Bibliometric Indicators of Interdisciplinarity Exploring New Class of Diversity Measures

Alexis-Michel Mugabushaka[1], Anthi Kyriakou & Theo Papazoglou

[1] *Alexis-Michel.MUGABUSHAKA@ec.europa.eu*
European Research Council Executive Agency[1], Brussels, (Belgium)

Abstract
In bibliometrics, interdsicsiplinatity is often measured in terms of the "diversity" of research areas in the references that an article cites. The standard indicators used are borrowed mostly from other research areas, notably from ecology (biodiversity measures) and economics (concentration measures). This paper discusses a new class of measures, which are used in the study of biodiversity and especially the Leinster-Cobbold diversity measure (Leinster Cobbold 2010). We present a case study based on previously published dataset of 12 journal articles from a group of five researchers from the bio-nano science described and published by Rafols and Meyer (2010). We replicate the findings of this study to show that the various interdisciplinarity measures are in fact special cases of the Cobbold-Leinster diversity measure. The paper discusses some interesting properties of the Cobbold-Leinster diversity measure, which makes it appealing in the study of disciplinary diversity than the standards diversity indicators used as proxy for interdisciplinarity.

Conference Topic
Indicators

Introduction
Considerable efforts have been made to operationalize and measure the concept of interdisciplinarity in bibliometrics (Porter et al., 2007; Rafols & Meyer, 2010). The most commonly used indicators of interdisciplinarity are mostly borrowed from other research areas, notably from ecology (biodiversity measures) and economics (concentration measures).

The purpose of this paper is to bring to discussion a relatively new class of diversity indicators which are used in ecology but so far not been used to investigate disciplinary diversity. Drawing from the literature in ecology, the paper highlights important properties of those measures and discusses how they can help the bibliometric study of interdisciplinarity.

The paper is divided in three parts. The next section briefly presents indicators of interdisciplinarity in bibliometrics. The second section discusses the development of new class of diversity measures used in ecology and presents the Leinster-Cobbold diversity measure, highlighting its properties and why they are relevant for bibliometric usage. The third section presents a case study to illustrate the potential of Leinster-Cobbold diversity indicators as a measure of disciplinary diversity.

Currently used Bibliometric indicators of interdisciplinarity
Bibliometric analyses of interdisciplinarity take as unit of analysis a scientific paper and assume that the extent to which it integrates elements of different disciplines is reflected in the references it cites. References in scientific papers are expected to reflect various aspects of interdisciplinary because researchers will credit what they are indebted to other disciplines: conceptually (concepts, ideas and approaches from other disciplines); analytically (methods for defining, collecting and analyze data) and technically (tools developed in other fields).

---

[1] The views expressed in this paper are the authors'. They do not necessarily reflect the views or official positions of the European Commission, the European Research Council Executive Agency or the ERC Scientific Council.

Porter et al. (2007) developed the integration score as measure of interdisciplinary which takes into account not only the distribution of the cited references in different subject categories but also how closely related those subject categories are (see also Porter et al., 2006; Porter et al., 2008). In line with Porter's conceptualization, Rafols and Meyer (2006, 2010) introduced a new set of bibliometric indicators to quantify the disciplinary diversity of references as a proxy measure of interdisciplinarity. They are mostly based on the general framework for analyzing diversity developed by Stirling (2007). The most commonly used indicators are summarized in table 1. We note that there are also efforts to use network based measures (Rafols & Meyer, 2010; Karlovčec & Mladenić, 2015) but here we focus on diversity measures.

Table 1. Most common indicators of interdisciplinarity in bibliometric studies .

Indicators	Definition/description	
Variety	The number of different disciplines that a given paper cites**	N
Shannon entropy	As measure of diversity the Shannon Entropy quantifies how diverse the subject categories in the references are.	$$H_{SH} = -\sum_{i=1}^{S} p_i \log p_i$$ Where p_i is the proportion of elements in a system and S the number of elements in the system.
Simpson diversity	It measures how references are distributed (or concentrated) in subject categories.	$$H_{GS} = 1 - \sum_{i=1}^{S} p_i^2$$ Where p_i is the proportion of elements in a system and S the number of elements in the system
Rao-Stirling index	Can be understood as the Simpson diversity which takes into account distance/similarity (between disciplines).	$$= \sum_{i,j} d_{i,j}\, p_i p_j$$ Where di,j is the distance between the ith and jth element in the distance matrix and pi is the proportion of element i

Source: Rafols & Meyer 2010, p. 267 **Its variants includes normalization by the total numbers of subject categories or the shares of references outside a given subject category

New classes of diversity measures in ecology

Effective numbers

The diversity measures listed in table are also among the commonly used indicators of biodiversity in ecology. However, they have recently faced strong criticisms (Jost, 2006; Chao & Lou, 2012).

The main criticism is that those measures fail to satisfy the most basic property that ecologist would expect from a meaningful measure of diversity, namely the replication principle. In simple term, the "replication principle" states that if you have two completely distinct communities (i.e. without any overlap in the species) with each community having a diversity measure X, one would expect that combining those two communities would result in a community with a diversity measure 2X.

One category of diversity measures, which satisfy this replication principle is the so called "Hill-numbers" (also called "effective numbers of species"). They can be interpreted as the "number of equally abundant specifies that are needed to give the same value of the diversity measure (Chao & Lou, 2012, p. 204).

The Hill numbers have some properties that other measures of diversity based on entropy lack:

- They satisfy the replication principles. i.e. two communities with each 4 effective numbers of species will – if pooled together – result in a community whose effective number equal 8. They therefore give logically consistent answers.
- Their linear scale makes it easier to interpret the magnitude of their change.
- In addition to this this advantage of intuitive consistency, they have another interesting property that we call "unifying framework status". Jost (2006) has shown that practically all traditional measures of diversity can be easily converted to "Hill numbers/ "effective numbers" and vice-versa.

Leinster-Cobbold Diversity Measure

Leinster and Cobbold (2012) developed a measure, which extends the Hill numbers to include the similarities/differences between species. Their measure – called here the Leinster-Cobbold Diversity Measure - can be used with any similarity coefficient between each pair of the species. This extends the scope of its usage to other contexts such disciplinary diversity in bibliometrics. In the following, we first provide its formal definition and discuss its properties as well as its relation to other diversity measures. In the next section we provide a case study of its use in the study of disciplinary diversity.

Consider a system with S elements with relative frequencies translating in estimated probabilities p = (p1, …, pS) so that $\sum_{i=1}^{S} p_i = 1$

The similarity between the elements is encoded in an S x S Matrix Z.

Z = (Z_(i,j)), with Z_(i,j) measuring the similarity between the ith and jth elements.

Whereby 0 ≤Z_(i,j) ≤1, with 0 indicating total dissimilarity and 1 indicating identical elements.

The Leinster-Cobbold diversity measure is defined as

$$
{}^{q}D^{Z}(\boldsymbol{p}) = \begin{cases} \left(\sum_{i:p_i>0} p_i (Z\boldsymbol{p})_i^{q-1} \right)^{\frac{1}{1-q}} & q \neq 1, \\ \prod_{i:p_i>0} (Z\boldsymbol{p})_i^{-p_i} & q = 1, \\ \min_{i:p_i>0} \frac{1}{(Z\boldsymbol{p})_i} & q = \infty. \end{cases}
$$

where

$$
(Z_p)i = \sum_{j=1}^{S} Z_{i,j} p_i
$$

q is in number in range 0 ≤ q ≤ Infinity. It is called a sensitivity parameter and control the relative emphasize that the user wishes to place on common and rare species.

Case Study: Using the Leinster-Cobbold Diversity as a measure of disciplinary diversity

In our view, there are three main advantages in adopting the Leinster and Cobbold diversity measure in the study of disciplinary diversity as well:

- First, Leinster and Cobbold (2012) have discussed the relation between this measure and other diversity measures and showed that they can be seen as its special cases. The advantage here would be to have a single formula which would replace the Shannon entropy, the Simpson Diversity and the Rao-Stirling Index used in bibliometrics.
- Second, because the Leinster and Cobbold measure quantifies diversity on a spectrum which depends on how much emphasis should be given to relatively rare elements (sensitivity parameter q), it provides potentially more information than measures which consider only one value of this sensitivity parameter.
- The third advantage is the intuitive consistency of the Leinster and Cobbold measure. Because it directly produces "effective numbers" which obey the replication principle, the values can be easily interpreted and compared. Consider two publications: one with references from 2 (unrelated) categories and the other with reference from 4 (unrelated) categories. With the Leinster and Cobbold measure, they can be compared to say that the second has a twice as large diversity in references as the first one.

In the following, we present a case study to illustrate the potential of Leinster-Cobbold diversity profiles in quantifying disciplinary diversity.

Disciplinary diversity of selected papers in bio-nanoscience (Rafols & Meyer 2010)

The case study is based on a dataset of 12 journal articles from a group of five researchers from the bio-nano science described and published by Rafols and Meyer (2010). For those 12 papers, Rafols and Meyers published the distribution of their references in Web of Science Categories (Rafols & Meyers, 2010; p. 276, Table 3) as well as the scores on various indicators of diversity (ibid. p. 277, Table 4). The similarity/distance measures between the Web of Science subject categories are taken from the supplementary materials to the paper[2] by Chavarro et al. (2014).

Table 2. Diversity measures for the 12 papers in Rafols and Meyer (2010).

	not considering distance/similarity						considering distance/similarity					
sensitivity parameter q	**0**	**1**	**2**	**3**	**4**	**Inf**	**0**	**1**	**2**	**3**	**4**	**Inf**
Column no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Papers												
Fun95	16	6,452	4,553	3,989	3,740	3,106	1,656	1,422	1,329	1,288	1,266	1,188
Koj97	17	5,526	4,232	3,848	3,652	2,880	1,479	1,284	1,225	1,203	1,192	1,143
Ish98	15	5,003	3,499	2,990	2,741	2,156	1,342	1,229	1,192	1,176	1,167	1,108
Noj97	16	4,532	3,120	2,665	2,447	1,967	1,280	1,172	1,141	1,128	1,122	1,077
Yas98	16	4,466	3,003	2,537	2,327	1,890	1,231	1,158	1,133	1,122	1,115	1,072
Oka99	16	4,857	3,814	3,557	3,439	3,062	1,253	1,190	1,165	1,154	1,148	1,108
Kik01	14	4,944	3,857	3,534	3,364	2,673	1,251	1,195	1,169	1,155	1,148	1,102
Sak99	14	5,103	4,040	3,764	3,641	3,184	1,245	1,181	1,159	1,149	1,143	1,098
Bur03	14	4,697	3,536	3,230	3,086	2,571	1,178	1,142	1,127	1,120	1,115	1,082
Tom00	15	4,841	3,846	3,625	3,530	3,028	1,227	1,165	1,145	1,136	1,132	1,095
Tom02	14	4,849	3,864	3,630	3,531	3,192	1,242	1,180	1,159	1,149	1,143	1,103

This case study illustrates that the various diversity measures are in fact special cases of the Leinster-Cobbold diversity profiles. We do this by replicating the diversity measures computed by Rafols and Meyer 2010 using the Leinster-Cobbold diversity profiles. We first compute the values of the Leinster Cobbold measure using different values for the sensitivity parameters (0, 1, 2, 3, 4 and infinity) and in two variants: without taking into account the

[2] http://www.interdisciplinaryscience.net/topics/interdisciplinarity-and-local-knowledge

distance/similarity between the subject categories (i.e. the matrix Z is an identity matrix) and by taking into account the distance/similarity between the subject categories (using the similarity data provided in supplementary materials of Chavarro et al. (2014). Using the conversion formulas in the first row of Table 3, we use those Leinster Cobbold values to derive the diversity measures provided in Rafols and Meyer 2010 (table 4 on page 277). The Table 3 below replicates the diversity values reported in Rafols and Meyer 2010. There are some differences, which are due to rounding but also to the fact that some indicators in Rafols and Meyer (2010) were given in normalized form.

Table 3. Deriving diversity measures commonly used in bibliometrics from the Leinster-Cobbold values.

	Variety	**Gini-Simpson**	**Shannon**	**Rao**
computation	**Col 1**	**1- (1/Col 3)**	**ln(Col 2)**	**1- (1/Col 9)**
Papers				
Fun95	16	0,78	1,86	0,25
Koj97	17	0,76	1,71	0,18
Ish98	15	0,71	1,61	0,16
Noj97	16	0,68	1,51	0,12
Yas98	16	0,67	1,5	0,12
Oka99	16	0,74	1,58	0,14
Kik01	14	0,74	1,6	0,14
Sak99	14	0,75	1,63	0,14
Bur03	14	0,72	1,55	0,11
Tom00	15	0,74	1,58	0,13
Tom02	14	0,74	1,58	0,14
Yil04	16	0,76	1,68	0,16

Concluding remarks

In bibliometrics, the interdisciplinarity is operationalized in terms of the diversity of the references in a scholarly article. The most commonly used indicators are derived from the fields of ecology (biodiversity measures) and from the fields of economics (concentration measures). We discuss a new class of biodiversity measures – the "effective numbers" - which not only generalize most of other diversity measures but also have some proprieties which make their interpretation intuitively consistent with the concept of diversity Jost (2006). They were further developed by Leinster-Cobbold (2012) to take into account the similarity/distance of elements (species) in a system (community). We provide an example on how the bibliometric indicators of interdisciplinarity are in fact special cases of this more general Leinster Cobbold indicator.

Future work should not only take a closer look at their statistical properties (distribution, parameters etc.) but also test their reliability and validity. In particular, it would be of interest to analyze how sensitive the indicators are to various degree of granularity of different classifications of research disciplines and to assess extent to which they depend on measures of distances used.

Acknowledgments

We thank Ismael Rafols for helpful comments on an earlier draft of the paper and Diego Chavarro for making the similarity matrix freely available.

References

Chao, A., & Jost, L. (2012). *Diversity measures. In Encyclopedia of Theoretical Ecology* (Eds. A. Hastings and L. Gross), pp. 203-207, Berkeley: University of California Press.

Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: evidence from a developing country. *Research Evaluation*, *23*(3), 195-209.

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*, 363-375.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, *88*, 2427–2439.

Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics*, *68*,925–928.

Karlovčec, M., & Mladenić, D (2015) Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics*, *102*(1), 433-454.

Leinster T., & Cobbold CA. (2012). Measuring diversity: the importance of species similarity. *Ecology*, *93*(3):477-489.

Porter et al., (2006). Interdisciplinary research: meaning, metrics and nurture. *Research Evaluation*, *15*(3), 187-195.

Porter, A.L. & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719-745.

Porter, A.L., Cohen, A.S., Roessner, J.D., & Perreault, M. (2007), Measuring researcher interdisciplinarity, *Scientometrics*, *72*(1), 117-147.

Rafols, I. & Meyer, M. (2006). Diversity measures and network centralities as indicators of interdisciplinarity: case studies in bionanoscience. *SPRU working paper*.

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

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707-719.