Attention decay in science

Pietro Della Briotta Parolo[a], Raj Kumar Pan[a,*], Rumi Ghosh[b],
Bernardo A. Huberman[c], Kimmo Kaski[a], Santo Fortunato[a]

[a] *Complex Systems Unit, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland*
[b] *Robert Bosch LLC, Palo Alto, CA 94304, USA*
[c] *Mechanisms and Design Lab, Hewlett Packard Enterprise Labs, Palo Alto, CA, USA*

A R T I C L E I N F O

A B S T R A C T

The exponential growth in the number of scientific papers makes it increasingly difficult for researchers to keep track of all the publications relevant to their work. Consequently, the attention that can be devoted to individual papers, measured by their citation counts, is bound to decay rapidly. In this work we make a thorough study of the life-cycle of papers in different disciplines. Typically, the citation rate of a paper increases up to a few years after its publication, reaches a peak and then decreases rapidly. This decay can be described by an exponential or a power law behavior, as in ultradiffusive processes, with exponential fitting better than power law for the majority of cases. The decay is also becoming faster over the years, signaling that nowadays papers are forgotten more quickly. However, when time is counted in terms of the number of published papers, the rate of decay of citations is fairly independent of the period considered. This indicates that the attention of scholars depends on the number of published items, and not on real time.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Scientific publications in peer reviewed journals serve as the standard medium through which most of the progress of science is recorded. Besides offering a mechanism for claiming priorities and exposing results to be checked by others, publishing is also a way to attract attention of other scientists working on related problems. Attention, measured by the number and lifetime of citations, is the main currency of the scientific community, and along with other forms of recognition forms the basis for promotions and the reputation of scientists (Petersen et al., 2014). As Franck (Franck, 1999), Klamer and van Dalen (Klamer & Dalen, 2002) have pointed out, there is an attention economy at work in science, in which those seeking attention through the production of new knowledge are rewarded by being cited by their peers, whose own standing is measured by the amount of citations they receive.

The attention economy is also at work in many other fields besides science, ranging from entertainment to marketing, and is responsible for the phenomenon of stars, i.e., people whose income in attention far exceeds the norm in their own endeavors. Moreover, attention is a strong motivator of productivity. Recently, it has been shown that the productivity of YouTube videos exhibits a strong positive dependence on the attention they receive, measured by the number of downloads (Huberman, Romero, & Wu, 2009). Conversely, a lack of attention leads to a decrease in the number of videos uploaded and the consequent drop in productivity, which in many cases asymptotes to no uploads whatsoever.

---

* Corresponding author.
 E-mail address: rajkumar.pan@aalto.fi (R.K. Pan).

Table 1
Basic statistics of the different scientific fields we considered: Clinical Medicine, Molecular Biology, Chemistry and Physics. They represent the most active fields in terms of the total volume of publications. Here, N_P is the number of publications in a given field, c_{max} is the maximum number of citations to a given paper in that field and $\langle c \rangle$ is the average number of citations to all the papers in that field.

Field	N_P	c_{max}	$\langle c \rangle$
Clinical Medicine	10833626	25604	11
Molecular Biology	2849144	296498	24
Chemistry	4565197	134441	14
Physics	5583183	31759	13

Decision making and marketing, among others, are based on the mechanisms ruling how attention is stimulated and maintained (Dukas, 2004; Kahneman, 1973; Pashler, 1998; Pieters, Rosbergen, & Wedel, 1999; Reis, 2006). Over the past years, thanks to the Internet, a huge amount of data has allowed a thorough investigation of the dynamics of collective attention to online content, ranging from news stories (Dezső et al., 2006; Ghosh & Huberman, 2014; Wu & Huberman, 2007), to videos (Crane & Sornette, 2008) and memes (Leskovec, Backstrom, & Kleinberg, 2009; Matsubara, Sakurai, Prakash, Li, & Faloutsos, 2012; Weng, Flammini, Vespignani, & Menczer, 2012). Here attention is measured by the number of users' views, visits, posts, downloads, tweets. It is also noted that the attention decays over time, not only because novelty fades, but also because the human capacity to pay attention to new content is limited. A typical temporal pattern is characterized by an initial rapid growth, followed by a decay. The decay turns out to be slower than exponential: power law fits give the best results, stretched exponentials being preferable in particular cases (Wu & Huberman, 2007).

In this paper we focus on the decay of attention in science, on the basis of scientific articles, which like any other content, become obsolete after a while. Typically this happens because their results are surpassed by those of successive papers, which then "steal" attention from them. The problem of the obsolescence of scientific contents has received a lot of attention in scientometrics. The typical approach is to study the evolution of the number of citations received by a paper in a given time frame (usually one year), since its publication. The nature of the decay has been controversial, between claims of an exponential trend (Avramescu, 1979; Medo, Cimini, & Gualdi, 2011; Nakamoto, 1988) and analyses supporting a slower power law curve (Bouabid, 2011; Bouabid & Larivière, 2013; Pollman, 2000; Redner, 2005). This is partly due to the different types of analysis and the use of distinct data sources. Note that patterns of individual papers are usually noisy, as one cannot count on the high statistics available for online contents: the number of tweets posted on a single popular topic may exceed the total number of scientific publications ever made.

On the other hand, in contrast to online sources, bibliographic databases enable one to perform a longitudinal study of the life cycles of papers. In this work we make a systematic analysis of papers' life cycles, across different scientific fields and historical periods. We find that the decay of attention for individual papers can be described both by exponential and power law behaviors. Exponential fits turn out to be preferable in the majority of cases. These results are compatible with a relaxation of attention modeled by ultradiffusion, as observed for the popularity of online content (Ghosh & Huberman, 2014). We also found that attention is dying out more rapidly with time. However, due to the ongoing exponential growth of scientific publications, which is known to influence citation patterns (Egghe, 2000; Yang, Ma, Song, & Qiu, 2010), we conjecture that the faster decay observed nowadays is a consequence of the much larger pool of papers among which attention has to be distributed. In fact, if time is renormalized in terms of the number of papers published in the corresponding period (e.g., in each given year), we find that the rescaled curves die out at comparable rates across the decades.

2. Material and methods

2.1. Data description

Our data set consists of all publications (articles and reviews) written in English till the end of 2010 included in the database of the Thomson Reuters (TR) Web of Science. For each publication we extracted its year of publication, the subject category of the journal in which it is published and the corresponding citations to that publication. Based on the subject category of the journal (determined by TR) of the publication, the papers were categorized in broader disciplines such as Physics, Medicine, Chemistry and Biology (see Table 1). Most analyses are carried out using the top 10% papers (based on their total number of citations), as it allows to include a sufficient number of papers from older times, but still keeping the number of yearly citations large enough to allow for a statistically valid analysis. The analysis of papers with relatively lower citations follow qualitatively similar behavior and is shown in the Appendix.

2.2. Data fitting and F-statistics

We measure the trend in the temporal evolution of the different plots using the least square method. We consider the *F*-statistics for a significant linear regression relationship between the response variable and the predictor variable. We used it to compare the statistical models that best fit the population from which the data were sampled. As the *F*-score takes into account both the number of data points available for the fit and the number of degrees of freedom of the model, it is possible to compare the accuracy of the fit for different models with different parameters or between data sets of different size.