Journal of Informetrics 11 (2017) 583–594

Contents lists available at ScienceDirect

Journal of Informetrics

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

Measuring cognitive distance between publication portfolios

Ronald Rousseau [a,b,*], Raf Guns [c], A.I.M. Jakaria Rahman [c], Tim C.E. Engels [c,d]

[a] KU Leuven, Dept. of Mathematics, Celestijnenlaan 200B, B-3001 Leuven, Belgium
[b] University of Antwerp, Faculty of Social Sciences, Middelheimlaan 1, B-2020 Antwerp, Belgium
[c] Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium
[d] Antwerp Maritime Academy, Noordkasteel Oost 6, B-2030 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 12 October 2016
Accepted 7 February 2017

Keywords:
Cognitive distances
Barycenters
Similarity matrices
Similarity-adapted publication vectors
Weighted cosine similarity
Bootstrapping
Research expertise

ABSTRACT

We study the problem of determining the cognitive distance between the publication portfolios of two units. In this article we provide a systematic overview of five different methods (a benchmark Euclidean distance approach, distance between barycenters in two and in three dimensions, distance between similarity-adapted publication vectors, and weighted cosine similarity) to determine cognitive distances using publication records. We present a theoretical comparison as well as a small empirical case study. Results of this case study are not conclusive, but we have, mainly on logical grounds, a small preference for the method based on similarity-adapted publication vectors.

1. Introduction

In this article, we address the research question: How can we obtain, using publication data, a meaningful distance or proximity measure which represents the cognitive distance or proximity between two units? This is in fact a rephrased version of a problem we discussed earlier (Rahman, Guns, Rousseau, & Engels, 2015), where we asked 'How can we quantify the overlap of expertise between two entities, e.g., a research group and a panel, using publication data?'.

In our investigation, entities or units are either experts, panels of experts, or research groups. One can easily think of other informetric contexts in which the calculation of cognitive distances is relevant, e.g. the search of suitable peer reviewers for the evaluation of journal submissions, for grant applications or in hiring/promotion decisions, the exploration of potential collaborations, and distinguishing between different 'modalities' of interdisciplinarity (Molas-Gallart, Rafols, & Tang, 2014). Rafols, Porter, and Leydesdorff (2010) suggest several possible uses of overlay maps in research management that depend on cognitive distance, such as benchmarking and comparing the research profiles of organizations, and exploring complementarities and possible collaborations. In this regard they point out that "successful collaborations tend to occur in a middle range of cognitive distance, whereupon collaborators can succeed at exchanging or sharing complementary knowledge or capabilities, while still being able to understand and coordinate with one another." Our quantitative approaches are complementary to visual approaches like overlay maps (Leydesdorff & Rafols, 2009; Leydesdorff, Carley, & Rafols, 2013; Rafols et al., 2010).

In this contribution, we focus on theoretical-logical aspects of the calculation of cognitive distance. As an application and to keep a clear link with our previous work we re-use the data and framework of (Rahman et al., 2015). In that article,

* Corresponding author at: KU Leuven, Dept. of Mathematics, Celestijnenlaan 200B, B-3001 Leuven, Belgium.
E-mail address: ronald.rousseau@kuleuven.be (R. Rousseau).

http://dx.doi.org/10.1016/j.joi.2017.03.001
1751-1577/

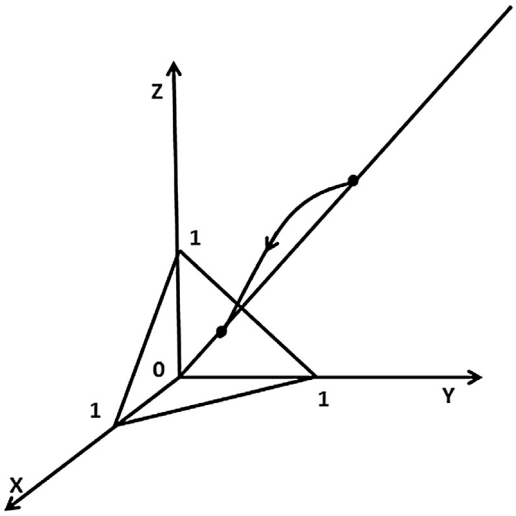

Fig. 1. Normalization, leading to a scale invariant approach.

publications were assigned to Web of Science Subject Categories, in short WoS SCs. We admit that the use of WoS SCs was a convenience approach, which has meanwhile been refined by applying a journal level approach (Rahman, Guns, Leydesdorff, & Engels, 2016). More precisely, instead of assigning publications to WoS SCs, publications were assigned to the journal in which they were published.

2. Measuring cognitive distance

Nooteboom (2000) defines cognitive distance as "a difference in cognitive function". He explains this as follows: "This can be a difference in domain, range, or mapping. People could have a shared domain but a difference of mapping: two people can make sense of the same phenomena, but do so differently". Hence, the term 'cognitive distance' refers to the way in which two persons, and by extension, two organizations or groups of persons, are different, not only in terms of knowledge, but also in the way they perceive and interpret external phenomena. Like many other notions used in the social sciences – the notions of impact, inequality, visibility come to mind –, the notion of cognitive distance must be operationalized. This operationalization can be done in many different ways.

Here, as in (Rahman et al., 2015; Rahman, Guns, Leydesdorff et al., 2016; Wang & Sandström, 2015) we consider the publication portfolio of the involved researchers to reflect the position of the unit in cognitive space and, hence, to determine cognitive distance. Expressed in general terms we measure cognitive distance between units based on how often they published in the same or similar journals. Similarity between journals can be measured in a direct way or via the WoS SCs to which they belong. Details are provided further on. In the case study presented in this paper, similarity is determined by the citation-based similarity of WoS SCs to which journals belong. The research groups are either research groups in physics or in chemistry working at the University of Antwerp, Belgium. For details we refer to Rahman et al. (2015).

One can think of other informetric ways to determine cognitive distance between scientists. Wang and Sandström (2015) for example use bibliographic coupling and topic modelling to determine cognitive distance between publication portfolios. Besides using publication portfolios, one could also measure cognitive distance between patent portfolios, in terms of conference participation, in terms of diplomas, and so on. Moreover, cognitive distance is relevant in many other social and political contexts as well, e.g. when hiring employees, when comparing the programs of political parties, or to understand cultural differences.

We recall (Rahman, Guns, Rousseau, & Engels, 2016) that in order to obtain meaningful cognitive distances these values must be scale-invariant. This means that the distance between points P and Q must be the same as the distance between the points P and cQ, where c is a strictly positive number. Indeed: the total output of a research group can be several orders of magnitude larger than that of one expert. For the applications we have in mind this difference must not play a role in determining cognitive distances. Scale-invariance can be obtained through normalization as illustrated (for 3 dimensions) in Fig. 1. All points situated on the straight line through the origin are represented by the same point in the plane with equation $x + y + z = 1$.

This is so-called L_1-normalization: by dividing each coordinate by the sum of all coordinates one obtains a new array for which the sum of all coordinates is one (taking into account that no coordinate is negative). One could equally well divide by an array's Euclidean length (so- called L_2-normalization) but as we do not see an advantage for any of the two approaches we applied L_1-normalization as is done in diversity studies.