Journal of Informetrics 6 (2012) 177–191

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

A preliminary investigation on diffusion through a layered system

Ronald Rousseau[a,b,c], Yuxian Liu[d], Fred Y. Ye[e,*]

[a] KHBO (Association K.U. Leuven), Faculty of Engineering Technology, Zeedijk 101, B-8400 Oostende, Belgium
[b] K.U. Leuven, Department of Mathematics, Celestijnenlaan 200B, B-3001 Leuven (Heverlee), Belgium
[c] Antwerp University (UA), IOIW, Venusstraat 35, B 2000 Antwerp, Belgium
[d] Tongji University, Library of Tongji University, Siping Street 1239, Shanghai 200092, China
[e] Zhejiang University, Department of Information Resource Management, Hangzhou 310027 China

ARTICLE INFO

Article history:
Received 14 November 2011
Received in revised form 19 January 2012
Accepted 19 January 2012

Keywords:
Diffusion
Layered systems
Gini evenness index
Counting methods

ABSTRACT

We introduce layered systems such as the citations–citing authors–citing institutes–citing countries one. Diffusion of scientific ideas flows through such layered systems. Our contribution contains three main topics: a fractional counting system for the number of different units in a layer; the fractional number of items of the same type, i.e. in the same layer, over which ideas have been diffused; and the evenness of diffusion over different layers. In this way we construct a coherent system to measure the extent to which scientific ideas are diffused.

1. Introduction

Diffusion of scientific ideas, usually operationalized by the study of the diffusion of citations received by an article or a group of articles containing these ideas, is one of the main topics in informetrics (Le Coadic, 1987). Recently the topic has received renewed interest in the context of interdisciplinarity and knowledge integration (Liu, 2011; Liu, Rafols, & Rousseau, 2012; Rafols & Meyer, 2010). The fact that the spread or diffusion of scientific ideas is nowadays considered as a factor in determining the innovativeness of these ideas is one of the reasons explaining renewed interest in the phenomenon of diffusion (Peri, 2005).

What is diffusion? Strictly speaking the term diffusion as such has no meaning. In any concrete context its meaning must be specified as diffusion always refers to the spread of something from somewhere to somewhere else in or via some medium. In physics diffusion refers to a process whereby particles of liquids, gases, or solids intermingle as the result of their spontaneous movement caused by thermal agitation (Merriam-Webster, http://www.merriam-webster.com). In the social sciences and humanities diffusion refers to the spread of cultural elements from one area or group of people to others and this through some medium such as newspapers, the Internet and gossip. In the context of the information sciences diffusion can be described as a virtual movement through cognitive space (Liu, 2011). Scientific results are diffused from one field to other ones, from laboratory to article, from science to technology, and from technology to society. The medium can be face-to-face apprenticeship, scientific articles, talks during scientific conferences and so on.

Diffusion has been studied in the past using several techniques. A simple way, is the use of diffusion factors, introduced in the informetric literature by Rowlands (2002) and further developed in Frandsen (2004) and Frandsen, Rousseau, and Rowlands (2006). Rowlands (2002) considered the journal diffusion factor as a measure of the transdisciplinary reception

* Corresponding author. Tel.: +86 571 87996929; fax: +86 571 87953814.
E-mail addresses: ronald.rousseau@khbo.be (R. Rousseau), yxliu@lib.tongji.edu.cn (Y. Liu), yye@zju.edu.cn (F.Y. Ye).

1751-1577/$ – see front matter
doi:10.1016/j.joi.2012.01.005

of a journal. He used the metaphor of throwing pebbles in a pond to describe the publication of new scientific ideas. The diffusion factor may then be understood as a measurement of the characteristics of the resulting ripples (citations). We do not recall Rowlands' definition but immediately proceed to the general forms as introduced in Frandsen et al. (2006).

The n-year diachronous relative diffusion factor of journal J in the year Y is defined as:

$$RDI_n(Y) = \frac{\sum_{j=0}^{n-1} U(Y+j, Y)}{\sum_{j=0}^{n-1} CIT(Y+j, Y)} \tag{1}$$

In this equation, $CIT(Y+j, Y)$ denotes the number of citations received in the year $Y+j$ by articles published in the year Y (in some fixed database). The number of unique, new journals for citations in the year $Y+j$, of articles published in this journal in the fixed year Y is denoted as $U(Y+j, Y)$. The phrase 'unique, new' refers to the fact that this journal has not cited an article published in the journal J in the year Y during the years $Y, Y+1, \ldots, Y+j-1$, but that it did cite (in the year $Y+j$) an article published in the year Y. Note that in the sum of the numerator any journal can contribute at most once. Also Frandsen (2004) defined a journal impact factor. Hers, denoted as $DI_n(Y)$ is defined as:

$$DI_n(Y) = \frac{\sum_{j=0}^{n-1} U(Y+j, Y)}{PUB(Y)} \tag{2}$$

The symbol $PUB(Y)$ denotes the number of articles published in journal J in the year Y. In the Rowlands–Frandsen–Rousseau approach the main part is the number of different, citing journals. In general the source of diffusion is a set of articles (not necessarily those published in one journal during one year). Frandsen et al. even provide an example in which the source is just one article. They also give an example of four articles published in different journals. Diffusion itself is measured by counting journals, where whole counting is used as a journal contributes (value: 1) or not (value: 0). A more sophisticated approach to diffusion in the context of research technologies can be found in Leydesdorff and Rafols (2011).

2. This contribution: layered systems

In this contribution we consider a layered system such as the "citing articles–citing authors–citing universities–citing countries" system. We prefer using layered systems as they lead to a coherent study. We will apply our ideas not only to hierarchical systems such as the previous example, but also to systems consisting of different levels which are not necessarily hierarchical such as published articles–authors–journals in which these articles are published (the published article layer is split into two non-hierarchical layers: the authors and the journal layer). We note that the 'citing author' layer has been studied in Ajiferuke and Wolfram (2009) and, in the context of h-indices, in Franceschini, Maisano, Perotti, and Proto (2010). Clearly if an article receives citations originating from multiple countries then this article's content is diffused more than when the article has received the same total amount of citations, but originating from just one country. It is this difference we intend to measure. We introduce a fractional counting system extending the whole counting system used in earlier publications on the subject of diffusion. We also want to study the evenness of this citations–authors–universities–countries system as a whole. The Gini evenness index will be our tool to realize this objective.

3. Methodology

3.1. Counting different items in a layered system

In this section we propose a fractional counting system for the number of different items. While doing this we assume that each unit contributes proportionally. If knowledge of the exact contribution of an actor is available, and if considered useful, a more sophisticated form of fractionalization may be used. Let article A be the object of our investigation; the scientific ideas contained in it are the source of diffusion.

Consider a layered system such as:

Layer 1: the set of all citations (received by article A) over a given citation window.
Layer 2: the set of all citing authors.
Layer 3: the set of all universities or institutes to which the citing authors belong.
Layer 4: the set of all countries to which these universities belong.

We are interested in the distributional characteristics of these authors, universities and countries. The underlying idea is that the more citing authors (different ones) the better, and similarly: the more citing universities and countries (again different ones), the better.

If the source of diffusion is a single article (as we assume here) then each citation corresponds to exactly one citing article. In this case the number of citations is equal to the number of citing articles (but see further in the next section). The total number of received citations is denoted as C. If a citing article has ten authors we do not want to give these ten authors a full citation credit. One citing article should only have a contribution equivalent to one citing author, one citing university and

Table 1
Layered system of received citations.

Citing articles	Authors	Universities	Countries
Citing article (CA1)	Author 1 (AU1)	University 1 (UNI1)	Country 1 (CO1)
	Author 2 (AU2)	University 1 (UNI1)	Country 1 (CO1)
Citing article (CA2)	Author 2 (AU2)	University 1 (UNI1)	Country 1 (CO1)
Citing article (CA3)	Author 1 (AU1)	University 1 (UNI1)	Country 1 (CO1)
	Author 3 (AU3)	University 2 (UNI2)	Country 2 (CO2)
	Author 4 (AU4)	University 2 (UNI2)	Country 2 (CO2)
	Author 5 (AU5)	University 3 (UNI3)	Country 2 (CO2)
		University 4 (UNI4)	Country 3 (CO3)

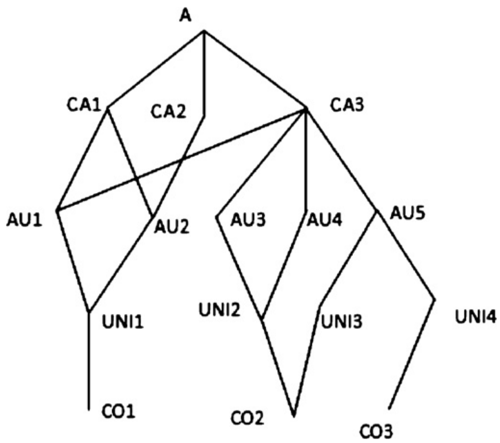

Fig. 1. Table 1 illustrated as a layered system.

one citing country. Recall moreover that we are not interested in the total number of citing units but in the total number of *different* citing units.

In order to realize this aim we propose the following counting scheme.

Assume that article A is cited three times as shown in Table 1.

Table 1 shows, among other things, that citing article 3 has 4 authors. The last of these (author 5) has two addresses; one at University 3 in Country 2, and one at University 4 in Country 3. This layered system is shown in Fig. 1. We note that, exceptionally, it is possible that one university has campuses in different countries or regions.

We want to assign counts to the four levels of this system. Clearly there are 3 received citations. The number of (weighted) different citing authors is obtained as shown in Table 2.

Using this counting method each article contributes a total of one author unit. If there is more than one author, contributions are divided equally (see for instance citing article 3, with four citing authors). This is the credit that one author obtains for his/her contribution to diffuse the scientific idea contained in article A. Of course, if for some reason one wants to divide credits among authors in another way (Hu, 2009), this is possible and does not change our approach. The total credit, denoted as CSum, is equal to max(1, sum of all contributions). In this way the maximum credit that a specific author is able to obtain is set to one. This truncation is performed because we study different diffusion units.

The sum of all credits that all authors obtain from the citing article is the credit of the author layer for the contribution to diffuse the scientific idea. If all articles were single-authored this would lead to the number of different citing authors in the normal sense. In real cases many articles are multi-authored, leading to a fractionally determined diffusion credit of the author layer. In this example this is 2.50.

Table 2
Counting method for the author layer.

Authors	Article 1	Article 2	Article 3	CSum
Author 1	0.5		0.25	0.75
Author 2	0.5	1		1
Author 3			0.25	0.25
Author 4			0.25	0.25
Author 5			0.25	0.25
Total author units	1	1	1	2.50

Table 3
Counting method for the university layer.

Universities	Article 1	Article 2	Article 3	CSum
University 1	1	1	0.25	1
University 2			0.5	0.5
University 3			0.125	0.125
University 4			0.125	0.125
Total university units	1	1	1	1.75

Table 4
Counting method for the country layer.

Countries	Article 1	Article 2	Article 3	CSum
Country 1	1	1	0.25	1
Country 2			0.625	0.625
Country 3			0.125	0.125
Total country units	1	1	1	1.75

Table 5
Layered system: special case 1.

Citing articles	Authors	Universities	Countries
Citing article 1	Author 1	University 1	Country 1
Citing article 2	Author 2	University 2	Country 2
Citing article 3	Author 3	University 3	Country 3

The credit of the different citing universities is derived from the author contributions and is obtained as illustrated in Table 3.

Again each citing article contributes a total of exactly one citing – university or institute – unit. If there are different authors, say n, each author's university receives a contribution of $1/n$. If, however, an author has k institutional addresses, then each institute contributes, via this author, a score of $1/nk$. Again CSum = max(1, sum of all contributions). Finally the credit score of citing countries is obtained as shown in Table 4. These sums too are derived from the author distribution (and not from the university distribution).

This leads to the following row of layered credits: 3–2.50–1.75–1.75.

Definition

A credit row as calculated above is called an LC-row, where LC stands for *layered credits*.

Often, and especially when there is a natural hierarchy in the used levels, such a row is non-increasing. Yet, an example of a row originating from a hierarchical layered system, for which this is not the case is given in Appendix.

3.2. Evenness and its relation with diffusion in layered systems

At any moment in time we want to know the relation between the numbers in different layers, i.e. the numbers constituting the LC-row. To gauge this difference we use the notion of evenness, concretely operationalized by one of the most-used evenness measures, namely the Gini evenness index (Nijssen, Rousseau, & Van Hecke, 1998). This index is calculated as follows: if $X=(x_j)$, $j=1, \ldots, N$ denotes an array of non-negative numbers, then the Gini evenness index, denoted as G_e, is calculated as:

$$G_e(X) = \frac{2\sum_{j=1}^{N} j x_j}{\mu N^2} - \frac{1}{N}$$

where the x_j's are ranked from high to low and μ denotes the mean of the set $\{x_j\}$. In our example $N=4$; $X=(3, 2.5, 1.75, 1.75)$; $\mu=2.25$. Hence the Gini evenness value is 0.875 which seems a high evenness value (but see further). Indeed, when all values are the same then $G_e=1$. In general the lowest Gini evenness value occurs for arrays of the form $(x, 0, 0, \ldots, 0)$ and is equal to $1/N$ if the length of this array is N. However, in our type of study zeros cannot occur and the least number of items at any level is 1. Hence, in our case we have to consider arrays of the form $(C, 1, 1, \ldots, 1)$, whose length depends on the number of levels considered. When the number of levels (N) is fixed the G_e-value of $(C, 1, 1, \ldots, 1)$ can be made as close to $1/N$ as one wishes by increasing C. If N and C are fixed then the lowest value is $(C+N^2-1)/((C+N-1)N)$. A relative Gini index, denoted as G_r, can then be defined which maps MIN = $(C+N^2-1)/((C+N-1)N)$ to zero and keeps 1 to 1. This relative Gini index is set equal to $(G_e - \text{MIN})/(1-\text{MIN})$. For the example of Table 1 MIN = (corresponding to $C=3$ and $N=4$) = 0.75 and hence $G_r=0.5$, which gives a better intuitive feeling of the evenness over layers than the value of 0.875 we obtained earlier.

Next we consider two extreme cases. In the first case (Table 5) each citing article is written by a different author, working in a different country; in the second (Table 6) all citing articles are written by the same author.

Table 6
Layered system: special case 2.

Citing articles	Authors	Universities	Countries
Citing article 1	Author 1	University 1	Country 1
Citing article 2	Author 1	University 1	Country 1
Citing article 3	Author 1	University 1	Country 1
Citing article 4	Author 1	University 1	Country 1
Citing article 5	Author 1	University 1	Country 1
Citing article 6	Author 1	University 1	Country 1

Table 7
Case study (data collected on August 19, 2011).

Citing articles	Authors	Universities	Countries
Citing article 1	Alonso	Univ. Granada	Spain
	Cabrerizo	Univ. Granada	Spain
	Herrera-Viedma	Univ. Granada	Spain
	Herrera	Univ. Granada	Spain
Citing article 2	Cabrerizo	Univ. Granada	Spain
	Alonso	Univ. Granada	Spain
	Herrera-Viedma	Univ. Granada	Spain
	Herrera	Univ. Granada	Spain
Citing article 3	Ye	Zhejiang Univ.	China
		ISTIC, Beijing	China
	Rousseau	KHBO	Belgium
		K.U. Leuven	Belgium
Citing article 4	Ye	Zhejiang Univ.	China
		iFQ, Bonn	Germany
Citing article 5	Lin	Guangxi Normal Univ.	China

Table 8
Counting different authors.

Authors	Art. 1	Art. 2	Art. 3	Art. 4	Art. 5	CSum
Alonso	0.25	0.25				0.5
Cabrerizo	0.25	0.25				0.5
Herrera-Viedma	0.25	0.25				0.5
Herrera	0.25	0.25				0.5
Ye			0.5	1		1
Rousseau			0.5			0.5
Lin					1	1
Total number of different authors	1	1	1	1	1	4.5

The citing level array of Table 5 is: (3, 3, 3, 3) and hence its evenness-value is 1. By definition this is also its relative evenness value.

The LC row of Table 6 is (6, 1, 1, 1). Its Gini evenness index is 0.5833 which is equal to $\text{MIN} = (C + N^2 - 1)/((C + N - 1)N)$ with $C = 6$ and $N = 4$. The relative Gini index is zero.

4. First case study

As an example of a real case we consider one of our own articles:

Ye, F. Y., & Rousseau, R. (2008). The power law model and total career h-indices sequences. *Journal of Informetrics, 2*(4), 288–297.

This article has been cited 5 times as shown in Table 7.

There are 5 citing articles. The citing authors are determined as shown in Table 8.

Citing institutes are determined as shown in Table 9.

Finally we come to the countries, see Table 10.

Table 9
Counting different institutes and universities.

Institutes	Art. 1	Art. 2	Art. 3	Art. 4	Art.5	CSum
Univ. Granada	1	1				1
Zhejiang Univ.			0.25	0.5		0.75
ISTIC, Beijing			0.25			0.25
KHBO			0.25			0.25
K.U. Leuven			0.25			0.25
iFQ, Bonn				0.5		0.5
Guangxi Normal Univ.					1	1
Total number of different institutes	1	1	1	1	1	4.00

Table 10
Counting different countries.

Countries	Art. 1	Art. 2	Art. 3	Art. 4	Art. 5	CSum
Spain	1	1				1
China			0.5	0.5	1	1
Belgium			0.5			0.5
Germany				0.5		0.5
Total number of different countries	1	1	1	1	1	3.0

All these calculations brought together yield the array (5, 4.5, 4, 3). Its G_e-value is 0.9015 and $G_r = 0.7374$ (using MIN = 0.625).

5. Other methodological considerations

5.1. Self-citations

How should self-citations be counted when studying diffusion? When knowledge is diffused it exists at a certain location (in the general sense): in a country, in a research area, in a journal, within a person's mind. The existence of knowledge is what we want to take into account. We mention four options (among many other ones). The first one consists in giving each co-author of the original article a full credit at the start. The rationale behind this is that the knowledge exists in the authors' head, and also in their institute, country or whichever unit they belong to. Alternatively one may credit each co-author with a diffusion value of $1/n$ (if the article is written by n co-authors), but then it must be accepted that their contribution to the total diffusion increases through self-citations. The method we have applied consists in considering a self-citation in exactly the same way as any other citation. The rationale behind this approach is that citing is an activity with respect to the knowledge included in the original article, while publishing is passive. Finally, one may remove all self-citations, but this approach does not seem fair as the knowledge in the original article does exist in the authors' institute, country or field. We consider our approach as a reasonable compromise, but have no objection if, in a diffusion context, colleagues prefer to take self-citations into account in a different way. We further note that when the source consists of several articles (see further), then other approaches become even more complicated.

5.2. Other potential layered systems

Besides the citations–citing authors–citing universities or institutes–citing country layered system (origination from one source article), other layered systems may similarly be studied. Examples are:

- Reference items (also originating from one source article), e.g. restricted to journal articles–journals in which these references are published–the publishers of these journals.
- One source article: citing articles (restricted to articles included in the WoS)–citing journals–JCR subject areas to which the journal belongs–ESI fields.

The first-mentioned layered system can be considered a kind of dual of the one we started from as references are used instead of received citations.

These examples make it clear that LC-rows depend on the particular layered system under consideration.

5.3. The case that the source consists of more than one article

Instead of a source consisting of a single article one may similarly study the diffusion of a set of articles as source. Examples are: The set of articles

- written by one scientist (over a given publication period), or
- published on a certain topic, or
- published by a research group, or
- published by a country.

In a scientific investigation this source may even be a random set of articles sampled from a larger set of interest.

When the source consists of several articles one naturally has another level, namely that of citing articles as with each citation a citing article is associated, but different citations may originate from the same citing article. The idea of studying citing articles instead of citations in research evaluation exercises has been applied by Rons and Amez (2009) in their impact vitality measure. From that point on the framework built in the previous section can be applied without further conceptual changes.

6. Second case study

As an example we consider all articles published jointly by Rousseau and Ye during the period [2008–2010] and indexed by the WoS (data collected on August 19, 2011). These articles are the following three:

A1. Ye, F. Y., & Rousseau, R. (2008). The power law model and total career h-indices sequences. *Journal of Informetrics, 2*(4), 288–297.
A2. Rousseau, R., & Ye, F. Y. (2008). A proposal for a dynamic *h*-type index. *Journal of the American Society for Information Science and Technology, 59*(11), 1853–1855.
A3. Ye, F. Y., & Rousseau, R. (2010). Probing the *h*-core: An investigation of the tail–core ratio for rank distributions. *Scientometrics, 84*(2), 431–439.

Citing articles of a least one of these three source articles are shown in Table 11.

There are 19 citations, coming from 16 different citing articles. The citing authors are determined as shown in Table 20; the citing institutes are determined as shown in Table 21, and finally the weighted number of different countries/regions is obtained from Table 22. Tables 20–22 are shown in the appendix.

This yields the LC-row (19, 16, 14.5, 14, 8.5). Its G_e-value is 0.8722 and G_r = 0.7959 (MIN = 0.3739). These values point to a high evenness, meaning a rather high relative diffusion over all levels.

7. Which type of questions did we pose and which answers did we get so far?

At this point we reflect a moment to see where we are and what can be done further. Focusing on a set of source articles (including the singleton set) we asked the following questions: how are citations to this set of source articles diffused over different layers? Concretely we studied how many different articles cited at least one article in the source set? How many different authors cited at least one article in the source set? How many different universities or research institutes cited at least one of the source articles? How many different countries cited at least one of the source articles? In these questions the word ‘different’ is the keyword as we focus on the diffusion of the ideas contained in the source articles over these units. Moreover, we used a fractional approach to the notion of ‘different items’. The obtained result is a row of numbers representing diffusion over different layers (here: articles, authors, universities, countries). One expects that these numbers decrease (although there might be exceptions) and the faster they decrease the less diffusion over the layers under study. We determined the evenness in a LC-row by the corresponding Gini index. In the next section we consider the source set itself.

8. A source set consisting of several items may contain several layers

When the source consists of different articles one may also study layers within this set. Examples are:

(1) the source set of articles–the journals in which these articles are published–the fields to which these journals belong.
(2) the source set of articles dealing with a certain topic–its authors–the institutes or universities they belong to–the corresponding countries.

If the set of articles is the set of articles published by a – large or small – research unit we are studying a form of diffusion by publication. Hence our approach is related to the ideas proposed in Liu and Rousseau (2010) and Liu et al. (2012) where it is explained that diffusion of scientific ideas may occur through citations (and hence mainly by others) or through publication activities (and hence mainly by the persons involved in the publication of the set of articles under study). Partially renaming our first example yields an illustration of this situation (see Table 12).

The corresponding source row *Y* consists of the same numbers as row *X* in our first example: *Y* = (3, 2.50, 1.75, 1.75), and has, of course, the same Gini values: G_e = 0.875 and G_r = 0.5.

Table 11
Measure for layered system on multi-source articles.

Cited article	Citing articles	Authors	Universities	Countries
A1. Power law	Citing article 1	Alonso	Univ. Granada	Spain
		Cabrerizo	Univ. Granada	Spain
		Herrera-Viedma	Univ. Granada	Spain
		Herrera	Univ. Granada	Spain
	Citing article 2	Cabrerizo	Univ. Granada	Spain
		Alonso	Univ. Granada	Spain
		Herrera-Viedma	Univ. Granada	Spain
		Herrera	Univ. Granada	Spain
	Citing article 3	Ye	Zhejiang Univ.	China
			ISTIC, Beijing	China
		Rousseau	KHBO	Belgium
			K.U. Leuven	Belgium
	Citing article 4	Ye	Zhejiang Univ.	China
			iFQ, Bonn	Germany
	Citing article 5	Lin	Guangxi Normal Univ.	China
A2. Dynamic *h*-index	Citing article 6	Dorta-Gonzalez Pablo	Univ. Las Palmas	Spain
		Dorta-Gonzalez Maria-Isabel	Univ. La Laguna	Spain
	Citing article 4	Ye	Zhejiang Univ.	China
			iFQ, Bonn	Germany
	Citing article 7	Hirsch	Univ. Cal. San Diego	USA
	Citing article 8	Perc	Univ. Maribor	Slovenia
	Citing article 9	Vieira	Univ. Porto	Portugal
		Gomes JANF	Univ. Porto	Portugal
	Citing article 10	Perc	Univ. Maribor	Slovenia
	Citing article 11	Dorta-Gonzalez Pablo	Univ. Las Palmas	Spain
		Dorta-Gonzalez Maria-Isabel	Univ. La Laguna	Spain
	Citing article 1	Alonso	Univ. Granada	Spain
		Cabrerizo	Univ. Granada	Spain
		Herrera-Viedma	Univ. Granada	Spain
		Herrera	Univ. Granada	Spain
	Citing article 12	Zhang	Tianjin Univ.	China
	Citing article 13	Burrell	Isle Man Int. Business School	Isle of Man
	Citing article 14	Jacso	Univ. Hawai	USA
	Citing article 15	Rousseau	K.U. Leuven	Belgium
		Jin	Chinese Acad. Science	China
A3. Probing *h*-core	Citing article 5	Lin	Guangxi Normal Univ.	China
	Citing article 16	Kuan, C. H.	Natl. Taiwan Univ.	Taiwan
		Huang, M. H.	Natl. Taiwan Univ.	Taiwan
		Chen, D. Z.	Natl. Taiwan Univ.	Taiwan

Table 12
Layered source system: articles studying a given topic.

Articles	Authors	Universities	Countries
Article 1	Author 1	University 1	Country 1
	Author 2	University 1	Country 1
Article 2	Author 2	University 1	Country 1
Article 3	Author 1	University 1	Country 1
	Author 3	University 2	Country 2
	Author 4	University 2	Country 2
	Author 5	University 3	Country 2
		University 4	Country 3

The questions we ask here are: how many different articles are there in the source set? Written by how many different authors? From how many different universities in how many different countries? Again we use a fractional approach.

If the source set consists of the articles written by the members of a department over a given publication window then one may ask: how many different articles are there? Published in how many different journals? Classified in how many different JCR-categories?

9. Third case study

As a real case we consider articles discussing “diffusion factors” as a topic in the area of Information and Library Sciences (see Tables 13–17).

Hence the LC row is (16, 12.17, 11.37, 8.42). Its G_e-value is 0.8772 and G_r = 0.7928 (MIN = 0.4079).

Table 13
Articles on "diffusion factors".

A1.	Rowlands, I. (2002). Journal diffusion factor: A new approach to measuring research influence. *Aslib Proceedings, 54*(2), 77–84
A2.	Frandsen, T. F. (2004). Journal diffusion factors: A measure of diffusion? *Aslib Proceedings, 56*(1), 5–11
A3.	Egghe, L. (2005). Journal diffusion factors and their mathematical relations with the number of citations and with the impact factor. In *ISSI 2005: Proceedings of the 10th international conference of the international society for scientometrics and informetrics*, 109–120
A4.	Rousseau, R. (2005). Robert Fairthorne and the empirical power laws. *Journal of Documentation, 61*, 194–202
A5.	Frandsen, T. F. (2005). Geographical concentration – The case of economics journals. *Scientometrics, 63*, 69–85
A6.	Lewison, G., Rippon, I., & Wooding, S. (2005). Tracking knowledge diffusion through citations. *Research Evaluation, 14*, 5–14
A7.	Frandsen, T. F., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation, 62*(1), 58–72
A8.	Sapa, R. (2007). International contribution to library and information science. *Scientometrics, 71*, 473–493
A9.	Haddow, G. (2008). Quality Australian journals in the humanities and social sciences. *Australian Academic and Research Libraries, 39*, 79–91
A10.	Rummler, G. (2009). Characterizing the dissemination of "Bibliometrics" in Brazilian Biomedical Journals from 1992 to 2007. In *ISSI 2009: Proceedings of the 10th international conference of the international society for scientometrics and informetrics*, 988–989
A11.	Jacso, P. (2009). Five-year impact factor data in the Journal Citation Reports. *Online Information Review, 33*, 603–614
A12.	Haddow, G., & Genoni, P. (2009). Australian education journals: Quantitative and qualitative indicators. *Australian Academic and Research Libraries, 40*, 88–104
A13.	Liu, Yx., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology, 61*(2), 340–351
A14.	Haddow, G., & Genoni, P. (2010) Citation analysis and peer ranking of Australian social. *Scientometrics*, *85*, 471–487
A15.	Sanni, S. A., & Zainab, A.N. (2011). Measuring the influence of a journal using impact and diffusion factors. *Malaysian Journal of Library & Information Science 16,* 127–140
A16.	Liu, Yx., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44

Table 14
The layered system of the diffusion articles.

A1	Rowlands, I.	City Univ. London	England	2002
A2	Frandsen, T. F.	Royal Sch. Lib. & Informat. Sci.	Denmark	2004
A3	Egghe, L.	Limburgs Univ. Ctr.	Belgium	2005
A4	Rousseau, R.	KHBO	Belgium	2005
		Univ. Antwerp	Belgium	
A5	Frandsen, T. F.	Royal Sch. Lib. & Informat. Sci.	Denmark	2005
A6	Lewison, G.	City Univ. London,	England	2005
	Rippon, I.	City Univ. London	England	
	Wooding, S.	RAND Europe	England	
A7	Frandsen, T. F.	Royal Sch. Lib. & Informat. Sci.	Denmark	2006
	Rousseau, R.	KHBO	Belgium	
	Rowlands, I.	City Univ. London	England	
A8	Sapa, R.	Jagiellonian Univ.	Poland	2007
A9	Haddow, G.	Curtin Univ. Lib.	Australia	2008
A10	Rummler, G.	Univ. Estadual Feira de Santana	Brazil	2009
A11	Jacso, P.	Univ. Hawaii	USA	2009
A12	Haddow, G.	Curtin Univ. Technol.	Australia	2009
	Genoni, P.	Curtin Univ. Technol.	Australia	
A13	Liu, Yx.	Tongji Univ.	China	2010
		Univ. Antwerp	Belgium	
	Rousseau, R.	KHBO	Belgium	
		K.U. Leuven	Belgium	
		Univ. Antwerp	Belgium	
A14	Haddow, G.	Curtin Univ. Technol.	Australia	2010
	Genoni, P.	Curtin Univ. Technol.	Australia	
A15	Sanni, S. A.	Lekan Salami Complex, Ibadan	Nigeria	2011
	Zainab, A. N.	Univ. Malaya	Malaysia	
A16	Liu, Yx.	Tongji Univ.	China	2012
		Univ. Antwerp	Belgium	
	Rafols, I.	University of Sussex	England	
		Georgia Institute of Technology	USA	
	Rousseau, R.	KHBO	Belgium	
		K.U. Leuven	Belgium	
		Univ. Antwerp	Belgium	

Table 15
Calculation of the credits per author.

16 authors	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	CSum
Egghe, L.			1														1
Frandsen, T. F.		1			1		0.34										1
Genoni, P.												0.5		0.5			1
Haddow, G.									1			0.5		0.5			1
Jacso, P.											1						1
Lewison, G.						0.34											0.34
Liu, Yx.													0.5			0.34	0.84
Rippon, I.						0.33											0.33
Rafols, I.																0.33	0.33
Rousseau, R.				1			0.33						0.5			0.33	1
Rowlands, I.	1						0.33										1
Rummler, G.										1							1
Sanni, S. A.															0.5		0.5
Sapa, R.								1									1
Wooding, S.						0.33											0.33
Zainab, A. N.															0.5		0.5
Total author units	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	12.17

Table 16
Calculation of the credits per institute.

16 institutes	A1	A2	A3	A4	A5	A 6	A 7	A8	A9	A 10	A 11	A 12	A 13	A 14	A 15	A 16	CSum
City Univ. London	1					0.67	0.33										1
Curtin Univ. Technol.									1			1		1			1
Georgia Institute of Technology																0.17	0.17
Jagiellonian Univ.								1									1
K.U. Leuven													0.17			0.11	0.28
KHBO				0.5			0.33						0.17			0.11	1
Lekan Salami Complex															0.5		0.5
Limburgs Univ.			1														1
RAND Europe						0.33											0.33
Royal Sch. Lib. & Informat. Sci.		1			1		0.34										1
Tongji Univ.													0.25			0.17	0.42
Univ. Antwerp				0.5									0.41			0.27	1
Univ. Estadual										1							1
Univ. Hawaii											1						1
Univ. Malaya															0.5		0.5
University of Sussex																0.17	0.17
Total university units	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	11.37

Table 17
Calculation of the credits per countries.

10 countries	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	CSum
Australia									1			1		1			1
Belgium			1	1			0.33						0.75			0.49	1
Brazil										1							1
China													0.25			0.17	0.42
Denmark		1			1		0.34										1
England	1					1	0.33									0.17	1
Malaysia															0.5		0.5
Nigeria															0.5		0.5
Poland								1									1
USA											1					0.17	1
Total country units	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.42

10. Discussion

We considered layered systems starting from a scientific article (or a group of articles) and its authorship and established counting methods for their study. In this approach we assume that scientific ideas are diffused via articles. In reality ideas are also diffused through formal and informal talks, e-mails, blogs and other electronic media but we assume that, eventually they all end up in scientific articles. The authors of these articles are from different institutes belonging to different countries. In this way a layered system and indicators associated with it can show how a scientific idea is diffused geographically (i.e. over the institutes or over the countries).

Table 18
Gini indices as a function of time.

	G_e	G_{min}	G_r
2002	1	1	1
2003	1	1	1
2004	1	0.85	1
2005	0.8681	0.5833	0.6835
2006	0.8431	0.55	0.6513
2007	0.8696	0.5227	0.7268
2008	0.8885	0.5	0.7769
2009	0.8934	0.45	0.8062
2010	0.8727	0.4265	0.778
2011	0.8839	0.4167	0.801
2012	0.8772	0.4079	0.7928

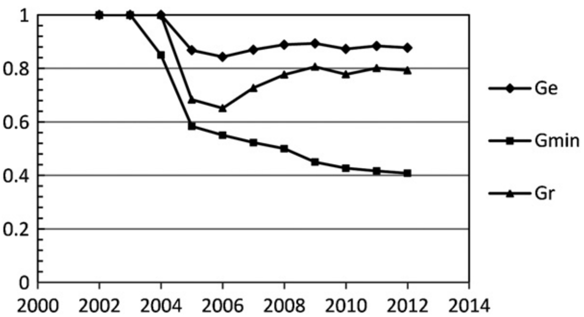

Fig. 2. Temporal Gini indices.

In order to obtain information on trends and evolutions we calculated the Gini evenness index and its relative Gini index for each year. This is done for the third case study (about diffusion factors). If $G_r = 0/0$ it is set equal to one. See Table 18 and Fig. 2.

Fig. 2 shows an interesting evolution over time. G_r starts at a value of 1 as expected. However, in the end, as there are much more authors than institutes, and much more institutes than countries, one expects that a successful article (or group of articles) will lead to a very concentrated (small evenness) situation. Relative recent or unsuccessful articles (like the ones studied) may converge to an intermediate value.

Table 19
Counterexample that an LC-row is not always decreasing.

Citing article	Citing authors	Citing institute
Article 1	Author 1	University 1
		University 2
		University 3
		University 4
	Author 2	University 5
		University 6
		University 7
		University 8
Article 2	Author 1	University 1
		University 2
		University 3
		University 4
	Author 2	University 5
		University 6
		University 7
		University 8
Article 3	Author 1	University 1
		University 2
		University 3
		University 4
	Author 2	University 5
		University 6
		University 7
		University 8

Table 20
Author table.

Authors	Art. 1	Art. 2	Art. 3	Art. 4	Art. 5	Art. 6	Art. 7	Art. 8	Art. 9	Art. 10	Art. 11	Art. 12	Art. 13	Art. 14	Art. 15	Art. 16	Sum
Alonso	0.25	0.25															0.5
Cabrerizo	0.25	0.25															0.5
Herrera-Viedma	0.25	0.25															0.5
Herrera	0.25	0.25															0.5
Ye			0.5	1													1
Rousseau			0.5												0.5		0.5
Lin					1												1
Dorta-Gonzalez Pablo						0.5					0.5						0.5
Dorta-Gonzalez Maria-Isabel						0.5					0.5						0.5
Hirsch							1										1
Perc								1		1							1
Vieira									0.5								0.5
Gomes									0.5								0.5
Zhang												1					1
Burrell													1				1
Jacso														1			1
Jin															0.5		0.5
Kuan, C. H.																0.333	0.333
Huang, M. H.																0.333	0.333
Chen, D. Z.																0.333	0.333
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14.5

Table 21
Tables of institutes.

Univ./institute	Art. 1	Art. 2	Art. 3	Art. 4	Art. 5	Art. 6	Art. 7	Art. 8	Art. 9	Art. 10	Art. 11	Art. 12	Art. 13	Art. 14	Art. 15	Art. 16	Sum
Univ. Granada	1	1															1
Zhejiang Univ.			0.25	0.5													0.75
ISTIC, Beijing			0.25														0.25
KHBO			0.25														0.25
K.U. Leuven			0.25												0.5		0.75
iFQ, Bonn				0.5													0.5
Guangxi Normal Univ.					1												1
Univ. Las Palmas						0.5					0.5						1
Univ. La Laguna						0.5					0.5						1
Univ. Cal. San Diego							1										1
Univ. Maribor								1		1							1
Univ. Porto									1								1
Tianjin Univ.													1				1
Isle Man Int. Business School														1			1
Univ. Hawai															0.5		1
Chinese Acad. Science																1	0.5
Natl. Taiwan Univ.																	1
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14

Table 22
Table of countries.

Countries/regions	Art. 1	Art. 2	Art. 3	Art. 4	Art. 5	Art. 6	Art. 7	Art. 8	Art. 9	Art. 10	Art. 11	Art. 12	Art. 13	Art. 14	Art. 15	Art. 16	Sum
Spain	1	1				1					1						1
China			0.5	0.5	1							1			0.5		1
Belgium			0.5												0.5		1
Germany				0.5													0.5
USA							1							1			1
Portugal									1								1
Slovenia								1		1							1
Isle of Man													1				1
Taiwan																1	1
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.5

11. Conclusion

Starting from publications or citations, we introduced layered systems and studied the diffusion of scientific ideas flows through such layered systems. Our contribution contains three main ideas:

- a fractional counting system for the number of different items,
- the fractional number of items of the same type, i.e. in the same layer, over which ideas have been diffused, is used as a basic diffusion indicator,
- inequality in the diffusion over the different layers of this system is characterized by an evenness value.

In this way we have constructed a coherent system to measure the extent to which scientific ideas are diffused.

Diffusion over layered systems begins with an evenness value of 1. We conjecture that depending on the source set, the type of layers and the time period, this evenness value may converge to a (small) steady state value, as articles may receive many more citations, than there are citing authors. These in turn may be in much larger numbers than there are universities, which in turn are more numerous than the number of countries. This is especially true for topics considered over an 'infinite' publication (or citation) window. For this reason one may say that the evenness indicator studied here is an indicator for transient behavior.

We finally like to point out that successful diffusion in a fixed group will lead to a high evenness value, as the notion under study will have reached each member of the group. However, when studying layered systems as we did, successful diffusion leads to a low evenness value.

Acknowledgments

The authors acknowledge the National Natural Science Foundation of China (NSFC Grant Nos. 7101017006, 71173187 and 71173154) for financial support and thank Star X. Zhao and Alice M. Tan for preliminary stimulating conversations. Fred Y. Ye thanks KHBO, Faculty of Engineering Technology for its hospitality during his visit. The authors thank anonymous reviewers and Raf Guns (UA) for constructive criticism.

Appendix.

Counterexample that an LC-row derived from a hierarchical system is not always decreasing.

Consider source article A and its citing articles, citing authors and citing institutes as shown in Table 19. Citing articles: 3; citing authors: 2; citing universities: $[(1/8) \times 3] \times 8 = 3$, leading to the LC-row: (3, 2, 3).

References

Ajiferuke, I., & Wolfram, D. (2009). Citer analysis as a measure of research impact: Library and information science as a case study. In B. Larsen, & J. Leta (Eds.), *Proceedings of the 12th international conference of the international society for scientometrics and informetrics (ISSI)* Rio de Janeiro, (pp. 798–808).
Franceschini, F., Maisano, D., Perotti, A., & Proto, A. (2010). Analysis of the *ch*-index: An indicator to evaluate the diffusion of scientific research output by citers. *Scientometrics*, *85*(1), 203–217.
Frandsen, T. F. (2004). Journal diffusion factors: A measure of diffusion? *Aslib Proceedings*, *56*(1), 5–11.
Frandsen, T. F., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation*, *62*(1), 58–72.
Hu, X. J. (2009). Loads of special authorship functions: Linear growth in the percentage of 'equal first authors' and corresponding authors. *Journal of the American Society for Information Science and Technology*, *60*, 2378–2381.
Le Coadic, Y. F. (1987). Modelling the communication, distribution, transmission or transfer of scientific communication. *Journal of Information Science*, *13*, 143–148.
Leydesdorff, L., & Rafols, I. (2011). Local emergence and global diffusion of research technologies: An exploration of patterns of network formation. *Journal of the American Society for Information Science and Technology*, *62*(5), 846–860.
Liu, Yx. (2011). *The diffusion of scientific ideas in time and indicators for the description of this process* (Doctoral thesis). Antwerp University.
Liu, Yx., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44.
Liu, Yx., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology*, *61*(2), 340–351.
Merriam-Webster. (n.d.). http://www.merriam-webster.com (last visited 15.01.12).
Nijssen, D., Rousseau, R., & Van Hecke, P. (1998). The Lorenz curve: A graphical representation of evenness. *Coenoses*, *13*(1), 33–38.
Peri, G. (2005). Determinants of knowledge flows and their effect on innovation. *The Review of Economics and Statistics*, *87*(2), 308–322.
Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287.
Rons, N., & Amez, L. (2009). Impact vitality: An indicator based on citing publications in search of excellent scientists. *Research Evaluation*, *18*(3), 233–241.
Rowlands, I. (2002). Journal diffusion factors: A new approach to measuring research influence. *Aslib Proceedings*, *54*(2), 77–84.