4: (Color online) Correlation between EC's publication count and cumulative citation count for three co-authorship buckets at four later time periods after publication, $\Delta t = 5, 8, 10, 12$. For each time period, first three bars represent correlation for non-influential EC ($PC_P < 21$) whereas the next three bars represent correlation for influential EC ($PC_P >= 21$). Influential immediate co-authors (Bucket 2) seem to badly affect the citation of the candidate paper P in the long term.

2.3 Relay-Linking Models for Prominence and Obsolescence in Academic Networks

How do actors in a social network pass from prominence to obsolescence and obscurity? Is ageing intrinsic, or informed and influenced by the local network around actors? And how does the ageing process affect properties of social networks, specifically, the tension between entrenchment of prominence (aka "rich gets richer" or the Matthew effect) vs. obsolescence? These are fundamental questions for any evolving social network, but particularly well-motivated in bibliometry.

So where does reality lie between entrenchment and obsolescence? Chakraborty *et al.* [7] present a nuanced analysis that naturally clusters papers into the ephemeral and the enduring. This gives hope that not all creativity is lost in the sands of time; but neither do older papers capture all our attention. Others [26, 27] model ageing as intrinsic to a paper, reducing the probability of citing it as it ages, but do not prescribe where the diverted citations end up.

In an interesting work on explaining ageing by attention stealing, Waumans *et al.* [28] present several evidences of attention stealing from parent paper by child paper. They show that the ArXiV[17] article entitled "Notes on D-Branes" [18] published in the year 1996 start losing its citation in the very next year (1997). The reason for attention stealing is attributed to four papers that cite [18] and go further on the same topic. In an another example, paper titled "Theory of Bose-Einstein condensation in trapped gases" [8] from the American Physical Society dataset[18] suffers from similar stealing effect. This paper starts losing attention to its three child papers six years after publication. In all the three cases, the title clearly indicates the scientific content continuity in the child paper. We attempt to understand and model this complex behavior of citation dynamics. Our specific contributions are summarized as:

2.3.1 Reconciling obsolescence vs. entrenchment. We propose several measurements on evolving networks that constitute a *temporal bucket signature* summarizing the coexistence between entrenchment and obsolescence. **Temporal bucket signature** denotes a stacked histogram of the relative age of target papers cited in a source paper (see Figure 5). Natural social networks (e.g., various research communities) show diverse and characteristic temporal bucket signatures. Surprisingly, many standard models of network evolution — and even obsolescence — fail to fit the temporal signatures of real bibliometric data. We establish this with temporal bucket signatures and two associated novel measures: **distance** and **turnover**. We also propose **age gap count histograms** to represent citation age distribution. Similar to temporal bucket signature, standard models fail to fit age gap count histogram of real data as well. We establish this fitness using another novel metric termed as **farness**. We proved that models with $O(1)$ parameters find it very challenging to pass all these stringent tests for temporal fidelity.

2.3.2 Insufficiency of intrinsic obsolescence. Albert and Barabasi's remarkable scale-free model (preferential attachment or **PA**) [1] "explained" power law degrees, but failed to simulate many other natural properties, such as bipartite communities. The "**copying model**" [15] gave a better power law fit and explained bipartite communities. Wang *et al.* [27] propose that the probability of citing paper p at time t is proportional to the product $k_p(t)e^{-\lambda(t-b_p)}$, where $k_p(t)$ is the number of citations p has at time t, b_p is its birth epoch, and λ is a global decay parameter. To our surprise, this model improve only modestly upon PA or copying models at matching age gap count histograms and temporal bucket signatures. A more sophisticated model by Wang *et al.* [26] involves three model parameters η_p, μ_p, σ_p *per paper*. In effect, this model is just a reparameterization to achieve *data collapse* [4] — collapsing apparently diverse citation trajectories into one standard function of age. We hypothesize that the reason is that ageing papers lose probability of getting cited, but the model do not *use the graph structure* to predict where these citations are likely to be redistributed.

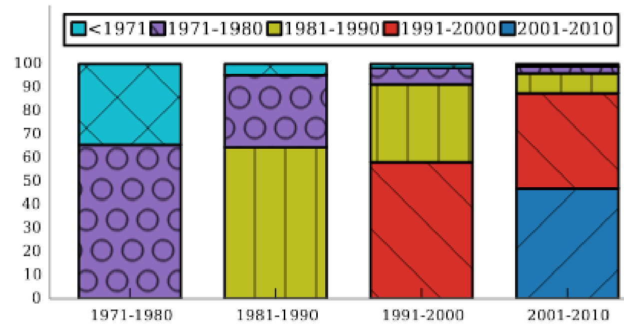

Figure 5: Temporal bucket signature representing citation distribution across 10-year buckets for computer science dataset. Each vertical bar represents a decade of papers. Within each bar, colored/textured segments represent the fraction of citations going to preceding decades. The bottommost segment is to the same decade, the second from bottom to the previous decade, etc. On one hand, the volume of citations to the current decade (bottommost segment) is shrinking to accommodate "old classics" (entrenchment). On the other hand, any given color/texture shrinks dramatically over decades (most papers fade away).

2.3.3 Triad uncompletion and relay-linking. Triad completion (*viz.*, if links (u, v) and (v, w) are present, consider adding (u, w)) has long been established [14] as a cornerstone of link prediction. The above observations led us to look for the *reverse micro-dynamic pattern*: whether a popular older paper p_0, at a given time, starts losing citations in favor of a newer paper p_1 citing p_0. Of course, we may only get to see the final decision to cite p_1 and not the process of "dropping" p_0. Therefore, it is a delicate process to tease apart such "relaying" (from p_0 to p_1) effects from myriad other reasons for increase or decrease in popularity. But we succeeded in designing high-precision filters that gathered strong circumstantial evidence that this effect is real..

This study led to a family of **relay-linking** models that are the central contribution of this work, roughly speaking: to add a citation in a new paper, choose an existing paper p_0, but if it is too old, walk back along a citation link to p_1 and (optionally) repeat the process. We call this hypothesized process *triad uncompletion* and the

[17] http://www.cs.cornell.edu/projects/kddcup/datasets.html
[18] http://journals.aps.org/datasets

associated generative model *relay-linking*. This new family of frugal ageing models has no per-node parameters and only 1–2 global parameters. Despite very few parameters, the new family of models show remarkably better fit with real data. **This work was presented at international conference on Knowledge Discovery and Data Mining (KDD 2017) [24].**

2.4 Conflict resolution framework for conference categorization

Classifying publication venues into top-tier or non top-tier is quite subjective and can be debatable at times. In this work, we propose *ConfAssist*, a novel assisting framework for conference categorization that aims to address the limitations in the existing systems and portals for venue classification.

A common belief in the research community is that researchers are confident about the category of the conference in the area of their expertise. We perform four small experiments to refute this intuition. Interestingly, in one among many of the experiments conducted, we refuted claims of Vasilescu et al. (2014) that observed strong negative linear correlation, suggesting that conferences with higher acceptance rates indeed have lower scientific impact. We observe that, it is not always true that all the top-tier conferences have low acceptance rate, and non top-tier conferences have high acceptance rate. There are many cases where clear demarcation of acceptance rate between top-tier and non top-tier is not found (see Figure 6).

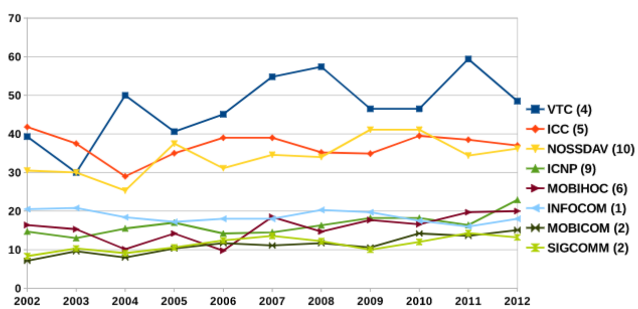

Figure 6: Acceptance rate for the top ten computer networks conferences over the years. Number inside brackets represents rank of the conference assigned by Microsoft academic search. Two conferences in top five, namely, ICC and VTC have high acceptance rate (~30). VTC (rank=4) has significantly higher acceptance rate (> 37%) than NOSSDAV (rank=10). Similarly, acceptance rate of ICNP (rank=9) is significantly low (~15). Note that two conferences (IPSN and SenSys) are not present due to unavailability of data.

While there are many clear cases where expert agreement can be almost immediately achieved as to whether a conference is a top-tier or not, there are equally many cases that can result in a conflict even among the experts. *ConfAssist* tries to serve as an aid in such cases by increasing the confidence of the experts in their decision. We start with the hypothesis that top-tier conferences are much more stable than other conferences and the inherent dynamics of these groups differs to a very large extent. We identify various features related to the stability of conferences that might help us separate a top-tier conference from the rest of the lot. The features have been grouped in two main categories; features based on diversity pattern in the accepted papers and features based on the co-authorship network of authors of the accepted papers. Accordingly, we identified nine distinct quantities, that gave us 27 different features (mean, median and standard deviation for each of the quantities). Figure 7 presents the comparison between these categories using four representative stability features, ΔCRDI, ΔCADI, ΔEDI and ΔABC's average and standard deviation profiles. X-axis denotes 11 consecutive year-differences and y-axis denotes the mean of the difference values with error bars showing standard deviation of the difference values. The corresponding values for the top-tier and non top-tier conferences are plotted using green and blue bars respectively. An analysis of these plots gives a clear indication that for all the four example quantities, the blue bars are higher than the green bars, i.e. there is a higher fluctuation for the non top-tier conferences (yearwise differences denoted by the height of blue bar are higher) as compared to the top-tier conferences (yearwise differences denoted by the green bar are lower).

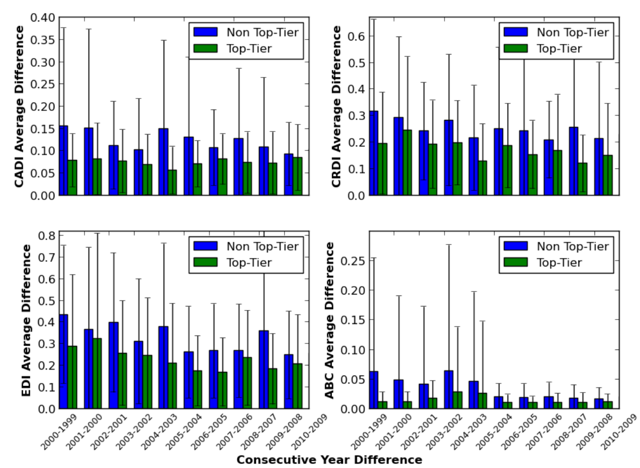

Figure 7: Comparison between top-tier and non top-tier using ΔCRDI, ΔCADI, ΔEDI and ΔABC's average and standard deviation profiles. X-axis denotes 11 consecutive year-differences and y-axis denotes the mean of the difference values across various conferences in a category, with error bars showing standard deviation of the difference values.

An analysis of 110 conferences from 22 sub-fields of computer science clearly favors our hypothesis as the top-tier conferences are found to exhibit much less fluctuations in the stability related features than the non top-tier ones. We evaluate our hypothesis using systems based on conference categorization. For the evaluation, we conducted human judgment survey with 28 domain experts. The results are impressive with 85.18% classification accuracy. We also presented comparison of dynamics of a new conference with matured top-tier and non top-tier conferences, which confirms that the proposed features can help in obtaining some initial signals of future popularity of the new conference. The system is applicable to any conference with atleast 5 years of publication history.

This work was presented as a poster in JCDL 2015 [20] and the extended version was published in Journal of Informetrics [21].

3 FUTURE DIRECTION AND EXPECTED CONTRIBUTION

Understanding the effect of citation context features of early citers on Long-Term Scientific Impact: Our previous works described in Section 2.1 and 2.2 have thoroughly studied the problem of estimation of Long-Term Scientific Impact. In future, we aim to study the distinguishing behavior of citation contexts generated from influential and non-influential early citers. We hope that similar to previously defined three properties of early citers, the citation context based properties would further extend the prediction results. We also plan to understand the effect of early citers on six citation profiles described in [7]. Further, the six citation profiles can be incorporated in the flat prediction model to develop a stratified early citer based citation prediction model.

Modeling early citers influence in relay-linking models: Relay-linking models described in Section 2.3 entirely depends on the citation count of research papers. Motivated by the study of influence of early citers on $LSTI$ (described in Section 2.2), in future, we plan to bring the author popularity aspect. More specifically, we plan to model early citers‘ influence in gaining and losing citations under relay-linking framework.

REFERENCES

[1] Réka Albert and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of modern physics* 74, 1 (2002), 47.

[2] Bader Aljaber, Nicola Stokes, James Bailey, and Jian Pei. 2010. Document clustering of scientific texts using citation contexts. *Information Retrieval* 13, 2 (2010), 101–131.

[3] Carl T Bergstrom, Jevin D West, and Marc A Wiseman. 2008. The Eigenfactor? metrics. *The Journal of Neuroscience* 28, 45 (2008), 11433–11434.

[4] Somendra M Bhattacharjee and Flavio Seno. 2001. A measure of data collapse for scaling. *J. Physics A: Mathematical and General* 34, 33 (2001), 6375. http://arxiv.org/pdf/cond-mat/0102515v2.pdf

[5] Lutz Bornmann and HansâĂŘDieter Daniel. 2008. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* 64, 1 (2008), 45–80.

[6] Tanmoy Chakraborty, Suhansanu Kumar, Pawan Goyal, Niloy Ganguly, and Animesh Mukherjee. 2014. Towards a Stratified Learning Approach to Predict Future Citation Counts. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '14)*. IEEE Press, 351–360.

[7] Tanmoy Chakraborty, Suhansanu Kumar, Pawan Goyal, Niloy Ganguly, and Animesh Mukherjee. 2015. On the Categorization of Scientific Citation Profiles in Computer Science. *Commun. ACM* 58, 9 (Aug. 2015), 82–90. DOI:http://dx.doi.org/10.1145/2701412

[8] Franco Dalfovo, Stefano Giorgini, Lev P Pitaevskii, and Sandro Stringari. 1999. Theory of Bose-Einstein condensation in trapped gases. *Reviews of Modern Physics* 71, 3 (1999), 463.

[9] P. Della Briotta Parolo, R. K. Pan, R. Ghosh, B. A. Huberman, K. Kaski, and S. Fortunato. 2015. Attention decay in science. *ArXiv e-prints* (March 2015). arXiv:physics.soc-ph/1503.01881

[10] Leo Egghe. 2006. Theory and practise of the g-index. *Scientometrics* 69, 1 (2006), 131–152.

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

[12] Jorge E Hirsch. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America* (2005), 16569–16572.

[13] Jorge E Hirsch and Gualberto Buela-Casal. 2014. The meaning of the h-index. *International Journal of Clinical and Health Psychology* 14, 2 (2014), 161–164.

[14] Petter Holme and Beom Jun Kim. 2002. Growing scale-free networks with tunable clustering. *Phys. Rev. E* 86 (2002), 026107–(1–5).

[15] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D Sivakumar, Andrew Tomkins, and Eli Upfal. 2000. Random graph models for the web graph.. In *FOCS*. 57–65.

[16] Cyril Labbé. 2010. Ike Antkare one of the great stars in the scientific firmament. *ISSI newsletter* 6, 2 (2010), 48–52.

[17] Bertrand Meyer, Christine Choppy, Jørgen Staunstrup, and Jan van Leeuwen. 2009. Viewpoint research evaluation for computer science. *Commun. ACM* 52, 4 (2009), 31–34.

[18] Joseph Polchinski, Shyamoli Chaudhuri, and Clifford V Johnson. 1996. Notes on D-branes. *arXiv preprint hep-th/9602052* (1996).

[19] Cagan H. Sekercioglu. 2008. Quantifying Coauthor Contributions. *Science* 322, 5900 (2008), 371.

[20] Mayank Singh, Tanmoy Chakraborty, Animesh Mukherjee, and Pawan Goyal. 2015. ConfAssist: A Conflict Resolution Framework for Assisting the Categorization of Computer Science Conferences. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '15)*. ACM, New York, NY, USA, 257–258. DOI:http://dx.doi.org/10.1145/2756406.2756963

[21] Mayank Singh, Tanmoy Chakraborty, Animesh Mukherjee, and Pawan Goyal. 2016. Is this conference a top-tier? ConfAssist: An assistive conflict resolution framework for conference categorization. *Journal of Informetrics* 10, 4 (2016), 1005 – 1022. DOI:http://dx.doi.org/10.1016/j.joi.2016.08.001

[22] Mayank Singh, Ajay Jaiswal, Priya Shree, Arindam Pal, Animesh Mukherjee, and Pawan Goyal. 2017. Understanding the Impact of Early Citers on Long-Term Scientific Impact. In *Digital Libraries (JCDL), 2017 ACM/IEEE Joint Conference on*. IEEE, 1–10.

[23] Mayank Singh, Vikas Patidar, Suhansanu Kumar, Tanmoy Chakraborty, Animesh Mukherjee, and Pawan Goyal. 2015. The Role Of Citation Context In Predicting Long-Term Citation Profiles: An Experimental Study Based On A Massive Bibliographic Text Dataset. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1271–1280.

[24] Mayank Singh, Rajdeep Sarkar, Pawan Goyal, Animesh Mukherjee, and Soumen Chakrabarti. 2017. Relay-Linking Models for Prominence and Obsolescence in Evolving Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1077–1086.

[25] Alex Verstak, Anurag Acharya, Helder Suzuki, Sean Henderson, Mikhail Iakhiaev, Cliff Chiung-Yu Lin, and Namit Shetty. 2014. On the Shoulders of Giants: The Growing Impact of Older Articles. *CoRR* abs/1411.0275 (2014). http://arxiv.org/abs/1411.0275

[26] Dashun Wang, Chaoming Song, and Albert-László Barabási. 2013. Quantifying long-term scientific impact. *Science* 342, 6154 (2013), 127–132.

[27] Mingyang Wang, Guang Yu, and Daren Yu. 2009. Effect of the age of papers on the preferential attachment in citation networks. *Physica A: Statistical Mechanics and its Applications* 388, 19 (2009), 4273 – 4276. DOI:http://dx.doi.org/10.1016/j.physa.2009.05.008

[28] Michaël Charles Waumans and Hugues Bersini. 2016. Genealogical trees of scientific papers. *PloS one* 11, 3 (2016), e0150588.

[29] Rui Yan, Congrui Huang, Jie Tang, Yan Zhang, and Xiaoming Li. 2012. To better stand on the shoulder of giants. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*. ACM, 51–60.

[30] Chun-Ting Zhang. 2009. A proposal for calculating weighted citations based on author rank. *EMBO reports* 10, 5 (2009), 416–417. DOI:http://dx.doi.org/10.1038/embor.2009.74