Calculating the outgrow index and similar structural indicators: A simple software program for visualizing outcomes

Brendan Rousseau[1] and Ronald Rousseau [2,3]

[1] OAK-IT, Eindhoven, the Netherlands

Email: info@oak-it.nl

[2] Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, 2000 Antwerpen, Belgium

[3] KU Leuven, 3000 Leuven, Belgium

Email: ronald.rousseau@uantwerpen.be

Abstract

It is shown that structural indicators, such as the outgrow index, can also be used in the context of diffusion or interdisciplinarity studies. We, moreover, provide a simple software program to calculate and visualize the results.

Keywords: outgrow index; multi-generation citation network; diffusion; interdisciplinarity; software program

1. Introduction: the outgrow index (C-R index) and related indices

We introduced the outgrow or C-R (citations of references) index in (Rousseau & Hu, 2010). This indicator has been put in a broader context and related indicators (R-R, R-C, C-C) were introduced in (Rousseau, 2011; Hu et al., 2012). We briefly recall the definition of these indicators. Consider an article citation network and focus on one specific target article: the ego, as it is called in network theory (Wasserman & Faust, 1994). We refer to this target article as A and consider A's reference list, denoted as Ref(A). The length of A's reference list, i.e. its number of references, is denoted as TRef(A). Article A and all articles in its reference list form a set, denoted as ER(A) = Ref(A) ∪ {A}, where ER stands for the extended reference list. We will attach a positive number to each element of ER(A), and this in different ways. This will allow us to rank the members ER(A) and we will characterize the relative position of A in this ranked list. Note that references correspond to outlinks in the 'cites' network. As a first case we determine for each element of ER(A) the number of articles by which it is cited. Next we rank all elements in ER(A) according to its number of received citations. Finally the position of A in this list is characterized by its citations-of-references or outgrow number, defined as:

$$CR(A) = 1 - \frac{R(A)}{TRef(A) + 1}$$

where R(A) denotes the rank of A in this list. In case of ties we use an average rank. CR(A) is always a number between zero (included) and one (not included). When A is ranked first its CR(A) is equal to 1 – 1/(TRef(A)+1). This implies that being the first among 100 is considered ‘better’ than being the first among 10. Recall that articles that cite an article from A’s reference list form the set of all articles which are bibliographically coupled with A (this set may, or may not, depending on the application, include article A itself).

Instead of considering the number of citations received by each element in ER(A) we may also take the number of references in each of these articles. This is yet another number associated to ER(A). In this way we obtain another ranked list and we can determine a reference-reference index, RR(A), defined as RR(A) = 1 – R’(A)/(TRef(A)+1), where R’(A) is the rank of A in the list determined by the number of references. This approach uses references of references, hence two generations of references, as considered e.g. in (Hu et al., 2011).

Next we consider all articles that cite A and denote this set as Cit(A). The number of elements in Cit(A) is denoted as TCit(A). This means that, taking A’s point of view, we now follow the ‘is cited by’ relation. Article A and all citing articles form a set, denoted as EC(A) = Cit(A) ∪ {A}. Again we will attach a positive number to each element of EC(A), leading to a ranking of the elements of EC(A). As was the case for the ‘cites’ relation, a number between zero and one will be used to characterize the relative position of A in this ranked list.

First we determine for each element in EC(A) the number of articles by which it is cited. Next we rank each element in EC(A) according to its number of received citations. Finally the position of A in this list is characterized by its citation-citation number CC(A) = 1 – R”(A)/(TCit(A)+1), where R”(A) denotes the rank of A in this new list. Again an average rank is used in the case of ties. Finally, instead of considering the number of citations received by each element in EC(A) we can also take the number of references in each of these articles. In this way we obtain another ranked list and determine a reference-citation index RC(A) using the relation RC(A) = 1 – R’”(A)/(TCit(A)+1), where R’”(A) is the rank of A in the list determined by the number of references. Articles that are cited by an article citing A form the set of all articles which are co-cited with A. The meaning of these indicators has been discussed in (Hu et al., 2012).

We recall that the number of elements in Ref(A) and in EC(A) is fixed, while the number of elements in Cit(A) and EC(A) may – and usually does – increase. This implies that the denominator of the indicators CC(A) and RC(A) may increase.

2. Diffusion

Instead of number of different citing articles (usually referred to as number of citations), one may count different citing authors, different citing journals, different citing institutions, countries, WoS categories, ethnicities, and similar entities. Doing this, rankings, now based, e.g. on different WoS categories, and the resulting indices connect our approach to diffusion theory (Liu & Rousseau, 2010; Liu et al.,2012) and the study of interdisciplinarity (Rafols & Meyer, 2010), as determined by references coming from different origins.

It is easy to see how this is done in the case of a diffusion outgrow or C-R index. An example is given in the following sections. One considers all elements in ER(A) and determines for each article in this set the number of different authors citing it, the number of different countries these authors belong to, the number of different journals, or, more generally, sources, in which citations occur and the WoS categories to which these sources belong.

In the case of the R-R index one again considers all elements in ER(A) and determines for each article the number of different cited authors, different cited sources and so on. For the C-C index one considers the set EC(A). For each of these elements one considers the number of different authors citing it, the number of different countries these authors come from (based on their institutional address or, maybe, their ethnicity), the number of different sources citing this document and so on. Finally, an R-C diffusion index is determined by considering the set EC(A) and for each of its elements the number of different cited authors, sources, countries and so on.

3. An example: data

In this section we provide an example of the calculations of the outgrow index for other citing items than citing articles, illustrating the link with diffusion theory.

Inspired by recent progress in three-dimensional printing we performed the following search in Thomson Reuters' Web of Science on June 12, 2013:

TS = ("3D printing" OR "3-D printing") AND WC = Materials Science

The first term retrieved 391 items. "AND"-ing with the WoS category of *Materials Science* resulted in 141 items (other ones were mostly related to the medical sciences). As an example for further calculations we choose the most-cited article in this set and observe that also this one is related to medicine. Note that the only purpose of this is to provide an example of how our approach works. We have no intention to obtain results in the field of 3D printing and accept data as provided by Thomson Reuters without questioning their validity.

Table 1. WoS data (partial) for most-cited 3D printing article in the WoS category *Materials Science*

Scaffold development using 3D printing with a starch-based polymer

Author(s): Lam, CXF; Mo, XM; Teoh, SH; Hutmacher, DW

Source: MATERIALS SCIENCE & ENGINEERING C-BIOMIMETIC AND SUPRAMOLECULAR SYSTEMS Volume: 20 Issue: 1-2 Special Issue: SI Pages: 49-56 Article Number: PII S0928-4931(02)00012-7 DOI: 10.1016/S0928-4931(02)00012-7 Published: MAY 31 2002

Times Cited: 152 (from Web of Science); but only 130 available in the WoS available in Flanders. As in Flanders we do not have access to the Book Citation Index, we assume that the remaining 22 citations occurred there.

Cited References: 18

Conference: International Conference on Materials for Advanced Technologies (ICMAT2001) Location: SINGAPORE, SINGAPORE Date: JUL 01-06, 2001

Sponsor(s): Mat Res Soc; Int Union Mat Res Soc

Document Type: Article; Proceedings Paper

Language: English

Author Keywords: three-dimensional printing; rapid prototyping; starch-based scaffolds; tissue engineering

KeyWords Plus: Amylopectin; amylose

Reprint Address: Hutmacher, DW; Natl Univ Singapore, Dept Orthoped Surg, Singapore 117548, Singapore.

Using our terminology we see that here # ER(A) = 19. In the – for us - available part of the Web of Science we found for this particular article:

130 citing documents

446 different citing authors

33 citing countries/territories

84 citing sources (mainly journals and conference proceedings)

41 citing WoS categories

When the citing documents were determined, the remaining data were obtained by using the “Analyze Results” functionality. In order to have a short time series we performed the analysis for WoS categories in three steps: first the period strictly before 2002 (resulting in zero categories, but this will not be the case for the article’s references); next for the period strictly before 2008 (27 citing categories) and finally for the whole period (41 citing categories). Observe that this type of time series is not included in the typology introduced in (Liu & Rousseau, 2008).

Then we collected the corresponding data for the 18 references of our target = ego article. Results are shown in Table 2. One may observe that there are always more citing authors than citing papers, although this is not a logical requirement. Clearly, many citing papers have multiple authors and, in general, there are not many re-citations, using the terminology of (Cronin & Shaw, 2001; White, 2001). Table 3 shows the results for the short time series related to different citing categories. We artificially started with a number zero for the target article and a number one for articles that are only cited by the target article.

Table 2. Numbers of different citing entities, related to the article shown in Table 1. (From the third column on data refer to the WoS as it is available in Flemish universities)

	all WoS citations	citing papers	citing authors	citing countries	citing sources	citing categories
Target article	152	130	446	33	84	41
1. Title: Three-dimensional printing techniques	18	16	49	9	15	14
2. Title: Scaffolds in tissue engineering bone and cartilage	1516	1452	4513	60	465	92
3.Title: [not available] Jane	1	1	4	1	1	1
4. Title: Incompatibility of amylose and amylopectin in aqueous-solution	140	135	274	32	49	22
5. Title: Survival and function of hepatocytes on a novel three-dimensional synthetic biodegradable	199	163	644	24	112	52
6. Title: Preparation of interpenetrating networks of gelatin and dextran as degradable biomaterials	57	57	148	16	41	27
7. Title: Layer position accuracy in powder-based rapid prototyping	3	2	6	2	2	3
8. Title: The roles of amylose and amylopectin in the gelation and retrogradation of	560	548	1225	56	111	42

starch						
9. Title: [not available] Nishimura	1	1	4	1	1	1
10. Title: [not available] Paturau	1	1	4	1	1	1
11. Title: A comparison of rapid prototyping technologies	160	153	445	32	106	47
12. Title: Processing and in vitro degradation of starch/EVOH thermoplastic blends	85	80	159	20	44	25
13. Title: 3-dimensional printing – rapid tooling and prototypes directly from a CAD model	238	231	503	21	151	54
14. Title: [not available] Stapley	1	1	4	1	1	1
15. Title: Enhanced crystallization of poly(L-lactide-co-epsilon-coprolactone) during storage at room temperature	40	40	84	14	22	11
16. Title: Solid free-form fabrication of drug delivery devices	109	102	316	18	81	38
17. Title: Effect of pore size and void fraction on cellular adhesion, proliferation, and matrix deposition	297	282	1073	34	130	55
18. Title: [not available] Zobel	132	131	334	38	51	25

Table 3. Numbers of different citing categories in overlapping periods

Different citing categories	2002	2007	2013
target	0	27	41
1. Title: Three-dimensional printing techniques	5	7	14
2. Title: Scaffolds in tissue engineering bone and cartilage	33	69	92
3.Title: [not available] Jane	1	1	1
4. Title: Incompatibility of amylose and amylopectin in aqueous-solution	21	22	22
5. Title: Survival and function of hepatocytes on a novel three-dimensional synthetic biodegradable	31	44	52
6. Title: Preparation of interpenetrating networks of gelatin and dextran as degradable biomaterials	2	10	27
7. Title: Layer position accuracy in powder-based rapid prototyping	1	1	3
8. Title: The roles of amylose and amylopectin in the gelation and retrogradation of starch	30	34	42
9. Title: [not available] Nishimura	1	1	1
10. Title: [not available] Paturau	1	1	1
11. Title: A comparison of rapid prototyping technologies	18	36	47
12. Title: Processing and in vitro degradation of starch/EVOH thermoplastic	10	18	25

blends			
13. Title: 3-dimensional printing – rapid tooling and prototypes directly from a CAD model	30	47	54
14. Title: [not available] Stapley	1	1	1
15. Title: Enhanced crystallization of poly(L-lactide-co-epsilon-coprolactone) during storage at room temperature	2	9	11
16. Title: Solid free-form fabrication of drug delivery devices	18	29	38
17. Title: Effect of pore size and void fraction on cellular adhesion, proliferation, and matrix deposition	7	32	55
18. Title: [not available] Zobel	22	22	25

4. Example: results

The results based on Table 2 are illustrated in Figure 1. Outgrow indices were calculated for different citing papers, authors, countries/regions, sources and categories. The outgrow diffusion values are indicated under the rectangular bars and are, respectively equal to: 0.526, 0.684, 0.737, 0.632 and 0.632. The rectangular bars have a length equal to one (as indicated) and the outgrow diffusion values are also indicated on these bars by the horizontal line drawn through the small rectangles representing the ego. The number in these small rectangles is the position of the ego-value among the values for the other members of ER(A) (or EC(A), depending on the case). Note that this rank can be a decimal number as we use an average in case of ties.

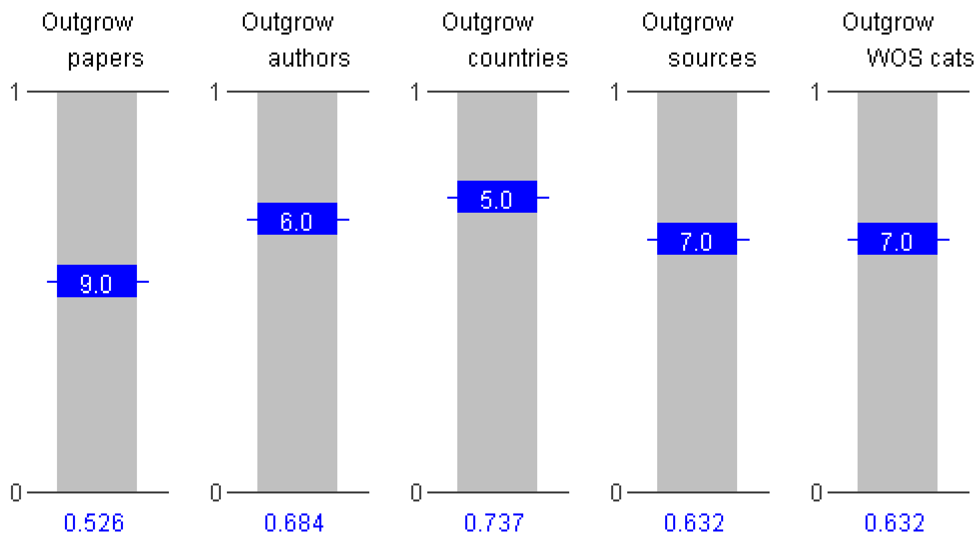

Figure 1. Different diffusion outgrow indices for the 3D-printing article A

Fig.2. illustrates the evolution of the diffusion outgrow index of article A over different time periods.

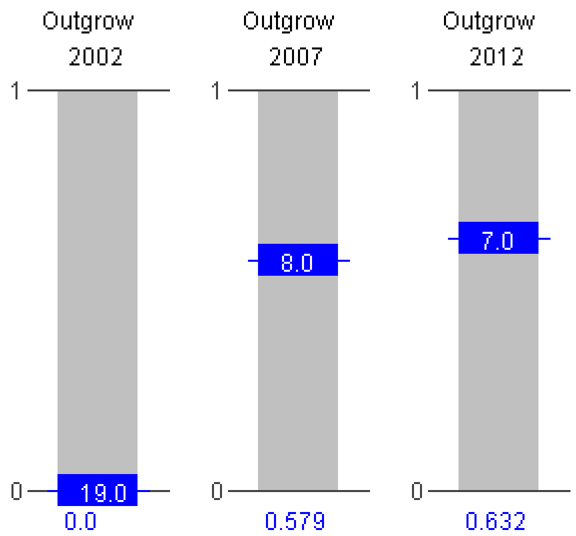

Figure 2: Evolution of diffusion outgrow index of A for WoS categories

5. A software program

We wrote a small program that led to the outputs shown in figures 1 and 2. It is available at http://crindex.com/. The input to this program is an Excel file in Microsoft Excel 97-2003 format. Note that this format is a requirement. Colleagues using a more recent version should save as 97-2003 format. Input must be given as shown in Fig.3, where the word 'Type' is in cell A1. Yet, the actual types and 'years' are just labels and may be replaced by any useful information that the user wants to display in the resulting figure. Our Fig.2 used years, but Fig.1 used categorical names.

Type	CC	CC	CC	CC
Date	2010	2011	2012	2013
Ego	0	5	10	30
Data	256	260	260	260
	456	1454	1454	1454
	10	11	11	11
	20	21	21	22
	1	2	4	6
		1	10	30
				30

Figure 3. Example of input data

We note that if an article is just published, it did not receive any citations, and this situation may continue for some years (or even forever). Then the C-C index is 1 – 1/1 = 0. This is also true for the R-C index. Hence, in this case, the ego row in Fig. 3 contains only zeros and the Data part in Fig. 3 stays empty. Also then our program yields the correct result as illustrated in Fig.4.

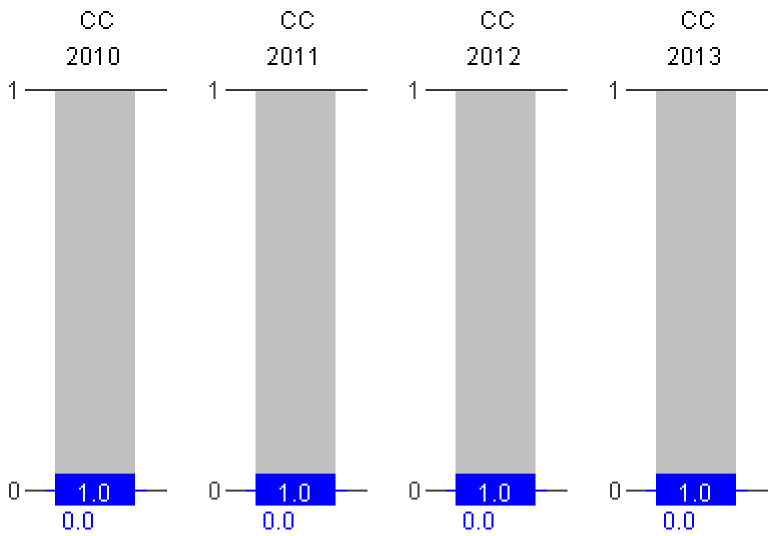

Figure 4; CC-index for an uncited article

6. Conclusion

In this article we have shown how structural indicators introduced in (Rousseau, 2011; Hu et al., 2012) can be used in the context of diffusion studies. We, moreover, provided a dedicated software program to calculate and illustrate these indicators.

References

Cronin, B., & Shaw, D. (2001). Identity-creators and image-makers: using citation analysis and thick description to put authors in their place. In: M. Davis & C. S. Wilson (Eds.), *8th International Conference on Scientometrics and Informetrics. Proceedings ISSI 2001* (pp.127-138). BIRG (UNSW), Sydney, Australia.

Hu, XJ., Rousseau, R., & Chen, J. (2011). On the definition of forward and backward citation generations. *Journal of Informetrics*, 5(1), 27-36.

Hu, XJ., Rousseau, R., & Chen, J. (2012). Structural indicators in citation networks. *Scientometrics*, 91(2), 451- 460.

Liu, YX., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, 68(1), 2012, 31- 44.

Liu, YX., & Rousseau, R. (2008). Definitions of time series in citation analysis with special attention to the h-index. *Journal of Informetrics*, 2(3), 202-210.

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.

Rafols, I. & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.

Rousseau, R. (2011). Algebraic structures in the ego article citation network. In: E. Noyons, P. Ngulube & J. Leta (Eds.), *Proceedings of the ISSI 2011 Conference* (pp.737-741). ISSI, Leiden University & University of Zululand.

Rousseau, R. & Hu, XJ. (2010). An outgrow index. *Annals of Library and Information Studies*, 57, 287-290.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis*. Cambridge: Cambridge University Press.

White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology,* 52(2), 87-108.