A tapered diffusion impact indicator: A preliminary exploration on the journal level

Lin Zhang[1,2]

[1] North China University of Water Conservancy and Electric Power,
Dept Management and Economics, Zhengzhou,
Beihuan Road 36, 450011 Zhengzhou, CHINA
[2] KU Leuven, ECOOM and MSI, Waaistraat 6, 3000 Leuven, BELGIUM
e-mail: linzhang1117@gmail.com

ABSTRACT

We introduce an indicator (D_T) to measure the diffusion impact of scientific research. The indicator focuses on the variety of citing sources by giving a full credit when a new citing source makes an appearance, and a tapered credit when a new citing paper appears in an already used existed source. As a preliminary study, we only used a simple example with individual journals as the diffusion units. A sample from a highly interdisciplinary journal, Chemometrics and Intelligent Laboratory Systems*, is used to illustrate this procedure. From the theoretical analysis and experimental examples, the D_T index proved an efficient function for differentiating the scope of citation diffusion.*

Keywords: Knowledge diffusion; Diffusion index; Tapered indicator; Citation analysis.

INTRODUCTION

The growth of scientific knowledge is largely due to a diffusion process in which new ideas are transmitted through some specific channels. The diffusion process of scientific knowledge can be considered as a layered system, namely, the citing authors - citing institutes - citing countries (Rousseau, Liu and Ye 2012), or the citing journals - citing sub-disciplines - citing disciplines. If the interaction between science and technology is taken into account then the analysis of knowledge diffusion can be expanded to include technology fields. The connection between basic research and technological innovations could then be traced by the scientific references cited in the patents, or vice versa.

Recently, promising approaches to knowledge diffusion analysis have been published. Chen and Hicks (2004) described an integrative approach to trace knowledge diffusion based on progressive and explanatory visualization of the evolution of patent citation networks. Bettencourt et al. (2006) applied epidemiological modes to capture the diffusion of research topics through scientific publications, and found a good fit between suitably adapted epidemic models and data for the spread of a specific research topic. Inspired by previous work on epidemic models, Kiss et al. (2010) demonstrated the feasibility of applying individual-based and weighted-network epidemic models to the spread of a research topic over the map of science. Bacchiocchi and Montobbio (2009) described the process of diffusion from university, public laboratories and corporate patents in six countries using data from the European Patent Office. The authors found that knowledge embedded in university and public research patents tends to diffuse more rapidly than knowledge originating in companies particularly in the USA, Germany, France and Japan. Yu, Wang and Yu (2010) investigated the knowledge diffusion patterns between nanoscience & nanotechnology and related subjects. Liu and Rousseau (2010) studied two forms of diffusion, namely, diffusion by publications which originates from the fact that a group publishes in different fields; and diffusion by citations, which originates from the fact that the group's publications are cited in different fields. Their approach was illustrated by a case study of mathematics at a Chinese university. Zhang, Thijs and Glänzel (2011) traced the diffusion of the h-related literature over a five-year period beginning with the introduction of the h-index, to explore how a new idea was spreading over different subject fields, authors and countries.

Most of the previous studies have been focusing on the knowledge diffusion process and network analysis, yet, very few investigations aimed on conducting an integrated analysis of the citation and diffusion impact.

TAPERED DIFFUSION INDEX

The traditional measures of citations are normally concerned with the (direct or relative) numbers of citations, such as the journal citation impact factor (Garfield and Sher 1963), or the more recently introduced h-index (Hirsch 2005). However, to date, very few indicators are able to differentiate the citing units (different citers, citing journals, citing subjects etc.) when calculating citation impact. One may ask whether two authors should be evaluated equally if they have received the same amount of citations, when the citations received by the first author are only from one citing unit (citers, journals, etc.), while the citations received by the other author are broadly distributed over different citing units.

Some recent studies tend to investigate on the citing side when evaluating citation impact. Ajiferuke and Wolfram (2010) proposed a ch-index to focus on the scope of "citing authors". These authors used the number of citers that an author's research has attracted. Later, Franceschini et al. (2010) further analyzed the characteristics of the "ch-index" and its relationship with the h-index.

In this study, we propose a diffusion impact indicator by which the diversity of citing units will be factored into the credits. The first step involves the classification of all citations received by one paper according to different citing units. As a preliminary study, we use individual journals as the citing units. A full credit will be assigned to a citation appearing in a new journal, and a tapered credit will be given to a citation appearing in a journal that has already cited the article under study. The idea of a tapered index was earlier proposed by Anderson, Hankin, and Killworth (2008), namely in the context of a "tapered h-index (h_T)". We agreed with the opinion of the authors that a bibliometric measure of publication output should be "strictly monotonic", that is assigning a positive score to each new citation as it occurs. However, different from Anderson, Hankin, and Killworth's (2008) research which focused on adjusting the h-index for evaluating individual authors, we are interested in differentiating the citing units (like journals) when measuring the diffusion impact of individual papers.

A simple demonstration of how to calculate the tapered diffusion index, denoted as D_T, is given in Figure 1.

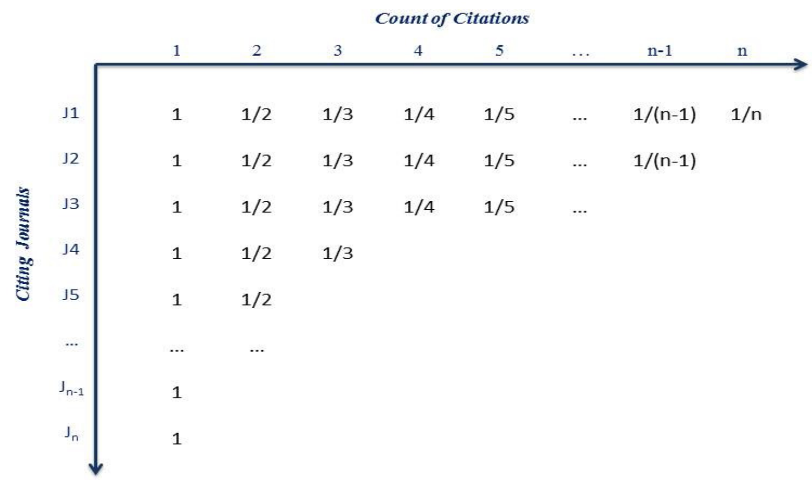

Figure 1: Calculation of the Tapered Diffusion Index

To illustrate further, we consider two articles A and B. Article A has received 20 citations, and article B has received 18 citations. Using the normal way of counting citations, article A has more impact. However, if we subject the citation counts to the diffusion impact, we may obtain a different picture. Assume that among the 20 citations received by article A, 5 are from journal J1, 3 from journal J2, 2 from journal J3 and one each from journals J4 to J13. While article B has received 18 citations: 4 from journal J1, 2 from journal J2 and 12 from other journals.

According to the definition of the tapered diffusion index, the diffusion impacts of A and B are:

D_T (A) = (1+1/2+1/3+1/4+1/5) + (1+1/2+1/3) + (1+1/2) + 10 = 14.62

D_T (B) = (1+1/2+1/3+1/4)+(1+1/2)+12 = 15.58

Thus, article B has a higher diffusion impact than article A.

A DATA SET FROM THE *JOURNAL CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS*

Chemometrics is the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analyzing chemical data. *Chemometrics and Intelligent Laboratory Systems* (CILS) is a highly interdisciplinary journal (classified in 6 Journal Citation Report categories) focusing on the methodologies and applications of chemometrics and related fields. All papers published in 2006 in CILS have been used as cited publications. WoS-indexed citations received by each paper are analyzed (data retrieved on March 15, 2012). Furthermore, citations received from the journal itself (CILS) and citations from conference proceedings are removed to avoid bias.

Table 1 lists the top 30 papers in CILS (2006) according to their D_T values. We can observe that D_T provides a different ranking result compared with that of pure citation counts. For instance, the third paper (*Melssen, W. et al., 2006, 83(2), 99-113*) in the D_T ranking with a high D_T value (24.58) has actually received fewer citations (27 citations) than "*Morales, AH. et al., 2006, 81(2), 180-187*" (30 citations) and "*Mok, DKW. et al., 2006, 82(1-2), 210-217*" (30 citations). The latter two papers have lower D_T due to the fact that their citation distributions are more concentrated. Another interesting pair of papers are "*Ustun, B. et al., 2006, 81(1), 29-40*" and "*Wang, GQ. et al., 2006, 82(1-2), 137-144*". These two papers have both received 22 citations. However, there is a big difference between their D_T values (20.33 vs. 16.25). A closer look reveals that there are 19 individual journals among the citing list of the first paper, and only 12 journals have cited the latter one.

Table 1: Top 30 Papers in CILS (2006) based on D_T

Paper	Rank	D_T	Rank	Citations
Rossi, F. et al., 2006, 80(2), 215-226.	1	32.90	1	43
Anderssen, E. et al., 2006, 84(1-2), 69-74.	2	31.12	2	38
Melssen, W. et al., 2006, 83(2), 99-113.	3	24.58	5	27
Morales, AH. et al., 2006, 81(2), 180-187.	4	22.62	3	30
Brown, CD. et al., 2006, 80(1), 24-38.	5	22.00	6	24
Mok, DKW. et al., 2006, 82(1-2), 210-217.	6	20.73	4	30
Ustun, B. et al., 2006, 81(1), 29-40.	7	20.33	8	22
Puxty, G. et al., 2006, 81(2), 149-164.	8	20.08	7	24
Chiang, LH. et al., 2006, 81(2), 109-119.	9	18.33	12	20
Halstensen, M. et al., 2006, 84(1-2), 88-97.	10	18.00	13	20
Zeaiter, M. et al., 2006, 80(2), 227-235.	11	17.83	10	21
Ni, YN. et al., 2006, 82(1-2), 241-247.	12	17.83	14	20
Gutes, A. et al., 2006, 83(2), 169-179.	13	17.50	17	19
Zhao, CX. et al., 2006, 82(1-2), 218-228.	14	16.83	21	18
Galvao, RKH. et al., 2006, 81(1), 60-67.	15	16.67	11	21
Davis, RA. et al., 2006, 81(1), 50-59.	16	16.67	15	20
Gan, F. et al., 2006, 82(1-2), 59-65.	17	16.50	22	18
Barrero, MA. et al., 2006, 80(1), 67-76.	18	16.50	25	17
Wang, GQ. et al., 2006, 82(1-2), 137-144.	19	16.25	9	22
Whelehan, OP. et al., 2006, 84(1-2), 82-87.	20	15.50	26	17
Shen, HL. et al., 2006, 82(1-2), 276-282.	21	15.50	29	16
Granitto, PM. et al., 2006, 83(2), 83-90.	22	15.00	30	16
Kasemsumran, S. et al., 2006, 82(1-2), 97-103.	23	14.50	18	19
Holm-Nielsen, JB. et al., 2006, 83(2), 114-126.	24	14.08	27	17
Aguado, D. et al., 2006, 84(1-2), 75-81.	25	14.00	31	16
Mazerolles, G. et al., 2006, 81(1), 41-49.	26	13.50	32	16
Liu, X. et al., 2006, 82(1-2), 8-14.	27	13.00	37	13
Durante, C. et al., 2006, 83(1), 54-65.	28	12.83	19	19
Harrington, PD. et al., 2006, 82(1-2), 283-293.	29	12.78	20	19
Camacho, J. et al., 2006, 81(2), 127-136.	30	12.76	16	20

Finally, the citation distributions of two individual papers are shown in Table 2. The two papers in Table 2 have received 19 citations each. However, the distribution of citations is quite different: on the left, almost half of the citations are from one particular journal (*Anal Chim Acta*); while on the right, the citations are more evenly spread. The differences of the citations distribution is according to the differences in their D_T values. The paper on the left has a D_T of 12.83 and is ranked at the 28th position, while the paper on the right reaches a D_T value of 17.50, with a ranking of 13th in Table 1.

Table 2: Distribution of Citing Sources of Two Individual Papers from CILS (2006)

Durante, C. et al., 2006, 83(1), 54-65.		**Gutes, A. et al., 2006, 83(2), 169-179.**	
Total citations: 19	D_T:12.83	Total citations: 19	D_T:17.50
Citing Source Titles	Citations	Citing Source Titles	Citations
Anal Chim Acta	9	*Analyst*	2
Anal Biochem	1	*Anal Chim Acta*	2
Cement Concrete Comp	1	*Talanta*	2
Chem Rev	1	*Adv Intel Soft Comput*	1
Electroanal	1	*Anal Sci*	1
Environ Monit Assess	1	*Appl Soft Comput Prax*	1
Food Res Int	1	*Appl Soft Comput*	1
Int J Food Prop	1	*Biolo Sign Pro Chem Sens*	1
J Exp Bot	1	*Electroanal*	1
J Food Eng	1	*Electrochim Acta*	1
J Food Sci	1	*Expert Syst Appl*	1
		J Food Eng	1
		Neural Comput Appl	1
		Neural Netw World	1
		Rev Sci Instrum	1
		Stud Comput Intel	1

CONCLUSION AND DISCUSSION

We proposed a tapered diffusion index to measure the diffusion impact of individual papers. From the theoretical analysis and experimental examples, the D_T index proved an efficient function for differentiating the scope of citation diffusion. The more citing journals (large diffusion of scientific ideas) the higher the D_T index. As a preliminary study, we only used a simple example with individual journals as the diffusion units. The disparity of the citing journals, namely, how similar or dissimilar these individual journals are, however, was not taken into account. It means, a citing journal from a new subject field could be given a higher credit than that from the same subject fields, and since the subject system is a hierarchical one, the credits could even be assigned accordingly. Furthermore, we intend to use other citing units, for instance, the individual subject disciplines, or the individual authors, as a citation diffusion factor, in future studies. Finally, if the tapered diffusion index would be used as a measure on a research evaluation exercise, field-normalization should be applied.

ACKNOWLEDGEMENT

I am deeply grateful to Prof. Ronald Rousseau for his insightful discussions and suggestions. I would like to acknowledge the support from the National Natural Science Foundation of China under Grant 71103064.

REFERENCES

Anderson, T.R., Hankin, R.K.S. and Killworth, P.D. 2008. Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, Vol. 76, no. 3: 577-588.

Ajiferuke, I. and Wolfram, D. 2010. Citer analysis as a measure of research impact: library and information science as a case study. *Scientometrics*, Vol. 83, no. 3: 623-638.

Bettencourt, L.M.A., Cinron-Arias, A., Kaiser, D.I. and Castillo-Chavez, C. 2006. The power of a good idea : quantitative modeling of the spread of ideas from epidemiological models. *Physica A*, Vol. 364: 513-536.

Bacchiocchi, E. and Montobbio, F. 2009. Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations. *Journal of Technology Transfer*. Vol. 34, no. 2: 169-181.

Chen, C.M. and Hicks, D. 2004. Tracing knowledge diffusion. *Scientometrics*, Vol. 59, no. 2: 199-211.

Franceschini, F., Maisano, D., Perotti, A. and Proto, A. 2010. Analysis of the ch-index: an indicator to evaluate the diffusion of scientific research output by citers. *Scientometrics*, Vol. 85, no. 1: 203-217.

Garfield, E., and Sher, I.H. 1963. New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, Vol. 14, no. 3: 195-201.

Hirsch, J.E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, Vol.102: 16569-16572.

Kiss, I.Z., Broom, M., Craze, P.G. and Rafols, I. 2010. Can epidemic models describe the diffusion of topics across disciplines? *Journals of Informetrics*, Vol. 4, no. 1: 74-82.

Liu, Y.X. and 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*, Vol. 61, no. 2: 340-351.

Rousseau, R., Liu, Y.X. and Ye, F.Y. 2012. A preliminary investigation on diffusion through a layered system. *Journal of Informetrics*, Vol. 6, no. 2: 177-191.

Yu, G., Wang, M.Y. and Yu, D.R. 2010. Characterizing knowledge diffusion of nanoscience & nanotechnology by citation analysis. *Scientometrics*, Vol. 84, no. 1: 81-97.

Zhang, L., Thijs, B. & Glänzel, W. 2011. The diffusion of H-related literature. *Journal of Informetrics*, Vol. 5, no. 4: 583-593.