Journal of Informetrics 11 (2017) 608–621

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

The time dimension of science: Connecting the past to the future

Yian Yin[a,b], Dashun Wang[a,b,c,*]

[a] Northwestern Institute on Complex Systems and Data Science (NICO), Northwestern University, Evanston, IL 60208, USA
[b] McCormick School of Engineering and Applied Sciences, Northwestern University, Evanston, IL 60208, USA
[c] Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 4 January 2017
Accepted 17 April 2017
Available online 15 May 2017

ABSTRACT

A central question in science of science concerns how time affects citations. Despite the long-standing interests and its broad impact, we lack systematic answers to this simple yet fundamental question. By reviewing and classifying prior studies for the past 50 years, we find a significant lack of consensus in the literature, primarily due to the coexistence of retrospective and prospective approaches to measuring citation age distributions. These two approaches have been pursued in parallel, lacking any known connections between the two. Here we developed a new theoretical framework that not only allows us to connect the two approaches through precise mathematical relationships, it also helps us reconcile the interplay between temporal decay of citations and the growth of science, helping us uncover new functional forms characterizing citation age distributions. We find retrospective distribution follows a lognormal distribution with exponential cutoff, while prospective distribution is governed by the interplay between a lognormal distribution and the growth in the number of references. Most interestingly, the two approaches can be connected once rescaled by the growth of publications and citations. We further validate our framework using both large-scale citation datasets and analytical models capturing citation dynamics. Together this paper presents a comprehensive analysis of the time dimension of science, representing a new empirical and theoretical basis for all future studies in this area.

1. Introduction

The increasing availability of large-scale datasets that capture major activities in science has created an unprecedented opportunity to explore broad and important patterns underlying the scientific enterprise, catalyzing in a drastic fashion a recent multidisciplinary shift in our quantitative understanding of science (Evans, 2008; Evans & Foster, 2011; Guevara, Hartmann, Aristarán, Mendoza, & Hidalgo, 2016; Guimera, Uzzi, Spiro, & Amaral, 2005; Hirsch, 2005; Jones, Wuchty, & Uzzi, 2008; Ke, Ferrara, Radicchi, & Flammini, 2015; Newman, 2009; Petersen et al., 2014; Radicchi, Fortunato, & Castellano, 2008; Redner, 2005; Sinatra, Deville, Szell, Wang, & Barabási, 2015; Sinatra, Wang, Deville, Song, & Barabási, 2016; Stringer, Sales-Pardo, & Amaral, 2008; Szántó-Várnagy, Pollner, Vicsek, & Farkas, 2014; Uzzi, Mukherjee, Stringer, & Jones, 2013; Wang, Song, & Barabási, 2013). Nowhere are these new advances more apparent than in the studies of citations (Barabási, Song, & Wang, 2012; Bornmann & Daniel, 2008; Eom & Fortunato, 2011; Moreira, Zeng, & Amaral, 2015; Radicchi et al., 2008; Redner, 2005;

* Corresponding author at: Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA
E-mail address: dashun.wang@northwestern.edu (D. Wang).

http://dx.doi.org/10.1016/j.joi.2017.04.002
1751-1577/

Uzzi et al., 2013; Wang et al., 2013), owing largely to their widespread applications, from science policy to promotions (Lane, 2010; Radicchi, Fortunato, Markines, & Vespignani, 2009; Shen & Barabási, 2014), hiring (Clauset, Arbesman, & Larremore, 2015; Duch et al., 2012), assignment of grants (Bromham, Dinnage, & Hua, 2016; Lane & Bertuzzi, 2011; Li & Agha, 2015) and prizes (Mazloumian, Eom, Helbing, Lozano, & Fortunato, 2011). The aim of this paper is to carry out a comprehensive analysis on one of the most fundamental dimensions of citations: Time.

Time plays a central role in science. Indeed, while some papers stay relevant and continue to dominate the scientific discourse over a long period, most papers unfortunately 'die out', after collecting their fair share of citations (Stringer, Sales-Pardo, & Amaral, 2010; Wang et al., 2013). Understanding the time dimension is critical for a wide range of reasons. It not only helps us understand the rise and fall of scientific paradigms (Kuhn, 1962), tracing knowledge horizons and hotspots in science and technology (Mukherjee, Romero, Jones, & Uzzi, 2017; Orosz, Farkas, & Pollner, 2016; Sinatra et al., 2015), it is also critical for allocating investment (Bromham et al., 2016; Ma, Mondragón, & Latora, 2015; Szell & Sinatra, 2015), identifying crucial yet initially unrecognized sleeping beauties (Ke et al., 2015; Van Raan, 2004), assessing and even predicting future impact of inventions and discoveries (Acuna, Allesina, & Kording, 2012; Newman, 2014; Redner, 1998; Sinatra et al., 2016; Uzzi et al., 2013; Wang et al., 2013). This has become ever more so with the increasing scale, cost and complexity of science (Šubelj & Fiala, 2016; Börner, Maru, & Goldstone, 2004; Michels & Schmoch, 2012; Van Noorden, Maher, & Nuzzo, 2014). Indeed, the exponential growth of science (Price de Solla, 1963; Sinatra et al., 2015; Van Noorden et al., 2014) suggests that there is now much more work for scientists to learn from, build upon, and cite, which is further exacerbated by intensifying specialization in science and engineering disciplines (Jones, 2011) and the inevitable dominance of interdisciplinary research spanning institutional and international boundaries (Adams, 2013; Deville et al., 2014; Evans & Reimer, 2009; Guimera et al., 2005; Jones et al., 2008). Furthermore, because citation systems are commonly treated as models of complex interconnected systems, understanding the time dimension of citations will also help deepen our quantitative understanding of complex systems by tightening models and observations that are highly generalizable to broad areas.

It is, therefore, not surprising that this question has been extensively investigated over the past several decades (Avramescu, 1979; Börner et al., 2004; Bouabid, 2011; Burrell, 2002; Burton & Kebler, 1960; de Solla Price, 1965; Egghe & Rao, 1992; Evans, 2008; Garfield & Sher, 1963; Glänzel, 2004; Golosovsky & Solomon, 2014; Gupta, 1997; Krauze & Hillinger, 1971; Larivière, Archambault, & Gingras, 2008; MacRae, 1969; Margolis, 1967; Matricciani, 1991; Motylev, 1989; Nakamoto, 1988; Pan, Petersen, Pammolli, & Fortunato, 2016; Parolo et al., 2015; Pollman, 2000; Redner, 2005; Sanyal, 2006; Simkin & Roychowdhury, 2007; Stinson & Lancaster, 1987; Tsay, 1999; Verstak et al., 2014; Wang et al., 2013), being one of most prolific lines of inquiry in science of science studies. Yet, despite its broad impact and rich historical context, we lack systematic answers to a simple yet fundamental question: How does time affect citations? In this paper, we systematically investigate this question by leveraging large-scale citation datasets and recent models capturing citations dynamics. We start by conducting a comprehensive review of existing literature, which reveals a significant lack of consensus on this matter. The main reason of this lack of consensus is that empirical measurements of temporal effect in citations have taken two related yet distinct measurement approaches (Fig. 1A). The first approach considers papers that are cited by a publication and analyzes *retrospectively* the age distribution of these references (Avramescu, 1979; Burrell, 2002; Burton & Kebler, 1960; de Solla Price, 1965; Egghe & Rao, 1992; Garfield & Sher, 1963; Glänzel, 2004; Golosovsky & Solomon, 2014; Gupta, 1997; Krauze & Hillinger, 1971; Larivière et al., 2008; MacRae, 1969; Margolis, 1967; Matricciani, 1991; Motylev, 1989; Nakamoto, 1988; Pan et al., 2016; Pollman, 2000; Redner, 2005; Stinson & Lancaster, 1987; Tsay, 1999; Verstak et al., 2014). The second approach, in contrast, studies *prospectively* the age distribution of citations that are gained over time by a paper (Avramescu, 1979; Bouabid, 2011; Glänzel, 2004; Golosovsky & Solomon, 2014; Krauze & Hillinger, 1971; MacRae, 1969; Motylev, 1989; Nakamoto, 1988; Parolo et al., 2015; Redner, 2005; Sanyal, 2006; Simkin & Roychowdhury, 2007; Stinson & Lancaster, 1987; Wang et al., 2013). The subtle difference between the two approaches creates a dramatic yet largely understudied effect in temporal citation patterns (Table 1), further confounded by the exponential growth in both the number of papers and references cited by them (Pan et al., 2016; Sinatra et al., 2015; Van Noorden et al., 2014).

The temporal behavior of citations has been measured and reported by independent research groups, each using specific datasets and measurement details (Table 1). The lack of consensus, and the coexistence of the two approaches, raise an important question: What is the most appropriate functional form describing the time dimension of science? Here, by introducing a general theoretical framework, we provide systematic answers to this question, which is then validated both analytically and empirically through citation models and large-scale datasets. As such, the paper makes several contributions to this prolific direction in the science of science. First and foremost, it serves as a detailed and much needed survey by reviewing and repeating results from most major studies in this domain. We then used the Web of Science (WoS) dataset to systematically measure and test results obtained by prior studies. We examine these results in the context of growth in science, systematically testing the most appropriate model describing aging measured by the prevailing two approaches. More importantly, despite the concomitant development of both approaches in the literature, we lack any known connections between the two. Here we introduced a general framework that helps us uncover that the two different approaches are connected through precise mathematical relationships, allowing us to derive and even predict one approach from the other. We further derived our empirical results using citation models, providing theoretical support for the observed temporal behavior. We conclude by a brief discussion on scaling relationships between aging structure and citation impact of papers. Together, this paper presents a comprehensive study on the time dimension of science, providing a new empirical and theoretical basis for all further studies on this topic. These results are not only relevant for the emerging field of the science