Journal of Informetrics 12 (2018) 656–678

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

The memory of science: Inflation, myopia, and the knowledge network

Raj K. Pan[a,1], Alexander M. Petersen[b,c,*,1], Fabio Pammolli[d,e], Santo Fortunato[f,g,**]

[a] Department of Computer Science, Aalto University School of Science, P.O. Box 15400, FI-00076, Finland
[b] Ernest and Julio Gallo Management Program, University of California, Merced, CA 95343, United States
[c] Department of Management of Complex Systems, School of Engineering, University of California, Merced, CA 95343, United States
[d] Department of Management, Economics, and Industrial Engineering, Politecnico di Milano, Milan 20156, Italy
[e] CADS, Center for Analysis, Decisions, and Society, Human Technopole, Milan 20157, Italy
[f] Center for Complex Networks and Systems Research, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA
[g] Indiana University Network Science Institute (IUNI), Indiana University, Bloomington, IN, USA

ARTICLE INFO

Article history:
Received 12 September 2017
Received in revised form 12 June 2018
Accepted 12 June 2018

Keywords:
Citation network
Reference distance
Models of science
Attention economy
Monte Carlo simulation
Citation inflation

ABSTRACT

Scientific production is steadily growing, exhibiting 4% annual growth in publications and 1.8% annual growth in the number of references per publication, together producing a 12-year doubling period in the total supply of references, i.e. links in the science citation network. This growth has far-reaching implications for how academic knowledge is connected, accessed and evaluated. Against this background, we analyzed a citation network comprised of 837 million references produced by 32.6 million publications over the period 1965–2012, allowing for a detailed analysis of the 'attention economy' in science. Our results show how growth relates to 'citation inflation', increased connectivity in the citation network resulting from decreased levels of uncitedness, and a narrowing range of attention – as both very classic and very recent literature are being cited increasingly less. The decreasing attention to recent literature published within the last 6 years suggests that science has become stifled by a publication deluge destabilizing the balance between production and consumption. To better understand these patterns together, we developed a generative model of the citation network, featuring exponential growth, the redirection of scientific attention via publications' reference lists, and the crowding out of old literature by the new. We validate our model against several empirical benchmarks, and then use perturbation analysis to measure the impact of shifts in citing behavior on the synthetic system's properties, thereby providing insights into the functionality of the science citation network as an infrastructure supporting the memory of science.

* Corresponding author at: Ernest and Julio Gallo Management Program, University of California, Merced, CA 95343, United States.
** Corresponding author at: Center for Complex Networks and Systems Research, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA.
E-mail addresses: petersen.xander@gmail.com (A.M. Petersen), santo.fortunato@gmail.com (S. Fortunato).
[1] These authors contributed equally.

https://doi.org/10.1016/j.joi.2018.06.005

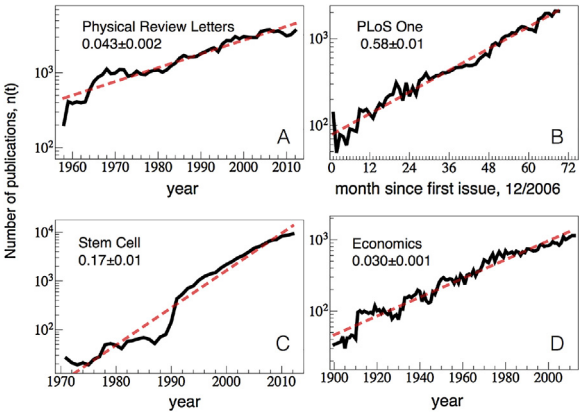

Fig. 1. Empirical growth of scientific output: journals, fields (A, B) Growth in two specific journals. Physical Review Letters (physics), and PLOS ONE (multidisciplinary open access). The remarkable growth rate for PLOS ONE – representative of the shift in scientific publication towards large online-only journals – dwarfs the others by an entire order of magnitude, with the exponential growth rate $g_n = 0.58$ corresponding to an annual growth rate $100(\mathrm{Exp}[0.58]-1) = 77.6\%$ and a publication doubling rate of $\ln 2/0.58 = 1.2$ years. (C, D) Growth in two specific research domains, Stem Cell research and Economics.

1. Introduction

Driven by public and private sector investment into people and projects (Stephan, 2012; Stokes, 1997), the rate of scientific production has exhibited persistent growth over the last century (Althouse, West, Bergstrom, & Bergstrom, 2009; Lariviere, Archambault, & Gingras, 2008). However, the extant empirical and theoretical literature provides little guidance for understanding the impact of these long-term growth trends on the structural properties of the science citation network. Only recently have studies shown that accounting for the growth of science can drastically change measured trends, e.g. in the decay rate of the citation life cycle of individual publications (Parolo et al., 2015). Insights such as this call for a better understanding of how the scientific attention economy (Franck, 1999; Klamer & van Dalen, 2002) is impacted by growth of the scientific system. Moreover, the recent proliferation of a new ecosystem of rapid-publication online-only "mega-journals" (Bjork, 2015; Petersen, 2018; Solomon, 2014; Solomon & Bjork, 2012; Wakeling et al., 2016) has further tipped the balance towards production over consumption of new research, making it a relevant and pressing issue.

To grasp the impact of this new publishing paradigm, consider the sole contribution by the online-only journal PLOS ONE, which grew over its first 6 years at an annual growth rate of 78.6%, corresponding to a doubling rate of only 1.2 years. To place this growth in real terms, after just 5 years since its inception, publications by PLOS ONE in 2012 (more than 23,000 articles) accounted for 1.4% of the entire volume of 2012 items indexed by the Web of Science (WOS) *Science Citation Index Expanded*; and calculated over the entire period 1900–2012, PLOS ONE publications represented 0.12% of all articles in this WOS index. Compared with other journals and fields in Fig. 1, which show typical growth around 3–4% annually, and even breakthrough fields such as Stem Cell research, which has grown at a thriving annual rate of 18.5%, the emergence of mega-journals appear ready to sustain long-term trends in the growth of scientific production well into the 21st century.

Applying methods from network science, complex systems, and data analytics, the 'science of science' (Fortunato et al., 2018) uses the millions of new research outputs produced each year by scientists around the world to illuminate knowledge production and innovation processes, thereby aiming to provide valuable insights for science policy guidance (Fealing, 2011). One particular link to the past that is preserved within each publication is the bibliographic list of references, which provide a means to measure how much today's research builds upon yesteryear's. As such, the citation network – where nodes are publications and links are the references within a publication to prior literature – has been used to conceptualize and measure the processes underlying the evolution of the scientific enterprise for more than half a century (de Solla Price, 1965; Garfield, 1955), and continues to be useful for making important insights into the long-term evolution of the scientific enterprise (Fortunato et al., 2018; Sinatra, Deville, Szell, Wang, & Barabasi, 2015).

Against this background we analyze the interplay between publication output growth and the attention to prior literature captured by the citation network. More specifically, we simultaneously identify and model three key features of this complex adaptive system:

1. The steady growth of the total number of references produced each year, arising from increasing publication output and reference-list lengths, and its relation to *citation inflation*.
2. The subsequent shifts in the concentration of citations received by publications at the lower and upper extremes of the citation distribution, providing perspective on *citation inequality*.
3. The distribution of references backwards in time which captures the historical breadth of *scientific attention*.