Journal of Informetrics 12 (2018) 510–517

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Ranking scholars: A measure representation☆

Casilda Lasso de la Vega[a], Oscar Volij[b,*]

[a] University of the Basque Country, Spain
[b] Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Received 13 February 2018
Received in revised form 17 April 2018
Accepted 17 April 2018

Keywords:
Scholars
Citation lists
Measure representation
Additive representation

ABSTRACT

A scholar can be identified with his citation list and a scholar ranking is a complete order on the set of scholars. We characterize those scholar rankings that admit a measure representation.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Several papers have recently proposed methods for ranking scholar citation lists. Such a list records the number of citations of each of the scholar's publications. One of the most prominent ranking methods is known as the h-index, which assigns to a scholar the largest number h such that there are h publications with at least h citations. This index was introduced by Hirsch (2005), and was later characterized, among others, by Woeginger (2008), Marchant (2009a), Quesada (2011), Miroiu (2013) and Bouyssou and Marchant (2014). Recently, Chambers and Miller (2014) characterized the family of step-based indices, which includes the h-index. Finally, Perry and Reny (2016) characterized the Euclidean index which is given by the Euclidean length of the citation vectors.

A scholar can be identified with his citation function, the step function that assigns to each of his publications the number of its citations. He can also be identified with the hypograph of this function, namely, with the set of points lying below its graph. As it is, all the above mentioned indices have a measure representation. That is, they select a measure and order scholar citation lists by the measure of their hypographs. In their survey of the literature on scholar rankings, Palacios-Huerta and Volij (2014) ask whether the class of rankings that can be represented in this way could be axiomatically characterized. The objective of this paper is to answer this question.

It turns out that the mathematical results necessary to address this issue have already been developed. Indeed, necessary and sufficient conditions for the existence of an additive representation on finite subsets of general product sets can be found in the literature in different forms (see, for instance, Fishburn, 1970; Krantz, Luce, Suppes, & Tversky, 1971; Scott, 1964; Wakker, 1989). These conditions are what we are looking for. To see this, notice that since a scholar can be identified with the hypograph of his citation function and since hypographs are subsets of $\mathbb{R}^2_+$, the question we are interested in can be

☆ We thank Peter Wakker for directing us to the relevant literature on additive representations. Oscar Volij thanks the Foundations of Economics Analysis I Department at the University of the Basque Country for its kind hospitality. This paper was written during his stay there. The authors also thank the Spanish Ministerio de Economía y Competitividad (project ECO2015-67519) and the Gobierno Vasco (project IT568-13) for research support.

* Corresponding author.
E-mail addresses: casilda.lassodelavega@ehu.es (C. Lasso de la Vega), ovolij@bgu.ac.il (O. Volij).

https://doi.org/10.1016/j.joi.2018.04.003
1751-1577/© 2018 Elsevier Ltd. All rights reserved.

reformulated, when the set of scholars is finite, as that of finding an additive and monotone representation of preferences over a finite subset of $\mathbb{R}^2_+$. In light of this fact, we can use the above-mentioned conditions to formulate an axiom (which we call strong irrelevance) that, together with a mild monotonicity condition, yields a characterization of those rankings that have a measure representation. We are aware that this formulation is merely a translation of the above-mentioned conditions. However, while these conditions are technical and hard to interpret when applied to abstract subsets of product sets, they yield a natural and appealing interpretation when translated into the language of scholar rankings.

To motivate the strong irrelevance axiom consider two university departments with the same number of scholars. Assume that on aggregate they have the same publication records in the following sense. For each scholar s in one department, we can find another scholar in the other department whose most cited publication has the same number of citations as the most cited publication of scholar s. And the same applies to all his other publications as well. Namely, we can find a (possibly different) scholar in the other department whose jth most cited publication has the same number of citations as the jth most cited publication of scholar s. In this case, the axiom will require that no matter how we match the scholars of one department to the scholars of the other, if there is a scholar in the former that is ranked strictly above the corresponding one in the latter, then there must also be a scholar in the latter that is ranked strictly above its match in the former. It turns out that this axiom together with a monotonicity condition characterizes the class of rankings with a measure representation.

Segal (1993) has also addressed the question of the existence of a measure representation but in the context of preferences over lotteries, and he has characterized the preferences that have a measure representation. That is, those for which there is a measure such that one lottery is preferred to another one if the measure of the hypograph of the former is greater than the measure of the hypograph of the latter.[1] The main axiom in Segal's characterization requires that certain slight alterations applied to two lotteries should not reverse their ranking in the preference relation. This seemingly weak axiom, along with continuity and monotonicity, is sufficient to provide the desired measure representation. In our case, though, this axiom is not powerful enough. In fact, it is possible to find robust examples of monotone rankings that satisfy it, but do not have a measure representation. The reason is that although both lotteries and citation lists can be identified with their hypographs, the class of lotteries that Segal (1993) considers is much larger than the class of citation lists we focus on. Indeed, while prizes and probabilities can take real values in Segal's context, the number of publications and citations take only integer values in ours.

We devote the remainder of the paper to applying the additive representation results to the standard scholar ranking problem.

2. Scholar rankings

A scholar is represented by a list $s=(s_1, \ldots, s_m)$ of m non-negative integers. This list can be given several interpretations. For instance, s can be a scholar's publication list where $\{1, \ldots, m\}$ is the set of journals, and for each j in it, s_j is the number of articles published in journal j. Alternatively, s can be a scholar's citation list, where m is the number of publications and for each $j=1, \ldots, m$, s_j is the number of citations obtained by the jth publication. Under a third interpretation the number of positive components represents the number of publications, and for each publication, namely for each $j=1, \ldots, m$ with $s_j>0$, s_j-1 represents the number of its citations. For ease of exposition we will adopt the third interpretation. Nevertheless, the axioms we will discuss are equally appealing under any of the alternative interpretations.

Since under the chosen interpretation it is usually assumed that all publications are equally important, all permutations of a given citation list are equivalent. For this reason we will restrict attention to citation lists that are written in non-increasing order. We denote by $\mathcal{S}_m = \{(s_1, \ldots, s_m) : s_1 \geq s_2 \geq \cdots \geq s_m\}$ the class of scholars with at most m publications, and by $\mathcal{S} = \cup_{m\geq 1}\mathcal{S}_m$ the class of all scholars. In this paper we will focus attention on the following subclasses of scholars. For any two integer numbers $m, n>0$, let $\mathcal{S}_{m\times n} = \{(s_1, \ldots, s_m) \in \mathcal{S}_m : n \geq s_j \ \text{ for } j = 1, \ldots, m\}$. Adopting our interpretation of a scholar, $\mathcal{S}_{m\times n}$ is the class of scholars with at most m publications, each having less than n citations.

We can associate with any scholar $s \in \mathcal{S}$ his *citation function* $F_s : \mathbb{R}_+ \to \mathbb{R}_+$ which is defined by $F_s(x)=s_{\lfloor x+1\rfloor}$, where for each $x \in \mathbb{R}$, $\lfloor x\rfloor$ denotes the greatest integer that is smaller than or equal to x. Fig. 1 depicts the citation functions of scholars $s=(6, 6, 5, 1, 0, 0)$ and $t=(5, 3, 2, 2, 2, 1)$.

The hypograph of a citation function F_s is the set of points in $\mathbb{R}^2_+$ that lie below its graph. Formally,

$$\mathrm{hyp}(F_s) = \{(x, y) \in \mathbb{R}^2_+ : y < F_s(x)\}.$$

We will sometimes refer to $\mathrm{hyp}(F_s)$ simply as the hypograph of s.

For any class of scholars $\mathcal{S}'$, a *scholar ranking on* $\mathcal{S}'$ is a complete and transitive binary relation on $\mathcal{S}'$. Scholar rankings will be denoted by $\succsim$ and the asymmetric and symmetric parts of $\succsim$ will be denoted by $\succ$ and $\sim$, respectively.

[1] Segal (1993) works with the epigraphs of the cumulative distribution functions of lotteries but his analysis could equivalently have been done using the hypographs of the decumulative distribution functions.

Download English Version:

https://daneshyari.com/en/article/6934109

Download Persian Version:

https://daneshyari.com/article/6934109

Daneshyari.com