Measuring cognitive distance between publication portfolios

Reference:

Rousseau Ronald, Guns Raf, Jakaria Rahman Abu Imran Mohammad, Engels Tim.- Measuring cognitive distance between publication portfolios
Journal of informetrics - ISSN 1751-1577 - 11:2(2017), p. 583-594
Full text (Publishers DOI): http://dx.doi.org/doi:10.1016/J.JOI.2017.03.001
To cite this reference: http://hdl.handle.net/10067/1423650151162165141

Highlights:

- Cognitive distance between publication portfolios of scientific units is determined.
- The importance of scale invariance in determining cognitive distance is explained.
- Two similarity-based methods in N dimensions are proposed.
- Low dimensional and N-dimensional methods are compared in a small case study.

Measuring cognitive distance between publication portfolios

Ronald Rousseau [a,b], Raf Guns [c] , A.I.M. Jakaria Rahman [c], and 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

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.

Keywords: cognitive distances; barycenters; similarity matrices; similarity-adapted publication vectors; weighted cosine similarity; bootstrapping; research expertise.

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 et al., 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; Rafols, Porter, & Leydesdorff, 2010; Leydesdorff, Carley, & Rafols, 2013).

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, 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, 2016a). 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, 2016a; 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 & 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 et al., 2016b) 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.

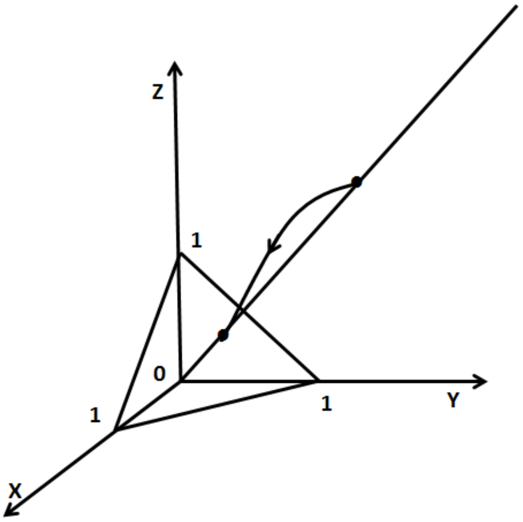

Fig. 1. Normalization, leading to a scale invariant approach

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.

3. Representing researchers' publication profiles

Researchers' publication profiles and their (dis)similarities will be represented in five different ways: a benchmark, two methods using barycenters (one in two and one in three dimensions), a fourth method using similarity-adapted publication vectors (in short: SAPVs) and a fifth one using weighted cosine similarities (in short: WCS). The benchmark and the last two values are applied in N dimensions, where N denotes the total number of SCs. In each case we start from a publication vector M = $(m_j)_j$, with j=1,…,N. The coordinates of this vector are the number of publications belonging to category j. Each panel member and each research group has a corresponding publication vector. In the applications only publications during a specific publication window and included in the Web of Science are considered, but the approach is independent of the used publication window or data source.

Throughout the remainder of the text, we will work with the example of determining cognitive distances between expert panels and their members on the one hand and research groups on the other (in the context of research evaluation). However, we stress the fact that the methods presented are more general and can also be applied in other contexts and for other purposes.

3.1 The benchmark

Scientists and research groups are represented as N-dimensional publications vectors. As a start (benchmark) we just calculate the Euclidean distance between the L_1-normalized arrays of each panel member and each research group. Recall that the Euclidean distance between two vectors a = $(a_n)_{n=1,\ldots,k}$ and b = $(b_n)_{n=1,\ldots k}$ in $\mathbf{R}^k$, for any strictly positive integer k, is given as:

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + \cdots + (a_k - b_k)^2} \qquad (1)$$

In this paper we will use formula (1) for k = 2, k=3 and k = N.

3.2 Second and third method: barycenters

To answer our research question the second method uses a 2-dimensional base map. We note that this base map can be considered to be universal and hence has nothing to do with the concrete data at hand. Each SC has a place on this map, characterized by corresponding coordinates, denoted as $(L_{j,1}, L_{j,2})$, $j = 1, \ldots, N$. In the application that will follow, the 2-dimensional barycenter approach is based on a VOS (visualization of similarities) (Van Eck & Waltman, 2007) map (taken from Leydesdorff et al., 2013), but other 2-dimensional mappings are feasible. Now for each panel member and for each research group a barycenter derived from their publication profiles is calculated. Coordinates of these barycenters (in 2 dimensions) are given as

$$C_1 = \frac{\sum_{j=1}^{N} m_j L_{j,1}}{T} \ ; \ C_2 = \frac{\sum_{j=1}^{N} m_j L_{j,2}}{T} \qquad (2)$$

where m_j is the number of publications of the unit under investigation (panel member, research group) belonging to category j; this category j has coordinates $(L_{j,1}, L_{j,2})$ in the base map; $T = \sum_{j=1}^{N} m_j$ is the total number of publications of the unit under investigation. We note that in the case study performed further on, T is larger than the total number of publications as full counting of WoS SCs has been used, which means that publications belonging to multiple WoS SCs are counted multiple times. Euclidean distances between units, as represented by their barycenters, can be calculated leading to quantitative results answering our research question.

The barycenter method explained above and in particular formulae (2) satisfy the scale-invariance requirement as multiplying all m_js with the same strictly positive factor leads to the same barycenter.

Although it is convenient to perform visualization and to determine cognitive distance in the plane, there is no theoretical reason to perform these acts in two dimensions. Likewise, there are no strong reasons to do both in the same dimension. The barycenter method can, at least in theory, be applied in any strictly positive dimension smaller than or equal to N. Not wanting to go too deep into this largely theoretical issue we will just check how results for our case studies compare in two and three dimensions, leading to the third method, namely the use of barycenters in three dimensions.

For three dimensions, we again use the VOS algorithm, but now resulting in a three dimensional base map. This map was based on the network in http://www.leydesdorff.net/overlaytoolkit/map10.paj and obtained using Pajek, which implements the VOS algorithm both in 2 and 3 dimensions.

Again each SC has a place on this map, characterized by corresponding coordinates, denoted as $(L_{j,1}, L_{j,2}, L_{j,3})$, $j = 1, \ldots, N$, and for each panel member and for each research group a barycenter derived from their publication profiles is calculated. Coordinates in 3 dimensions are given as

$$C_1 = \frac{\sum_{j=1}^{N} m_j L_{j,1}}{T} \; ; \; C_2 = \frac{\sum_{j=1}^{N} m_j L_{j,2}}{T} ; \; C_3 = \frac{\sum_{j=1}^{N} m_j L_{j,3}}{T} \qquad (3)$$

The meaning of the symbols T and m_j in formulae (3) is the same as in formulae (2).

3.3 Fourth method: Similarity-adapted publication vectors (SAPV)

In Rahman et al. (2015) we used another quantitative approach (mistakenly also referred to as a barycenter method, but corrected in Rahman et al., 2016b), this time in N dimensions. In that approach, we used a matrix of similarity values between the WoS SCs as made available by Rafols, Porter & Leydesdorff (2010) at http://www.leydesdorff.net/overlaytoolkit/map10.paj. These authors created a matrix of citing to cited SCs based on the Science Citation Index (SCI) and Social Sciences Citation Index

(SSCI), which was cosine-normalized in the citing direction. The result is a symmetric N×N similarity matrix (here, N=224) which we denote by S = $(s_{ij})_{ij}$.

The multiplication $S * M$, i.e. applying the linear map with matrix representation S to the publication vector M leads to a new vector which we termed a similarity-adapted publication vector, SAPV in short. If we ignore similarity then S is the identity matrix and publication columns stay unchanged. We consider the SAPV method to be quite interesting as it provides a solution to the problem that WoS SCs overlap and are sometimes poorly defined, the SC *Information Science & Library Science* being a well-known example.

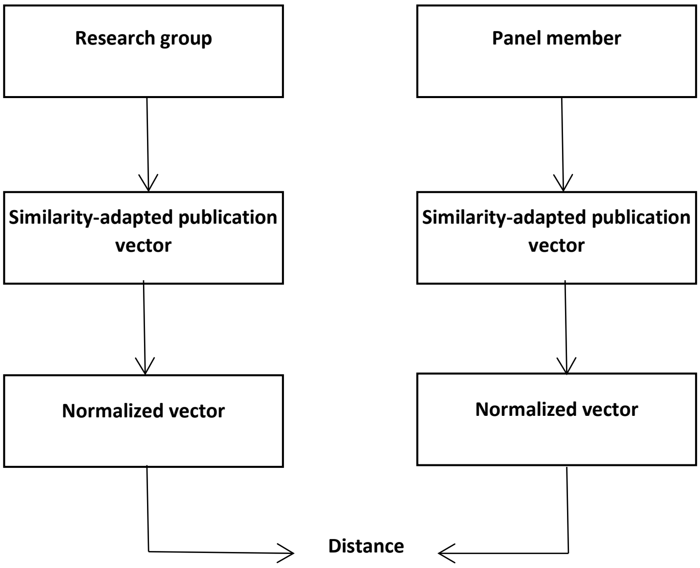
Research group
Panel member
Similarity-adapted publication vector
Similarity-adapted publication vector
Normalized vector
Normalized vector
Distance

Fig. 2. Workflow for determining distances between SAPVs

In Rahman et al. (2015), we determined the distance for SAPVs (although they were referred to as N-dimensional barycenters). As these vectors were not normalized the obtained results were not scale-invariant. It suffices though, to follow the workflow shown in Fig.2.

Hence, a normalized SAPV of a research group or panel member is determined as the vector $C = (C_1, C_2, \ldots, C_N)$, with coordinates C_k determined as:

$$C_k = \frac{\sum_{j=1}^{N} s_{kj} m_j}{\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} m_j} = \frac{(S * M)_k}{\|S * M\|_1} \tag{4}$$

where s_{kj} denotes the similarity value between the k-th and the j-th WoS SC, and m_j is the number of publications in WoS SC j of the research group or the panel member. The numerator of Equation (4) is equal to the k-th element of $S * M$, the multiplication of the similarity matrix S and the column matrix of publications $M = \left(m_j\right)_j$. The denominator is the L_1-norm of the unnormalized vector. We observe that the L_1-norm of the normalized vector C is indeed equal to 1.

3.4 Fifth method: Weighted cosine similarity

Finally, we mention a weighted cosine similarity method (in short: WCS). The WCS between panel member (PM) k and research group m, according to Zhou et al. (2012) is:

$$\frac{\sum_{i=1}^{N} M_i^k \left(\sum_{j=1}^{N} R_j^m s_{ji}\right)}{\sqrt{\left(\sum_{i=1}^{N} M_i^k \left(\sum_{j=1}^{N} M_j^k s_{ji}\right)\right) . \left(\sum_{i=1}^{N} R_i^m \left(\sum_{j=1}^{N} R_j^m s_{ji}\right)\right)}}$$

$$= \frac{\left(M^k\right)^t * S * R^m}{\sqrt{\left(M^k\right)^t * S * M^k} . \sqrt{\left(R^m\right)^t * S * R^m}} \qquad (5)$$

The numerator is nothing but the matrix multiplication: $\left(M^k\right)^t * S * R^m$, where ^{t} denotes matrix transposition, S is the similarity matrix, M^k denotes the column matrix of publications of panel member k and R^m denotes the column matrix of publications of research group m. Similarly, the two products under the square root in the denominator are: $\left(M^k\right)^t * S * M^k$ and $\left(R^m\right)^t * S * R^m$. The result is the WCS value between panel member k and research group m. Formula (5) is clearly scale-invariant: multiplying M^k or R^m with a fixed constant does not change the result. Note that if S is the identity matrix (similarity is not taken into account), formula (5) reduces to regular cosine similarity. A similarity or proximity can be considered as the opposite of a distance: the higher the similarity the better the match – the closer the distance – between a panel member and a research group. This value too is calculated for each panel member and each research group. We note that this fifth method may lead to mathematical problems when applied in general vector spaces, but that these do not occur in the particular framework used in this article (in mathematical terms: we work in the positive cone $(\mathbf{R}^+)^N$, where **R** denotes the real numbers). Details are provided in Appendix B.

4. Results

As in our previous paper (Rahman et al., 2015), we calculate the cognitive distance between different research groups and panel members. Group names have been standardized using the first four letters of the corresponding department, for example, CHEM-A for chemistry research group A, PHYS-B for physics research group B. The panel member names are standardized as PM1, PM2 etc. , but refer to different colleagues depending on the panel in question.

Yet, another problem must be solved before we can really state that one panel member is closer to a research group than another. Small differences in distance or similarity bear little meaning and should not be used to make claims that, for instance, one panel member is a 'better' choice than another. We therefore use a bootstrapping method (Efron & Tibshirani, 1998) leading to 95% confidence intervals for distances and similarities. Details of the bootstrapping method we applied are explained in (Rahman et al., 2016a). A more detailed explanation can be found online (http://nbviewer.jupyter.org/gist/rafguns/6fa3460677741e356538337003692389 and http://nbviewer.jupyter.org/gist/rafguns/faff8dc090b67a783b85d488f88952ba). If the confidence interval of the panel member who is closest to a given research group overlaps with that of the panel member who ranks second (and maybe even with the panel members ranking third or fourth) we say that there is no (statistical) difference in cognitive distance. In order to facilitate a comparison between the five methods, results for the barycenter method in 2D, although already published in (Rahman et al., 2015) are included in Appendix A. These results are recalculated (leading to small differences) and information about the calculated confidence intervals is added. Hence we begin the presentation of shortest distances between panel members and research groups with the benchmark case (Tables 1 and 2), followed by the 3D barycenter case (Tables 3 and 4), the SAPV method (Tables 5 and 6) and finally the WCS method (Tables 7 and 8). For each research group we determine the panel member at the shortest distance. The number in the row corresponding to this panel member is indicated in bold and underlined. Distances whose confidence intervals overlap with that of the shortest distance are in bold (same column). We will use the same way of showing results for all the tables.

Table 1: Euclidean distances in N dimensions between normalized publication arrays of research groups and panel members of the Chemistry department.

	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
PM1	0.607	0.697	0.646	**0.459**	0.627	0.743	0.656	0.652	0.674	0.646	0.607	0.667
PM2	0.507	0.565	**0.402**	0.588	**0.300**	**0.240**	0.316	**0.377**	**0.269**	**0.356**	**0.445**	0.531
PM3	0.540	0.573	**0.381**	0.598	**0.279**	**0.405**	**0.288**	**0.257**	**0.242**	**0.350**	**0.468**	0.561
PM4	0.542	0.601	**0.441**	0.608	**0.331**	**0.340**	**0.217**	**0.372**	**0.336**	**0.360**	**0.464**	0.556
PM5	**0.180**	**0.157**	0.482	0.604	0.500	0.659	0.547	0.499	0.515	0.520	**0.500**	**0.368**
PM6	0.715	0.762	0.726	**0.255**	0.693	0.809	0.738	0.731	0.749	0.729	0.693	0.745
PM7	0.684	0.770	0.741	0.758	0.732	0.825	0.746	0.744	0.761	0.741	0.713	0.739

Table 2: Euclidean distances in N dimensions between normalized publication arrays of research groups and panel members of the Physics department.

	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
PM1	0.716	0.793	0.699	**0.114**	0.519	0.786	0.730	0.806	0.662
PM2	0.953	0.466	0.788	1.048	0.801	1.008	0.956	0.457	0.899
PM3	0.639	0.741	0.654	0.819	0.634	0.759	0.701	0.705	0.621
PM4	0.600	0.663	0.476	0.738	0.481	**0.663**	**0.278**	0.662	0.523
PM5	**0.510**	0.376	**0.171**	0.667	**0.296**	**0.559**	0.494	0.410	**0.387**
PM6	0.618	**0.224**	0.388	0.736	**0.379**	**0.576**	0.568	**0.241**	0.531

Table 3: Euclidean distances between barycenters of research groups and panel members of the Chemistry department using the 3-dimensional WoS SCs map.

	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
PM1	**0.037**	**0.032**	**0.043**	**0.033**	0.064	**0.059**	**0.018**	**0.006**	**0.014**	**0.043**	0.103	**0.033**
PM2	0.110	0.108	0.114	**0.045**	**0.017**	**0.022**	0.062	0.075	0.063	**0.060**	**0.035**	0.110
PM3	0.051	0.047	0.056	**0.019**	**0.050**	**0.044**	**0.006**	**0.015**	**0.007**	**0.040**	0.090	0.048
PM4	0.069	0.063	0.074	**0.012**	**0.037**	**0.032**	**0.013**	**0.033**	**0.023**	**0.050**	0.084	0.064
PM5	**0.030**	**0.027**	**0.034**	**0.040**	0.069	**0.064**	0.028	**0.007**	**0.019**	**0.038**	0.103	**0.029**
PM6	0.057	0.052	0.062	**0.013**	**0.044**	**0.038**	**0.007**	**0.021**	**0.010**	**0.039**	0.085	0.054
PM7	**0.023**	**0.016**	**0.028**	**0.049**	0.080	**0.075**	0.034	**0.018**	**0.030**	**0.053**	0.117	**0.017**

Table 4: Euclidean distances between barycenters of research groups and panel members of the Physics department using the 3-dimensional WoS SCs map.

	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
PM1	0.453	0.054	0.084	**0.011**	0.067	0.064	0.162	0.048	0.257
PM2	**0.408**	**0.007**	0.032	0.043	**0.016**	0.044	**0.112**	**0.008**	**0.211**
PM3	**0.392**	0.024	0.037	0.050	0.026	**0.013**	**0.105**	0.026	**0.196**
PM4	**0.361**	0.049	**0.018**	0.091	0.035	0.061	**0.062**	0.054	**0.163**
PM5	**0.393**	0.014	**0.017**	0.056	**0.003**	0.041	**0.096**	0.019	**0.195**
PM6	**0.409**	**0.006**	0.034	0.040	**0.017**	0.041	**0.113**	**0.004**	**0.211**

The recalculation with respect to what was obtained in N dimensions in (Rahman et al.. 2015). leads to the distances reported in Tables 5 and 6 for the cases of chemistry and physics.

Table 5. Euclidean distances between SAPVs of research groups and panel members of the Chemistry department using the similarity matrix of WoS SCs.

	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
PM 1	0.081	0.079	0.108	**0.061**	0.124	0.119	0.116	0.104	0.093	0.129	0.141	0.085
PM 2	0.082	0.074	0.079	**0.054**	**0.036**	**0.032**	0.055	**0.046**	**0.036**	**0.075**	**0.071**	0.070
PM 3	0.082	0.074	0.080	0.066	**0.057**	**0.058**	**0.040**	**0.040**	**0.042**	**0.075**	**0.086**	0.073
PM 4	0.106	0.099	0.104	0.085	0.064	0.070	**0.027**	0.063	0.071	**0.085**	0.094	0.091
PM 5	**0.015**	**0.013**	**0.034**	0.074	0.100	0.102	0.077	**0.053**	**0.050**	**0.082**	0.096	**0.024**
PM 6	0.093	0.087	0.111	**0.025**	0.085	0.080	0.096	0.090	0.080	0.113	0.116	0.088
PM 7	0.068	0.068	0.097	0.072	0.128	0.125	0.113	0.099	0.089	0.125	0.140	0.075

Table 6. Euclidean distances between SAPVs of research groups and panel members of the Physics department using the similarity matrix of WoS SCs.

	PHYS-A	PHYS- B	PHYS-C	PHYS- D	PHYS-E	PHYS- F	PHYS- G	PHYS- H	PHYS-I
PM 1	0.376	0.358	0.373	**0.098**	0.328	0.301	0.371	0.358	0.367
PM 2	0.172	0.019	0.038	0.272	0.054	0.127	0.115	**0.019**	0.133
PM 3	**0.156**	0.065	0.080	0.256	0.069	**0.100**	0.116	0.063	**0.111**
PM 4	**0.144**	0.060	0.039	0.271	0.051	0.129	**0.066**	0.063	**0.103**
PM 5	**0.157**	0.023	**0.016**	0.271	**0.044**	0.125	**0.095**	0.027	**0.115**
PM 6	**0.165**	**0.012**	0.035	0.258	**0.037**	**0.111**	**0.106**	**0.015**	0.125

Tables 5 and 6 are analogues of respectively, Tables 1 and 3 of the supplementary online material (part 2) of Rahman et al. (2015). This ends the presentation of the results obtained by the barycenter and SAPV method. Tables 7 and 8 contain the WCS results, where we recall that this is a similarity approach (not a distance based one) and hence largest values refer to entities that are closest.

Table 7. WCS values of research groups and panel members of the Chemistry department using the similarity matrix of WoS SCs.

	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
PM1	0.709	0.667	0.445	**0.922**	0.469	0.449	0.395	0.440	0.507	0.323	0.273	0.661
PM2	0.670	0.713	0.726	0.675	**0.914**	**0.945**	0.837	**0.847**	**0.947**	**0.703**	**0.527**	0.713
PM3	0.594	0.655	0.673	0.569	**0.839**	0.831	0.866	**0.880**	**0.894**	**0.711**	**0.403**	0.604
PM4	0.459	0.517	0.504	0.484	0.781	0.777	**0.951**	0.758	0.769	**0.626**	0.315	0.549
PM5	**0.983**	**0.990**	**0.842**	0.669	0.581	0.475	0.614	0.747	0.758	**0.573**	**0.512**	**0.933**
PM6	0.613	0.600	0.377	**0.973**	0.545	0.519	0.391	0.410	0.484	0.294	0.280	0.603
PM7	0.758	0.713	0.503	0.850	0.460	0.439	0.440	0.494	0.550	0.373	0.290	0.700

Table 8. WCS values of research groups and panel members of the Physics department using the similarity matrix of WoS SCs.

	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
PM1	0.030	0.155	0.043	**0.996**	0.561	0.508	0.028	0.154	0.052
PM2	**0.151**	**0.982**	0.920	0.127	0.806	0.513	0.543	**0.977**	**0.497**
PM3	**0.220**	0.714	0.625	0.211	0.668	0.526	0.440	0.762	**0.544**
PM4	**0.182**	0.729	0.829	0.129	0.757	0.436	**0.895**	0.741	**0.479**
PM5	**0.182**	0.965	**0.986**	0.158	**0.852**	0.475	0.656	0.957	**0.567**
PM6	**0.164**	**0.989**	0.930	0.272	**0.903**	**0.643**	0.631	**0.985**	**0.516**

5. Correlations

We calculated the Pearson correlation coefficient (r) and the Spearman rank correlation coefficient (ρ) between distances/similarities based on the five methods, see Tables 9 and 10. These calculations are based on all distances between research groups and individual panel members. For calculations involving WCS we show absolute values, as distances and similarities are each other's opposites, and hence correlations are negative.

Table 9. Chemistry: Pearson and Spearman correlations for all cognitive distances between research groups and individual panel members.

Pearson Spearman	Benchmark	Barycenter 2D	Barycenter 3D	SAPV	WCS
Benchmark	1.00	0.38	0.09	0.72	0.72
Barycenter 2D	0.34	1.00	0.81	0.75	0.64
Barycenter 3D	0.06	0.82	1.00	0.42	0.31
SAPV	0.67	0.72	0.42	1.00	0.92
WCS	0.67	0.62	0.30	0.92	1.00

In Tables 9 and 10, the upper triangle refers to Pearson correlations while the lower triangle refers to Spearman correlations. Clearly SAPV and WCS results in Tables 9 and 10 are highly correlated.

Table 10. Physics: Pearson and Spearman correlation for all cognitive distances between research groups and individual panel members.

Pearson Spearman	Benchmark	Barycenter 2D	Barycenter 3D	SAPV	WCS
Benchmark	1.00	0.12 (0.34)	0.22 (0.27)	0.50 (0.56)	0.63 (0.54)
Barycenter 2D	0.37(0.48)	1.00	0.99 (0.99)	0.29 (0.87)	0.60 (0.89)
Barycenter 3D	0.34(0.38)	0.94 (0.96)	1.00	0.35 (0.81)	0.61 (0.85)
SAPV	0.60(0.56)	0.64 (0.94)	0.71 (0.86)	1.00	0.86 (0.97)
WCS	0.65(0.58)	0.71 (0.91)	0.74 (0.83)	0.94 (0.97)	1.00

Values between brackets in Table 10 are correlations calculated after removal of PHYS-D and PM1; an explanation for doing this is provided further. Correlations for the benchmark case (ignoring all similarities) and the other approaches are moderate at best. Not surprisingly, the two N-dimensional approaches (SAPV and WCS) are more correlated with the benchmark case than the lower dimensional ones. Correlations between the 2D and the 3D approach are high in all cases. This illustrates that the number of dimensions chosen has only limited influence on the results based on barycenters. Most other correlations can be described as moderate to high. For chemistry we note, however, that the correlations between barycenter 3D on the one hand, and SAPV and WCS on the other, are lower than expected. Moreover, these values are lower than for the 2D case. We were not able to find an explanation for this unexpected difference. We further note a low correlation between SAPV and the barycenter methods in physics. For this case, however, we found a convincing explanation. Fig. 3 illustrates what happened.

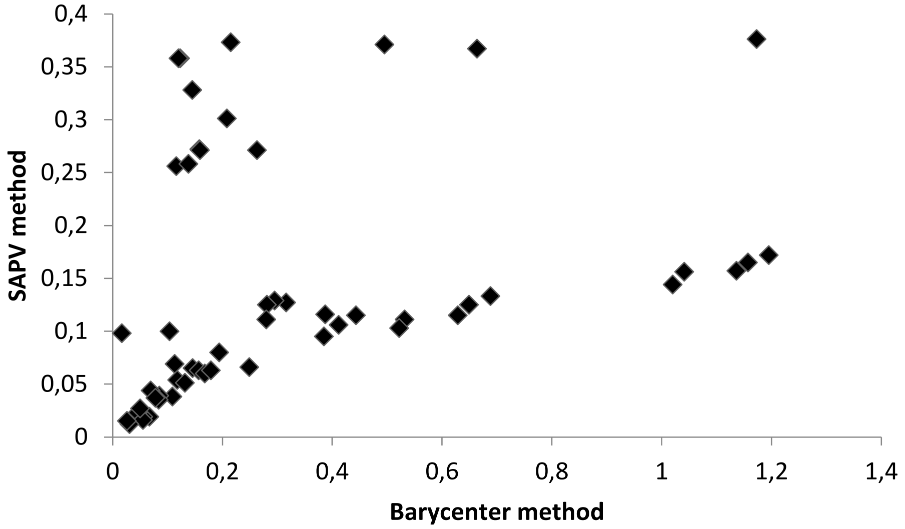

Fig.3. Scatter plot of the cognitive distances between research groups and individual panel members for the 2D barycenter and SAPV methods in the physics department.

This low Pearson correlation is due to the 13 points (including two times two points that overlap and cannot be seen) in the upper half of Fig.3. All these points correspond to distances involving research group PHYS-D and PM1 (but not both). This group and this panel member are active in the same field (*Physics, Particles & Fields*) and have different scientific interests than the other groups or panel members: 99.1% of PM1's publications belong to the SC *Physics, Particles & Fields,* while for PHYS-D, this SC covers 83.6% of its publications. Moreover, their publications cover only four (117 publications) and seven (269 publications) WoS SCs respectively while other panel members cover 12 to 26 WoS SCs, and other research groups 26 to 50 SCs. Fig. 4 presents the same data as Fig. 3, but leaves out distances involving PHYS-D and PM1. In this case, all correlations increase considerably.

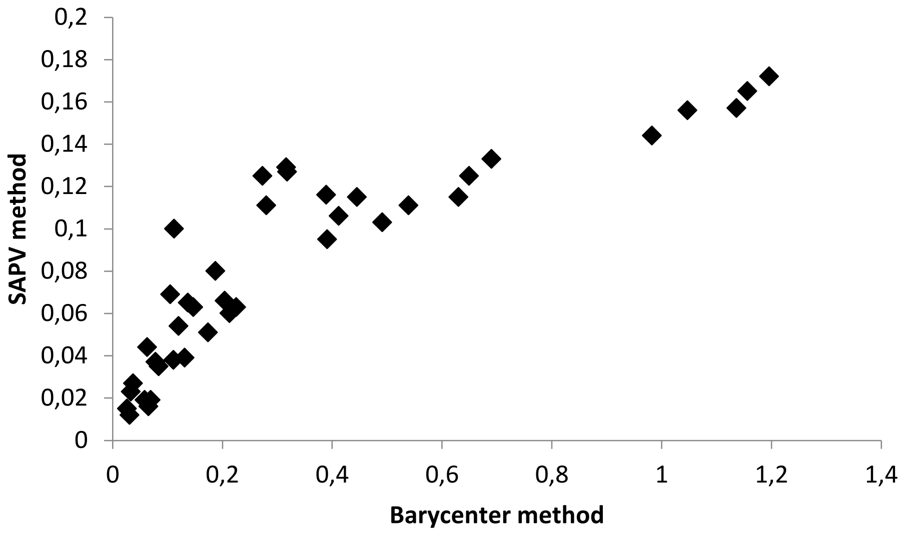

Fig. 4. Scatter plot of the cognitive distances between research groups and individual panel members obtained by the 2D barycenter and SAPV methods in the physics department excluding PHYS-D and PM1.

A more detailed comparison between the five methods follows in the next section.

6. Comparison between the five methods

A comparison would be easy if a gold standard existed. Clearly, it does not, but we used the labour division decided upon by the panel chair as a proxy. Prior to a site visit, see (Engels, Goos, Dexters & Spruyt, 2013) for details, the panel chair appointed a main assessor for each of the research groups to be evaluated. This main assessor studied the profile and performance of the research group in detail, asked the majority of questions during the site visit and wrote the (first draft of) the final assessment of the research group. Assuming that panel chairs assigned the best suited panel member as main assessor, a perfect method would always rank this main assessor first. However, remember that neither have panel members and research groups ever collaborated nor do they belong to the same university, so this assumption does not necessarily always hold in practice.

Tables 11 and 12 show the research groups, the corresponding main assessor, and the panel members with the closest distance (for the five methods). The first one in each cell is the panel member closest to the corresponding research group; the others are panel members whose distances are statistically not different from this shortest distance. We have to point out two extra problems for chemistry. The first is that although PM7 was indicated as the main assessor for CHEM-C, PM3 thought himself closest to this research group. The second problem was that PM3 was indicated as main assessor of CHEM-F but he himself doubted if he could assess this group as an expert.

Table 11. Chemistry: Top ranked panel members according to five methods

Research group	Main assessor	Benchmark	Barycenter 2D	Barycenter 3D	SAPVs	WCS
CHEM-A	PM6	PM5	PM5-PM7	PM7- PM5-PM1	PM5	PM5
CHEM-B	PM5	PM5	PM5-PM7-PM1	PM7- PM5-PM1	PM5	PM5
CHEM-C	PM7/PM3	PM3-PM2-PM4	PM5	PM7- PM5-PM1	PM5	PM5
CHEM-D	PM2	PM6- PM1	PM6-PM4-PM3-PM2-PM1	PM4- PM6-PM3- PM1-PM5- PM2-PM7	PM6-PM2-PM1	PM6-PM1
CHEM-E	PM2	PM3-PM2-PM4	PM2-PM4-PM6	PM2- PM4-PM6- PM3	PM2-PM3	PM2-PM3
CHEM-F	PM3	PM2-PM4-PM3	PM2-PM6-PM4-PM3	PM2- PM4-PM6- PM3-PM1- PM5-PM7	PM2-PM3	PM2
CHEM-G	PM3	PM4-PM3	PM3-PM4	PM3- PM6-PM4- PM1	PM4-PM3	PM4
CHEM-H	PM5	PM3-PM4-PM2	PM4-PM3-PM5	PM1- PM5-PM3- PM7-PM6- PM4	PM3-PM2-PM5	PM3-PM2-PM4
CHEM-I	PM4	PM3-PM2-PM4	PM3-PM5	PM3- PM6-PM1- PM5-PM4- PM7	PM2-PM3-PM5	PM2-PM3
CHEM-J	PM4	PM3-PM2-PM4	PM4-PM2-PM3-PM5	PM5- PM6-PM3- PM1-PM4- PM7-PM2	PM3-PM2-PM5-PM4	PM3-PM2-PM4-PM5
CHEM-K	PM6	PM2-PM4-PM3-PM5	PM2-PM4	PM2	PM2-PM3	PM2- PM5-PM3-
CHEM-L	PM1	PM5	PM5-PM7-PM1	PM7- PM5-PM1	PM5	PM5
score		**7/12 (2/12)**	**8/12 (4/12)**	**10/12 (3/12)**	**7/12 (2/12)**	**3/12 (2/12)**

Table 12. Physics: Top ranked panel members according to five methods

Research group	Main assessor	Benchmark	Barycenter 2D	Barycenter 3D	SAPVs	WCS
PHYS-A	PM3	PM5	PM4-PM3-PM5-PM6	PM4- PM3-PM5- PM2-PM6	PM4-PM3-PM5-PM6	PM3-PM5-PM4-PM6-PM2
PHYS-B	PM2	PM6	PM6-PM5	PM6- PM2	PM6	PM6-PM2
PHYS-C	PM5	PM5	PM5-PM4	PM5- PM4	PM5	PM5
PHYS-D	PM1	PM1	PM1	PM1	PM1	PM1
PHYS-E	PM4	PM5-PM6	PM5-PM6	PM5- PM2-PM6	PM6-PM5	PM6-PM5
PHYS-F	PM1	PM5-PM6-PM4	PM3-PM1	PM3	PM3-PM6	PM6
PHYS-G	PM4	PM4	PM4-PM3-PM5-PM6	PM4- PM5-PM3- PM2-PM6	PM4-PM5-PM6	PM4
PHYS-H	PM6	PM6	PM6-PM5	PM6- PM2	PM6-PM2	PM6-PM2
PHYS-I	PM3	PM5	PM4-PM3-PM5	PM4- PM5-PM3- PM2-PM6	PM4-PM3-PM5	PM5-PM3-PM6-PM2-PM4
Score:		**4/9 (4/9)**	**7/9 (4/9)**	**7/9 (4/9)**	**6/9 (4/9)**	**7/9 (4/9)**

In order to gauge the overall correspondence between the methods used by us and the chosen main assessor we count how often the method found the chosen assessor, once taking only the nearest panel member into account (sum between brackets) and once taking into account that some panel members could on statistical grounds (overlapping confidence intervals) not be separated, an approach which is assumed to be the better one. In most cases, WCS for chemistry being the exception, the benchmark case scores poorest, proving the benefit of taking similarities into account. For chemistry, the barycenter methods score slightly better than SAPV and WCS, while for physics there is hardly any difference between the four (even five) methods. Especially in the case of chemistry, we have several cases where most confidence intervals overlap. The barycenter method in 3D clearly has very low discriminatory power leading to cases where all confidence intervals overlap (CHEM-F and CHEM-J). In these cases the 3D barycenter cannot distinguish between panel members.

We see that for some research groups the five methods and the chosen assessor coincide (taking confidence intervals into account). This perfect result was attained for CHEM-B, CHEM-E, CHEM-J, PHYS-C, PHYS-D, PHYS-G and PHYS-H; while only the benchmark case missed PHYS-A and PHYS-L. Hence, this is the case for 3 of the 12 chemistry groups and for 4 (or 6) of the 9 physics groups. The smaller number of perfect results in chemistry is largely due to the WCS method. For some other groups no method leads to the chosen assessor. This is the case for CHEM-A, CHEM-K, and PHYS-E. Mainly due to the overlapping confidence intervals the barycenter method in 3D is the only one which included the main assessor for CHEM-C and CHEM-I (and the benchmark has PM3 as closest to CHEM-C). In all these negative cases, the results obtained by the five methods largely agree. A possible explanation for this surprising result might simply be that the panel chair included other factors - than pure scientific affinity - in the decision to assign a panel member to a research group. In the case of chemistry where the suggested labour division was partly contested by PM3, PM5 is identified as the closest to CHEM-C. A possible explanation for this specific case could be that PM5 was already the main assessor for two groups so that, for purely practical reasons, PM3 became the main assessor of CHEM-C.

Considering now the individual panel members we see that some are close to several research groups, while others are not close to any. For chemistry we see that, according to the 2D barycenter method PM4 and PM5 are close to seven research groups, while PM2, PM3 and PM5 are closest to seven research groups according to the SAPV method. PM5 is closest to six research groups according to the WCS method. Clearly, PM5 was an essential panel member. According to the two barycenter-based methods all chemistry panel members are closest to at least three groups, but according to the SAPV and the WCS method PM7 is closest to none.

For physics PM5 and PM6 are closest to at least four research groups, and this for the four similarity-based methods. PM2 is closest to none according to the 2D barycenter method, but closest to four groups according to the WCS method. We observe the special role of PM1 in physics who is the only one closest to PHYS-D and this according to the five methods. This observation confirms the results seen in the correlation analysis. It, moreover, contains a warning that correlation analyses may suggest wrong conclusions. In this case the poor correlations between the results obtained by the SAPV method and those obtained by the barycenter methods for groups and panel members that have no real importance (they are cognitively unrelated) should not distract from the generally better correlations for pairs that matter.

7. Conclusion

In this paper, we showed that, besides using barycenters in a two- and three dimensional base map, it is possible to derive cognitive distances in N dimensions using the SAPVs and WCS methods. Our approach is rather general: it can in principle be applied to all cases where units produce publications, which can be situated on a base map or counted in relation to a similarity matrix. Of course, other approaches are also possible, such as the one proposed by Wang and Sandström (2015), which is based on bibliographic coupling and topic modelling. Operationalizing the notion of cognitive distance is essential to several topics in informetrics, e.g. peer review processes, evaluation procedures, exploration of collaboration, and the study of interdisciplinarity. Indeed, cognitive distance could also be derived from other objects than publications, such as patents. Cognitive distance is also of essence in other contexts such as hiring decisions, political programs, and cultural differences.

As pointed out in this paper, calculating cognitive distances between units should be scale-invariant. Barycenters in a two- and three dimensional base maps satisfy this requirement. We note though that distances in a 2- or 3D map are artificial; for instance, Pajek uses coordinates in the interval [0, 1] (this also applies to its VOS implementation), whereas coordinates in VOSviewer may refer to a wider interval. Hence, only comparisons between distances and not their absolute values have meaning. Proper normalization in N dimensions also leads to scale-invariant distances.

We have shown that the barycenter method is relatively insensitive to the number of dimensions in which it is used. Yet, especially in 3D the barycenter method has little discriminatory power. Distances between normalized SAPVs in N dimensions are probably less distorted and hence more meaningful. A similar observation applies to the WCS method. Hence, our preference, based on mathematical logic, goes to the SAPVs and WCS methods. Yet, WCS scores badly in the case of chemistry, so that our final preference goes to the SAPV method. Admitting that in our case studies the barycenter methods score slightly better and that differences between the results obtained by different methods are rather small, it is obvious that the result of this comparison should not be generalized. In future research, we intend to make a similar empirical comparison for more disciplines.

In a previous approach, besides using a VOS map, we also investigated if a map based on the algorithm by Kamada and Kawai (1989) could be used. We found out however that a Kamada-Kawai map (in two and in three dimensions) can yield very different results, depending on the random seed used. For this reason, we turned to a VOS map, which is much more stable. We mistakenly mentioned in (Rahman et al., 2015) that the barycenter results in 2D were based on a Kamada-Kawai map. Also there we showed barycenter results based on a VOS map. We hope that this warning will prevent colleagues from making wrong inferences.

Finally our investigations led to two unsolved problems. The first one is the unexplained low correlation between the barycenter method in 3D and the SAPV and WCS methods for chemistry. We checked all calculations related to the barycenter method in 3D but did not detect any error. Moreover, consequent investigations related to other departments, in particular the biomedical sciences, gave similar low correlations. The second problem is the use of the main assessor, as appointed by the panel chair, as a “gold standard”. We admit that this is a problematic approach, since it relies on assumptions that are not always met. Yet, for the moment, we have not found a better solution.

Acknowledgments

This investigation has been made possible by the financial support of the Flemish government to ECOOM. The opinions in the paper are the authors’ and not necessarily those of the government. We thank reviewers and especially Ludo Waltman for constructive remarks, leading to a considerably improved manuscript.

References

Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Engels, T.C.E., Goos, P., Dexters, N., & Spruyt, E.H.J. (2013). Group size, h-index, and efficiency in publishing in top journals explain expert panel assessments of research group quality and productivity. *Research Evaluation*, 22(4), 224-236.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7-15.

Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589–593.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology,* 60(2), 348–362.

Molas-Gallart, J., Rafols, I., & Tang, P. (2014). On the relationship between interdisciplinarity and impact: different modalities of interdisciplinarity lead to different types of impact. *Journal of Science Policy and Research Management*, 29(2), 69-89.

Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management and Governance*, 4(1–2), 69–92.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9), 1871–1887.

Rahman, A.I.M.J., Guns, R., Leydesdorff, L., & Engels, T.C.E. (2016a). Measuring the match between evaluators and evaluees: cognitive distances between panel members and research groups at the journal level. *Scientometrics*, 109(3), 1639-1663.

Rahman, A.I.M.J., Guns, R., Rousseau, R., & Engels, T.C.E. (2015). Is the expertise of evaluation panels congruent with the research interests of the research groups: A quantitative approach based on barycenters. *Journal of Informetrics*, 9(4), 704–721.

Rahman, A.I.M.J., Guns, R., Rousseau, R., & Engels, T.C.E. (2016b). Corrigendum to "Is the expertise of evaluation panels congruent with the research interests of the research groups: A quantitative approach based on barycenters" [Journal of Informetrics 9 (4) (2015) 704–721]. *Journal of Informetrics*, 10(4), 1052-1054.

Van Eck, N.J., & Waltman, L. (2007). VOS: a new method for visualizing similarities between objects. In H.-J. Lenz. & R. Decker (Eds.). *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society* (pp. 299-306). Springer.

Wang, Q., & Sandström, U. (2015). Defining the role of cognitive distance in the peer review process with an explorative study of a grant scheme in infection biology. *Research Evaluation*, 24(3), 271-281.

Zhou, Q., Rousseau, R., Yang, L., Yue, T., & Yang, G. (2012). A general framework for describing diversity within systems and similarity between systems with applications in informetrics. *Scientometrics*, 93(3), 787-812.

Appendix A. Euclidean distances between barycenters

Table A1: Euclidean distances between barycenters of research groups and panel members of the Chemistry department using the 2-dimensional WoS SCs map.

	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
PM 1	0.167	**0.129**	0.217	**0.165**	0.329	0.337	0.179	0.165	0.111	0.394	0.454	**0.127**
PM 2	0.350	0.342	0.362	**0.129**	**0.079**	**0.090**	0.145	0.215	0.199	**0.259**	**0.228**	0.342
PM 3	0.171	0.161	0.192	**0.129**	0.252	**0.263**	**0.053**	**0.061**	**0.020**	**0.269**	0.330	0.161
PM 4	0.269	0.262	0.280	**0.108**	**0.158**	**0.170**	**0.063**	**0.134**	0.121	**0.232**	**0.250**	0.263
PM 5	**0.056**	**0.055**	**0.091**	0.232	0.367	0.378	0.154	**0.093**	**0.099**	**0.315**	0.411	**0.057**
PM 6	0.302	0.276	0.335	**0.027**	**0.175**	**0.181**	0.161	0.210	0.156	0.366	0.370	0.275
PM 7	**0.116**	**0.072**	0.172	0.235	0.395	0.404	0.216	0.178	0.144	0.410	0.491	**0.070**

Table A2. Euclidean distances between barycenters of research groups and panel members of the Physics department using the 2-dimensional WoS SCs map.

	PHYS-A	PHYS- B	PHYS-C	PHYS- D	PHYS-E	PHYS- F	PHYS- G	PHYS- H	PHYS-I
PM 1	1.173	0.123	0.215	**0.017**	0.145	**0.208**	0.495	0.120	0.664
PM 2	1.195	0.067	0.109	0.158	0.118	0.316	0.443	0.056	0.688
PM 3	**1.041**	0.146	0.194	0.116	0.113	**0.104**	**0.387**	0.157	**0.532**
PM 4	**1.020**	0.168	**0.085**	0.263	0.132	0.295	**0.249**	0.179	**0.522**
PM 5	**1.136**	**0.046**	**0.055**	0.159	**0.069**	0.281	**0.385**	**0.050**	**0.629**
PM 6	**1.157**	**0.031**	0.084	0.138	**0.078**	0.280	**0.412**	**0.026**	0.649

Appendix B. A mathematical caveat

In this appendix we show that weighted cosine similarity cannot be used with any similarity matrix but that the problem does not occur for the similarity matrices used by us. We illustrate this with the unweighted cosine similarity (the numerator of formula (5)).

In a general (real or complex) vector space it is possible that if expressions of the form $\left(M^{k}\right)^{t} * S * R^{m}$, with S a symmetric matrix, are used as similarity measures, some non-null vectors have similarity zero to themselves. This excludes this type of construction as a general method for calculating similarities.

We consider the symmetric matrix $S = \begin{pmatrix} 1 & 0.8 & 0.9 \\ 0.8 & 1 & 0 \\ 0.9 & 0 & 1 \end{pmatrix}$, see (Zhou et al., 2012). and want to find a vector X = $(u,v,w)^t$, (u,v,w: real numbers) such that $\left(X\right)^{t} * S * X$ = 0. Replacing X by $(u,v,w)^t$ leads to the requirement: $u^2+1.6uv+1.8uw+v^2+w^2 = 0$. Taking u = 1, v ≈ -1.44031 and w = -1.1 provides an (approximate) solution. In fact this is just one solution among infinitely many.

If u = 1 and w = K then v_1 = $-\left(\sqrt{-K^2 - 1.8 * K - 0.36} + 0.8\right)$ and v_2 = $\left(\sqrt{-K^2 - 1.8 * K - 0.36} - 0.8\right)$ always provide solutions (some of which may be complex numbers). The one given above is v_1 with K = -1.1. This solution was obtained using TI-*n*spire software.

We check now that v_1 and v_2 as given above, indeed lead to the perfect null solution. Writing $\sqrt{-K^2 - 1.8 * K - 0.36}$ as R and using v_1 we find:

u^2+1.6 uv+1.8uw+v^2+w^2 = 1 -1.6 R – 1.28 + 1.8 K + (-K^2 - 1.8K -0.36) + 1.6 R + 0.64 + K^2 = (1-1.28-0.36+0.64)+(1.8-1.8)K+ (-K^2+K^2) +R(-1.6+1.6) = 0

Similarly, with v_2 we obtain: u^2+1.6 uv+1.8uw+v^2+w^2 = 1 +1.6 R – 1.28 + 1.8 K + (-K^2 – 1.8K -0.36) – 1.6 R + 0.64 + K^2 = (1-1.28-0.36+0.64)+(1.8-1.8)K+ (K^2-K^2) +R(1.6-1.6) = 0.

However, this problem cannot occur when the matrix S has non-negative values and when, moreover, the vector X has only non-negative values, which is precisely the context in which we work. Indeed: under these circumstances the expression $(X)^t * S * X$ is always non-negative and only zero when X = 0 (the zero-vector) and this in any dimension. Note that the example presented above led to a vector X with two negative coordinates.