Journal of Informetrics 7 (2013) 603–610

Contents lists available at ScienceDirect

Journal of Informetrics

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

Two time series, their meaning and some applications

Ronald Rousseau[a,b,c], Xiaojun Hu[d,*]

[a] VIVES (Association KU Leuven), Faculty of Engineering Technology, Zeedijk 101, 8400 Oostende, Belgium
[b] Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, 2000 Antwerpen, Belgium
[c] KU Leuven, 3000 Leuven, Belgium
[d] Medical Information Centre, Zhejiang University School of Medicine, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 27 January 2013
Received in revised form 21 March 2013
Accepted 21 March 2013
Available online 23 April 2013

Keywords:
Time series
Social sciences
Informetrics
Chinese universities

ABSTRACT

Introducing and studying two types of time series, referred to as *R1* and *R2*, we try to enrich the set of time series available for time dependent informetric studies. In a first part we focus on mathematical properties, while in a second part we check if these properties are visible in real data. This practical application uses data in the social sciences related to top Chinese universities. *R1* sequences always increase over time, tending relatively fast to one, while *R2* sequences have a decreasing tendency tending to zero in practical cases. They can best be used over relatively short periods of time. *R1* sequences can be used to detect the rate with which cumulative data increase, while *R2* sequences detect the relative rate of development.

The article ends by pointing out that these time series can be used to compare innovative activities in firms. Clearly, this investigation is just a first attempt. More studies are needed, including comparisons with other related sequences.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction: time series in informetric investigations

Numerous colleagues have studied time series in the context of informetric or scientometric investigations (Abercrombie, Udoeyop, & Schlicher, 2012; Franses, 2003; Hood & Wilson, 2001; Ingwersen & Elleby, 2011; Mahbuba, Rousseau, & Srivastava, 2010; Vanclay, 2012; Ye & Rousseau, 2008). Most of these studies collect basic information, such as number of publications, number of citations, journal impact over a period of time, *h*-index values, and perform trend analyses, investigating temporal changes (increasing or decreasing trends, fluctuations) in the data. Most studies of the growth of science are of this type (Brookes, 1970; Price & de Solla, 1963). Some studies compare time series such as (Rousseau, 1999) in which the time series of a journal's self-citations is compared with its external citations. Note that when cumulative data are collected there is always "growth". Yet, the term 'growth' is sometimes used as a synonym for development or evolution, leading to the study of growth cycles (Swanson, 1993). Liang and Rousseau (2008) study time series of temporal data, namely the time it takes a journal volume to accumulate the same number of citations as the number of references included in that volume. This period is called the yield period and the authors study, as an example, the yield periods of the journals *Nature* and *Science*, over the period [1955–2003]. Franceschini and Maisano (2011) take an innovative approach by studying the stability of research output over time. Rousseau (2001) illustrates how to apply median filtering as a smoothing technique for irregular time series (hits in keyword searches on the Internet). Trend detection based on Internet data is performed e.g. in (Amitay, Carmel, Herscovici, Lempel, & Soffer, 2004; Payne & Thelwall, 2008).

* Corresponding author. Tel.: +86 057188208085.
E-mail addresses: ronald.rousseau@ua.ac.be (R. Rousseau), xjhu@zju.edu.cn (X. Hu).

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.03.006

Special cases are the works by McGrath (1996) and Decroos et al. (1997). These authors analyze series of library circulation data using spectral methods. By these methods periodicities, i.e. recurrent phenomena, hidden in the data, can be detected. Doré and Ojasoo (2001) demonstrate the use of correspondence factorial analysis to jointly study publication trends of many countries and disciplines. Generally time series in informetrics use one year as a basic step, but sometimes smaller or larger units are used too (Brunk, 2003). Liang and Rousseau (2009) even study the journal Science based on periods defined by editorship: the time before Daniel Koshland was editor-in-chief, the period of his editorship, and the period thereafter. Hu, Rousseau, and Chen (2011) present two synchronous and one diachronous time series related to the so-called outgrow index, an indicator related to the extent to which an article outgrows – in terms of received citations – the references on which it is based. They even propose a mixed time matrix, using two time dimensions instead of the common one dimension.

Continuous and discrete modeling of informetric time series is performed e.g. in (Bettencourt, Kaiser, & Kaur, 2009; Burrell, 1980; Chen et al., 2009; Franses, 2003; Glänzel, 2007; Hall, 1992). A classification of different discrete time series is given in Liu and Rousseau (2008). Animations of time slices as developed by Leydesdorff and Schank (2008) using Visone lead to dynamic visualization of journal network structures.

This section is not meant as a literature review of the study of time series in informetrics. It just provides a framework in which our work is situated. For more information on time series in informetrics we refer to (Bar-Ilan, 2008). For time series in the context of h-indices we refer to (Egghe, 2010). In this article the terms series and sequence are used as synonyms.

The remainder of the article is organized as follows: in the next section we formulate the purpose of the study; followed by the definitions of two time series and their meaning. Then we discuss practical applications of the two time series, ending with a conclusion and suggestions for further research, in particular in the context of comparing the competitive strength of companies.

2. Purpose of the study

Informetric studies are often described within a sources-items framework (Egghe, 2005). Typical examples are authors as sources producing articles as items; or articles "producing" citations. In this contribution we focus on one source and record the number of items produced per unit of time. Examples are: a scientist, a research group, a department, a university or a country as source and the number of articles published each year (as time unit) during a given period. This leads to a basic time series of numbers of articles published. Similarly one may consider the yearly number of patents applied for by a given company or granted to it. By applying restrictions on the data, one may obtain specific subsequences of interest such as a time series of numbers of patents applied for jointly by a university and a firm as a result of collaborative research (Van Looy, Callaert, & Debackere, 2006). In the context of knowledge diffusion one may collect data leading to a time series of different citing journals or ESI-fields of a given article or set of articles (Liu & Rousseau, 2010). Studying patents one may collect a time series of different fields of technology in which a company is actively patenting. Going outside our field one may collect time series of weekly (another time unit) book sales, daily sales of a convenience store or stock exchange transactions per hour. Associated with any time series of yearly events $(X_j)_{j=0,1,...}$, we associate a time series of cumulative events with elements

$(\mathrm{TX})_j$, $j = 0,1,\ldots$, where $(\mathrm{TX})_j = \sum_{k=0}^{j} X_k$.

Time series constructed by counting different, but related, items such as different citing journals, different citing ESI-fields, different technological fields in which a company becomes active are especially interesting as they are related to diffusion and innovation. Similarly one may collect for a research group the different journals or fields they themselves cite over time leading to studies of knowledge integration or interdisciplinarity (Liu, Rafols, & Rousseau, 2012; Rafols & Meyer, 2010).

Finally, we remark that modeling aspects, which are, in our field, usually studied in a continuous framework (Egghe, 2005) are outside the scope of this contribution.

3. A first derived series

Given a general time series $(X_j)_{j=0,1,...}$ (where each $X_j \geq 0$) we form two types of derived sequences. The first makes use of a base year Y (assuming that years are used as the unit of time). X_0 is the value in the year Y, for instance the number of publications by institute I or the number of citations received in the year Y by an article published in that same year. Then the sequence $(R1_j)_j$ or $(R1_j^Y)_j$ if we want to stress the base year Y, is defined as:

$$R1_j^Y = \frac{(\mathrm{TX})_j - (\mathrm{TX})_0}{(\mathrm{TX})_j} = \frac{(\mathrm{TX})_j - X_0}{(\mathrm{TX})_j} = 1 - \frac{X_0}{(\mathrm{TX})_j}$$

where X_0 is the value in the year Y. If X_0 is zero then this sequence is always equal to 1, and hence of no practical meaning. Henceforth we will always assume that $X_0 > 0$. The first element of this derived sequence $R1_0$ is always equal to zero. When there is no increase, $X_j = 0$ for $j > 0$, then the $R1$ sequence is always zero. Hence $0 \leq R1_j < 1$. If $X_{j+1} = 0$ then $R1_j^Y = R1_{j+1}^Y$. If $(X_j)_{j=0,1,...}$, is non-zero for every j and consisting of natural numbers then $(\mathrm{TX})_j$, $j = 0,1,\ldots$ is strictly increasing and the

Download English Version:

https://daneshyari.com/en/article/523168

Download Persian Version:

https://daneshyari.com/article/523168

Daneshyari.com