Determining cognitive distance between publication portfolios of evaluators and evaluees in research evaluation: A case study of Physics department

TECHNICAL REPORT

A. I. M. Jakaria Rahman and Raf Guns

jakaria.rahman@uantwerpen.be, raf.guns@uantwerpen.be

Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences

University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium

Antwerp, 2017

This technical report is prepared in the context of A. I. M. Jakaria Rahman's PhD project on *Determining cognitive distance between publication portfolios of evaluators and evaluees in research evaluation: Exploration of informetric methods.* Similar technical reports on Biology, Biomedical Sciences, Pharmaceutical Sciences, Chemistry and Veterinary Sciences department are also available at the institutional repository of the University of Antwerp (https://repository.uantwerpen.be).

Table of Contents

List of Tables

List of Figures

1 Introduction

We study the problem of composing an expert panel, such that the individual panel members' expertise covers the specific subdomains in the discipline where the units of assessment (in our case: research groups) have publications. We explore expertise overlap between panel and research groups through publishing in the same or similar Web of Science subject categories (WoS SCs) and journals. We use the data collected in the framework of completed research evaluations by the University of Antwerp (Belgium) through site visits by the expert panel members. We specifically focus on the situation where the expert panel needs to evaluate all the research groups of a department.

Research evaluations carried out at the University of Antwerp are organized by its Department of Research Affairs and Innovation (ADOC). At the start of a research evaluation, a department – typically encompassing several research groups – is invited to suggest potential panel chairs in addition to those suggested by the ADOC. Preferably, chairs are appointed as full professor, have an excellent publication record, have experience in research evaluations, are editors or board members of important journals, and possess academic management experience. The ADOC verifies whether proposed panel chairs and members have no prior involvement (i.e. no prior joint affiliations, no co-publications, no common projects) with the assessed research groups, and further checks if they are scholars with a prominent publication record in recent years, a proven track record of training young researchers, and sufficient experience in research policy, preferably in academic leadership positions. Furthermore, proposed panel chairs and members are preferably not affiliated with any Flemish institution of higher education and have no formal links to the University of Antwerp. The department that is being evaluated is also allowed to suggest potential panel members, but it should be noted that it is eventually the chair's prerogative to decide on the final composition of the panel.

The combined expertise of all panel members is to cover all subdomains in the discipline that is being evaluated and the panel is preferably balanced in terms of gender and nationality. When a sufficient number of professors have agreed to be on the panel, the university's research council ratifies the panel composition. Furthermore, all research groups belonging to a specific department (e.g., Physics) are to be evaluated by the same panel and the language of communication is English. Following the Dutch Standard Evaluation Protocol (VSNU, 2003; VSNU, KNAW, & NWO, 2014), the peer panels assess the quality, the productivity, the relevance and the viability of each research group.

These evaluations consider the entire research groups' scientific activity for a specific period, typically 8 years preceding the year of evaluation. All articles, letters, notes, proceeding papers, and reviews by the research groups published during the reference period are included in the evaluation. In this report, we consider only the publications that are index in Science Citation Index Expanded (SCIE) and Social Sciences Citation Index (SSCI) of WoS.

Research groups at the University of Antwerp (Belgium) consist of professors (of all ranks), research and teaching assistants, and researchers (PhD students and postdocs). A research group consists either of one professor assisted by junior and/or senior researchers, or of a group of professors and a number of researchers linked to them.

An expert panel typically consists of independent specialists, and is multidisciplinary and/or interdisciplinary in its composition; each of the members are recognized experts in at least one of the fields addressed by the department under evaluation. However, the degree to which the expertise of the panel (members) overlaps with the expertise of the research groups has not been quantified to date. The goal is therefore to present Informetric bibliometric methodologies to assess the congruence of panel expertise and research interests in the units under assessment. As such, we present a bibliometric analysis of the overlap of expertise between research groups in the Departments of Physics and the respective expert panels based on research evaluations carried out at the University of Antwerp.

In this technical report, we present the Physics department's research groups and panel members. We describe our methods step by step. This report is divided into four parts. Firstly, we describe the technical steps for all of our three methods (barycenter, similarity-adapted publication vector, and weighted cosine similarity) using WoS SCs (Section 2). Secondly, we present the three methods using journals (section 3). In the third and fourth part, we present a heat map of spearman's rank-order correlation coefficient between each pair of the six approaches (section 4) and the programming code for the main methods used respectively (section 5). Finally, we present overlay maps and location of similarity adapted publication vector of Physics individual research groups, all research groups together, panel members and panel (all panel members together) in WoS SCs and journals in the appendix.

2 Cognitive distance based on Web of Science subject categories

2.1 Data collection process

We collect data from the 2010 assessment of the 9 research groups of the Department of Physics, University of Antwerp. First, from ADOC, we collect all the WoS accession numbers of the publications of each research group. We replace the name of the research groups with code names PHYS-A, PHYS-B etc.

a) Research groups data retrieval

We remove the prefix 'WOS:' from the accession numbers and use a Python script to put 'OR' in between the accession numbers to create a long search string. We do a basic search in WoS with the accession numbers of each research group, keeping the time span to all years and searching SCIE and SSCI. We use the 'Analyze Results' option in the WoS, and rank the records by WoS SCs with the minimum set to 1. We save the resulting list as 'analyze.txt' and subsequently save a copy of the file named '[Research group code]_WoS SCs.txt', for example 'PHYS-A_WoS SCs.txt' and keep both files.

Table 1: Publication statistics of Physics research groups (2002-2009)

Group code	Number of Publications	Number of Journals	Number of WoS SCs
Physics research groups			
PHYS-A	125	53	44
PHYS-B	486	66	25
PHYS-C	525	147	46
PHYS-D	269	17	7
PHYS-E	159	55	28
PHYS-F	42	23	13
PHYS-G	43	26	12
PHYS-H	132	31	12
PHYS-I	115	63	49
All groups	**1739**	**353**	**108**

Table 1 lists the publication profile of the physics research groups during the eight years preceding their evaluation. The Physics research groups generated 1739 publications over 353 journals. Members of two research groups co-authored 150 publications and three research groups co-authored seven publications. In total, their publications are distributed over 108 WoS SCs.

We combine the search sets for each research group from the search history of the WoS, and get the data for the publications of the department as a whole. In this way, any publication that has been co-authored by members of two or more research groups is counted only once. We use the ‘Analyze Results’ option in the WoS, and rank the record by WoS SCs with the minimum set to 1. We save the resulting list as ‘analyze.txt’ and subsequently save a copy of the file named ‘Groups together_WoS SCs.txt’.

b) Panel members data retrieval

The Physics panel was composed of 6 panel members (including the chair). We have obtained the names and curricula vitae of the panel members from the ADOC. We replace the original name of each panel member with a code name: PM1, PM2 etc. We perform an advanced search for each panel member in WoS through checking the SCIE and SSCI. All the publications of the individual panel members up to the year of assessment (2010) were taken into account. We use the ‘Analyze Results’ option in the WoS, and rank the record by WoS SCs with the minimum set to 1. We save the resulting list as ‘analyze.txt’ and subsequently save a copy of the file named ‘[PM code]_WoS SCs.txt’ for example, ‘PM1_WoS SCs.txt’.

Table 2. Publication statistics of Physics panel members

Panel code	Number of Publications	Number of Journals	Number of WoS SCs
PM1	117	7	3
PM2	168	15	4
PM3	124	49	10
PM4	166	40	10
PM5	247	87	10
PM6	282	54	10
Panel	**1104**	**204**	**46**

Table 2 lists the publication statistics of the Physics panel members. The combined publication output of the Physics panel members consists of 1104 publications. None of which is co-authored publications between panel members. The number of publications per panel member ranges from 117 to 282. In total, these publications appeared in 204 different journals and are assigned to 46 different WoS SCs.

We combine the search sets for each panel member from the search history of the WoS, and get the result for the panel as a whole. In this way, any co-authored publication between two or more panel members is counted only once. Again, we use the ‘Analyze Results’ option in the WoS, and rank the record by WoS SCs with the minimum set to 1. We save the resulting list as ‘analyze.txt’ and subsequently save a copy of the file named ‘Panel_WoS SCs.txt’.

	A	B	C
1	**Web of Science Categories**	**records**	**% of 1732**
2	PHYSICS CONDENSED MATTER	515	29.734
3	PHYSICS APPLIED	252	14.55
4	PHYSICS MULTIDISCIPLINARY	231	13.337
5	MATERIALS SCIENCE MULTIDISCIPLINARY	226	13.048
6	CHEMISTRY PHYSICAL	193	11.143
7	PHYSICS PARTICLES FIELDS	154	8.891
8	NANOSCIENCE NANOTECHNOLOGY	111	6.409
9	MICROSCOPY	72	4.157
10	PHYSICS ATOMIC MOLECULAR CHEMICAL	66	3.811
11	OTORHINOLARYNGOLOGY	65	3.753
12	CHEMISTRY INORGANIC NUCLEAR	56	3.233
13	PHYSICS MATHEMATICAL	56	3.233
14	CHEMISTRY MULTIDISCIPLINARY	50	2.887
15	OPTICS	42	2.425

GroupsTogether | PHYS-B | PHYS-C | PHYS-D | PHYS-E | PHYS-I

Figure 1. Excerpt of Physics research groups and panel members_WoS SCs.xlsx file

The downloaded data files, ‘[Research group code]_WoS SCs. txt’, ‘[PM code]_WoS SCs. txt’, ‘Groups_WoS SCs.txt’ and ‘Panel_WoS SCs.txt’, have been exported to an MS Excel file. The sheets in the Excel file contain data on and are named after the research groups’ code names (PHYS-A, PHYS-B, PHYS-C, etc.), the panel members’ code names, (PM1, PM2, PM3, etc.), Panel together and Groups together. The Excel file is saved as ‘Physics research groups and panel_WoS SCs.xlsx’ (Figure 1).

2.2 Correlation between publication profiles of research groups together and panel

a) Pearson’s correlation coefficient and Spearman’s rank-order correlation coefficient

We determine the correlation between the publication output of research groups together and and panel, using Pearson’s correlation coefficient and Spearman’s rank-order correlation coefficient for the numbers of publications per WoS SC. We make an Excel file ‘Physics panel and research groups together_WoS SCs.xlsx' (Figure 2) and export data from 'Panel_WoS SCs.txt' and 'Groups together_WoS SCs.txt' in two different sheets.

A Python script 'join-sheets.py' is used to take the data of the two sheets and join it into one. We run the program as:

```
python join-sheets.py "Physics Panel and research groups together_WoS SCs.xlsx"
```

	A	B	C
1	**Web of Science Categories**	**records**	**% of 1732**
2	PHYSICS CONDENSED MATTER	515	29.734
3	PHYSICS APPLIED	252	14.55
4	PHYSICS MULTIDISCIPLINARY	231	13.337
5	MATERIALS SCIENCE MULTIDISCIPLINARY	226	13.048
6	CHEMISTRY PHYSICAL	193	11.143
7	PHYSICS PARTICLES FIELDS	154	8.891
8	NANOSCIENCE NANOTECHNOLOGY	111	6.409
9	MICROSCOPY	72	4.157
10	PHYSICS ATOMIC MOLECULAR CHEMICAL	66	3.811
11	OTORHINOLARYNGOLOGY	65	3.753
12	CHEMISTRY INORGANIC NUCLEAR	56	3.233
13	PHYSICS MATHEMATICAL	56	3.233
14	CHEMISTRY MULTIDISCIPLINARY	50	2.887
15	OPTICS	42	2.425
16	PHYSICS FLUIDS PLASMAS	39	2.252
17	ENGINEERING ELECTRICAL ELECTRONIC	38	2.194
18	METALLURGY METALLURGICAL ENGINEERING	38	2.194
19	ASTRONOMY ASTROPHYSICS	31	1.79
20	RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGIN	27	1.559
21	INSTRUMENTS INSTRUMENTATION	25	1.443
22	NEUROSCIENCES	25	1.443
23	SPECTROSCOPY	20	1.155
24	AUDIOLOGY SPEECH LANGUAGE PATHOLOGY	18	1.039
25	CRYSTALLOGRAPHY	17	0.982
26	ENGINEERING BIOMEDICAL	17	0.982
27	BIOPHYSICS	16	0.924
28	NUCLEAR SCIENCE TECHNOLOGY	16	0.924
29	BIOCHEMISTRY MOLECULAR BIOLOGY	15	0.866
30	CLINICAL NEUROLOGY	15	0.866
31	CARDIAC CARDIOVASCULAR SYSTEMS	13	0.751
32	COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	13	0.751

GroupsTogether PanelTogether

	A	B	C
1	**Web of Science Categories**	**records**	**% of 1104**
2	PHYSICS CONDENSED MATTER	469	35.557
3	PHYSICS APPLIED	220	16.679
4	OPTICS	214	16.224
5	CHEMISTRY PHYSICAL	213	16.149
6	MATERIALS SCIENCE MULTIDISCIPLINARY	154	11.676
7	PHYSICS MULTIDISCIPLINARY	150	11.372
8	PHYSICS ATOMIC MOLECULAR CHEMICAL	120	9.098
9	PHYSICS PARTICLES FIELDS	116	8.795
10	NANOSCIENCE NANOTECHNOLOGY	73	5.534
11	INSTRUMENTS INSTRUMENTATION	61	4.625
12	CHEMISTRY MULTIDISCIPLINARY	58	4.397
13	ASTRONOMY ASTROPHYSICS	55	4.17
14	PHYSICS NUCLEAR	53	4.018
15	SPECTROSCOPY	42	3.184
16	MICROSCOPY	27	2.047
17	ENGINEERING ELECTRICAL ELECTRONIC	26	1.971
18	METALLURGY METALLURGICAL ENGINEERING	20	1.516
19	MATERIALS SCIENCE COATINGS FILMS	11	0.834
20	PHYSICS MATHEMATICAL	10	0.758
21	MECHANICS	10	0.758
22	IMAGING SCIENCE PHOTOGRAPHIC TECHNOLOGY	10	0.758
23	BIOCHEMICAL RESEARCH METHODS	10	0.758
24	MATERIALS SCIENCE CHARACTERIZATION TESTING	6	0.455
25	PHYSICS FLUIDS PLASMAS	5	0.379
26	POLYMER SCIENCE	4	0.303
27	ENGINEERING MANUFACTURING	4	0.303
28	ENDOCRINOLOGY METABOLISM	4	0.303
29	CRYSTALLOGRAPHY	4	0.303
30	ENGINEERING MECHANICAL	3	0.227
31	ENGINEERING CHEMICAL	3	0.227
32	COMPUTER SCIENCE INTERDISCIPLINARY APPLICAT	3	0.227

GroupsTogether **PanelTogether**

Figure 2. Excerpt of the Physics panel and research groups together_WoS SCs.xlsx file

This produces a new Excel file called 'Physics panel and groups together_WoS SCs-joined.xlsx' (Figure 3). To calculate the correlation, the value zero was kept on the corresponding WoS SCs in which either the panel or the groups had no publications (but not both). Using the data from the file, we calculate correlation coefficient in SPSS (Statistical Package for the Social Sciences) and find value ($r = 0.92$, $\rho = 0.52$). Figure 4 shows a log-log plot of the number of publications per WoS SCs for the Physics panel and research groups together.

	A	B	C	D
1		**Web of Science Categories**	**records_x**	**records_y**
2	**0**	PHYSICS CONDENSED MATTER	515	469
3	**1**	PHYSICS APPLIED	252	220
4	**2**	PHYSICS MULTIDISCIPLINARY	231	150
5	**3**	MATERIALS SCIENCE MULTIDISCIPLINARY	226	154
6	**4**	CHEMISTRY PHYSICAL	193	213
7	**5**	PHYSICS PARTICLES FIELDS	154	116
8	**6**	NANOSCIENCE NANOTECHNOLOGY	111	73
9	**7**	MICROSCOPY	72	27
10	**8**	PHYSICS ATOMIC MOLECULAR CHEMICAL	66	120
11	**9**	OTORHINOLARYNGOLOGY	65	0
12	**10**	CHEMISTRY INORGANIC NUCLEAR	56	0
13	**11**	PHYSICS MATHEMATICAL	56	10
14	**12**	CHEMISTRY MULTIDISCIPLINARY	50	58
15	**13**	OPTICS	42	214
16	**14**	PHYSICS FLUIDS PLASMAS	39	5
17	**15**	ENGINEERING ELECTRICAL ELECTRONIC	38	26
18	**16**	METALLURGY METALLURGICAL ENGINEERING	38	20
19	**17**	ASTRONOMY ASTROPHYSICS	31	55
20	**18**	RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING	27	0
21	**19**	INSTRUMENTS INSTRUMENTATION	25	61
22	**20**	NEUROSCIENCES	25	0

Figure 3. Excerpt of the Physics panel and groups together_WoS SCs - joined.xlsx file

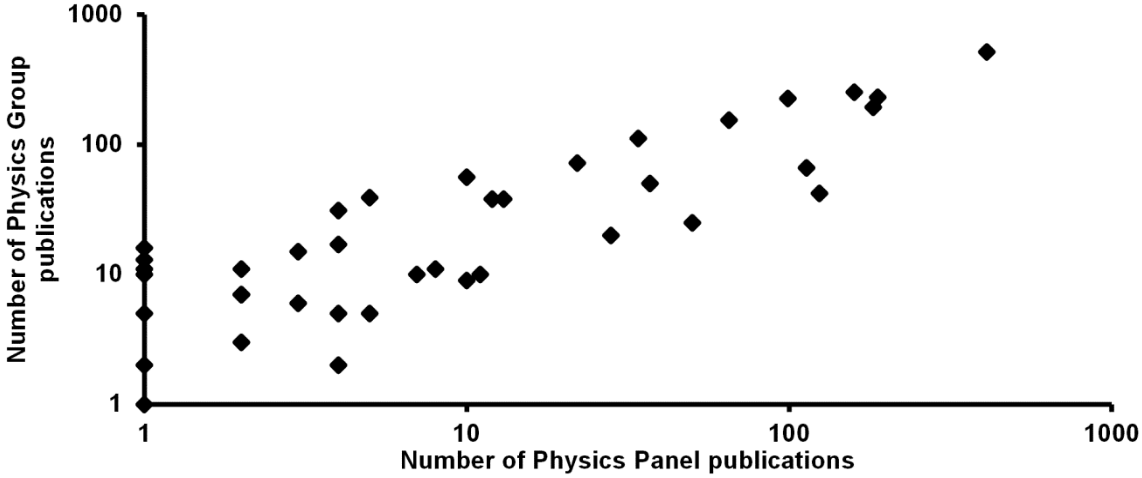


Figure 4. Log-log plot of the number of publications (log-log scale) per WoS SC for the panel (vertical axis) and research groups together (horizontal axis) of the Physics department

b) Top-Down correlation coefficient

In some cases, the panel publications belong to a WoS SC in which the research groups have not published or vice versa, i.e. there are many zeroes on both sides. Since traditional correlation coefficients like Pearson's and Spearman's are not well-suited to zero-inflated data (i.e., data with a large amounts of zeroes), we adopt the top-down correlation coefficient (Iman & Conover, 1987). This correlation coefficient was found to be an adequate rank correlation coefficient for zero-inflated data (Huson, 2007). For a full description of the Top-down correlation coefficient we refer to Iman and Conover (1987). This coefficient places emphasis on the higher ranked data by computing the correlation using Savage scores derived from the ranked data.

Savage scores are calculated as follows:

$$S_i = \sum_{j=i}^{n} 1/j \tag{1}$$

where i is an item's rank among a set of n items. For instance, if n = 3, the three Savage scores are $S_1 = 1 + \frac{1}{2} + \frac{1}{3}$, $S_2 = \frac{1}{2} + \frac{1}{3}$, and $S_3 = \frac{1}{3}$. The Top-down correlation coefficient is calculated as:

$$r_{td} = (\sum_{i=1}^{n} S_{R_i}\ S_{Q_i} - n)/(n - S_1) \tag{2}$$

where S is the Savage score, R_i and Q_i are the ranks of the data in the two samples, and n is the sample size. In case of ties, we use the average Savage score. We use a Python script 'calc_topdowncorr.py' (all core logic is in topdowncorr.py, see section 5) for top-down correlation taking into account formulas (1) and (2).

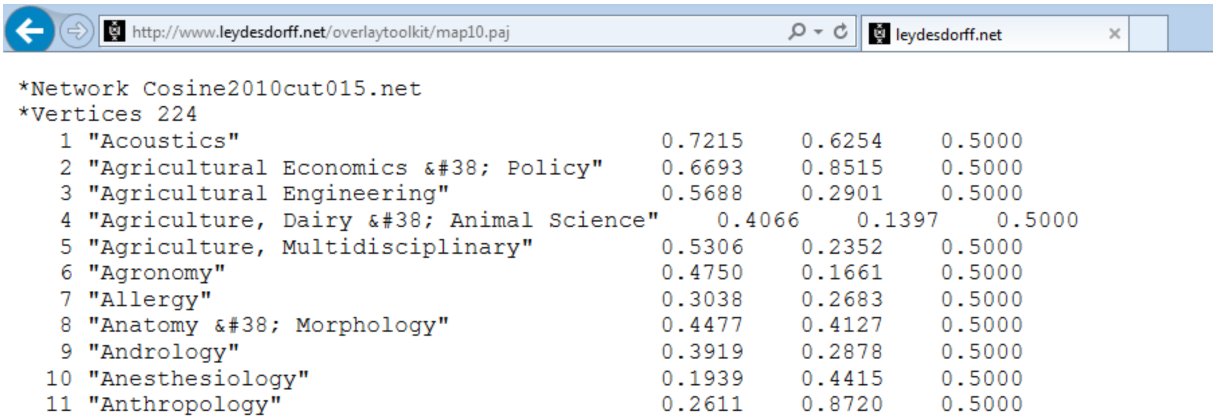
We reuse the 'Excerpt of the Physics panel and groups together_WoS SCs - joined.xlsx' (Figure 3) file, but keep the zeros in the WoS SCs where neither the panel nor the research groups have publications. We run the program as:

```
python calc_topdowncorr.py "Physics panel and research groups together_WoS SCs-
joined.xlsx"
```

The outcome shows that the top-down correlation between Physics research groups together and the panel's profile in the WoS SCs is low (0.39). In our opinion, the correlations are an insufficient measure in this case, as the similarity of WoS SCs is not taken into account here. This is reminiscent of the way diversity is sometimes studied using only the dimensions of variety and balance. As discussed by Stirling (2007), the additional dimension of disparity – the opposite concept of similarity – is needed to provide a complete picture. Likewise, a comparison of publication profiles based on WoS SCs that does not take WoS SC similarity into account might yield distorted results.

2.3 Web of Science subject categories similarity matrix

We download the global map of science based on WoS SCs data made available at http://www.leydesdorff.net/overlaytoolkit/map10.paj. These authors (Leydesdorff & Rafols, 2009; Rafols, Porter, & Leydesdorff, 2010; Leydesdorff, Carley, & Rafols, 2013) created a matrix of citing to cited WoS SCs based on the SCIE and SSCI, which was subsequently normalized in the citing direction. Only cosine values > 0.15 were retained. The result is a symmetric N×N similarity matrix (here, N=224). If we interpret it as an adjacency matrix, we see that it is equivalent to a weighted network, in which similar categories are linked (the higher the link weight, the stronger the similarity). The file 'map10.paj' contains this weighted network of WoS SCs.




```
*Network Cosine2010cut015.net
*Vertices 224
   1 "Acoustics"                               0.7215     0.6254     0.5000
   2 "Agricultural Economics &#38; Policy"     0.6693     0.8515     0.5000
   3 "Agricultural Engineering"                0.5688     0.2901     0.5000
   4 "Agriculture, Dairy &#38; Animal Science"     0.4066     0.1397     0.5000
   5 "Agriculture, Multidisciplinary"          0.5306     0.2352     0.5000
   6 "Agronomy"                                0.4750     0.1661     0.5000
   7 "Allergy"                                 0.3038     0.2683     0.5000
   8 "Anatomy &#38; Morphology"                0.4477     0.4127     0.5000
   9 "Andrology"                               0.3919     0.2878     0.5000
  10 "Anesthesiology"                          0.1939     0.4415     0.5000
  11 "Anthropology"                            0.2611     0.8720     0.5000
```

Figure 5. Excerpt of the map10.paj file

We download the 'map10.paj' (Figure 5) file and open the file in Pajek (available at http://mrvar.fdv.uni-lj.si/pajek) and save the network as 'map10.net'. The information in the network file can be visualized. The subfield of bibliometric mapping is dedicated to the visualization, clustering and interpretation of similarity matrices or networks like the one we use. Many different algorithms or layout techniques have been developed for this purpose. We have used two layout techniques:

i) Kamada-Kawai (Kamada & Kawai, 1989) is a spring-based layout algorithm for networks, which is implemented in, among others, Pajek (de Nooy, Mrvar, & Batagelj, 2012). Kamada-Kawai is the algorithm used by Rafols et al., (2010)

i) VOS (van Eck & Waltman, 2007) stands for ‘visualization of similarities’ and is a variant of multidimensional scaling (Borg & Groenen, 2005; van Eck, Waltman, Dekker, & van den Berg, 2010). It is implemented in VOSviewer and in recent versions of Pajek.

Figure 6 shows the transformation of WoS SC similarity matrix to Kamada-Kawai and VOS map. It provides an overview of the relations between similarity matrix, network and the two maps. Since the source data include all research fields included in the SCI and SSCI, the resulting maps are global maps of science (as opposed to local maps of science, which focus on one or a few disciplines).

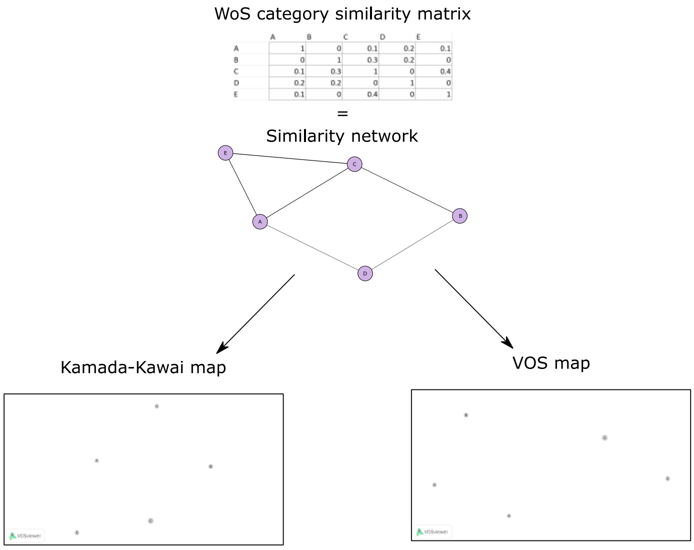


Figure 6. Transformation of WoS SCs similarity matrix to Kamada-Kawai map and VOS map

We run VOSviewer (http://www.vosviewer.com) and click on ‘Create’ from the action tab. It offers to create a map based on a network. We select this option and in the next step through Pajek tab, we choose the ‘map10.net’ file and click on the next button. It prompts us to choose whether we want to use the coordinates that are in the file or want to calculate new ones (Figure 7).

We choose ‘Yes’ to keep using the Kamada-Kawai coordinates. We save the map as ‘Kamada-Kawai.txt’ file, export the data to an Excel file, and save as ‘WoS SCs_Kamada-Kawai map.xlsx’ (Figure 8).

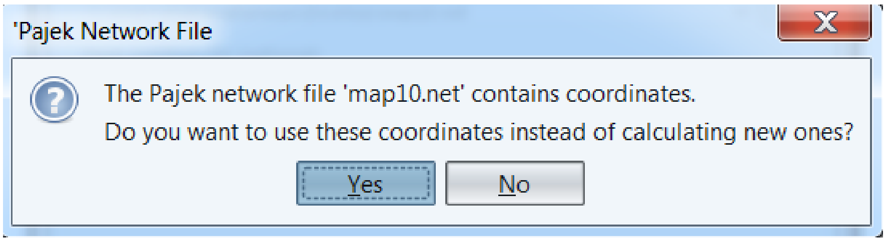


Figure 7. VOSviewer message before choosing Kamada Kawai map or VOS map data

Again, we run VOSviewer and click on 'Create' from the action tab. It offers to create a map based on a network. We select this option and in the next step through the Pajek tab, we choose the 'map10.net' file and click on the next button. It again prompts us to choose whether we want to use the coordinates that are in the file or want to calculate new ones (Figure 7). We choose 'No' to let VOSviewer calculate the coordinates according to its own VOS algorithm (Figure 9).

	A	B	C	D	E	F
1	**id**	**label**	**x**	**y**	**weight**	**cluster**
2	1	Acoustics	0.7215	0.6254	4.596	2
3	2	Agricultural Economics, Policy	0.6693	0.8515	5.484	3
4	3	Agricultural Engineering	0.5688	0.2901	15.708	4
5	4	Agriculture, Dairy, Animal Science	0.4066	0.1397	6.208	1
6	5	Agriculture, Multidisciplinary	0.5306	0.2352	15.422	4
7	6	Agronomy	0.475	0.1661	10.28	4
8	7	Allergy	0.3038	0.2683	8.034	1
9	8	Anatomy, Morphology	0.4477	0.4127	28.204	1
10	9	Andrology	0.3919	0.2878	16.946	1
11	10	Anesthesiology	0.1939	0.4415	9.528	1

Figure 8. Excerpt of WoS SCs Kamada-Kawai map data

	A	B	C	D	E	F
1	**id**	**label**	**x**	**y**	**weight**	**cluster**
2	1	Acoustics	-0.6842	-0.1274	4.596	2
3	2	Agricultural Economics, Policy	1.285	-0.1948	5.484	3
4	3	Agricultural Engineering	-0.6952	0.0854	15.708	4
5	4	Agriculture, Dairy, Animal Science	-0.2417	0.2277	6.208	1
6	5	Agriculture, Multidisciplinary	-0.4803	0.1691	15.422	4
7	6	Agronomy	-0.5502	0.1844	10.28	4
8	7	Allergy	0.104	0.2443	8.034	1
9	8	Anatomy, Morphology	0.0029	0.1922	28.204	1
10	9	Andrology	-0.0119	0.2208	16.946	1
11	10	Anesthesiology	0.3086	0.1703	9.528	1

Figure 9. Excerpt of WoS SCs VOS map data

	A	B	C	D	E	F
1	id	label	x	y	weight	cluster
2	1	Acoustics	0.4329	-0.1442	4.596	2
3	2	Agricultural Economics Policy	-0.9319	-0.2762	5.484	3
4	3	Agricultural Engineering	0.699	0.0866	15.708	4
5	4	Agriculture, Dairy Animal Science	0.2792	0.3684	6.208	1
6	5	Agriculture, Multidisciplinary	0.5097	0.2382	15.422	4
7	6	Agronomy	0.5917	0.2442	10.28	4
8	7	Allergy	-0.0432	0.5132	8.034	1
9	8	Anatomy Morphology	0.1085	0.3599	28.204	1
10	9	Andrology	0.0906	0.427	16.946	1
11	10	Anesthesiology	-0.2826	0.3844	9.528	1

Figure 10. Excerpt of WoS SCs VOS map

However, we have observed the coordinates of the VOS map that we derived from the map10.paj file is different that the VOS map available at http://www.leydesdorff.net/overlay toolkit while creating overlap map (Figure 10). We use this VOS map (Figure 10) as this map is readily available and applied for creating overlay maps (Leydesdorff, Carley, et al., 2013; Rafols et al., 2010). The details of obtaining this VOS map have been discussed in the next section. In this technical report, calculations of barycenters, Euclidean distance comparisons, and visual explorations are based on the VOS map of WoS SCs (Figure 10).

2.4 Web of Science subject categories overlay map creation

During data collection (see section 2.1, the resulting files are downloaded using the default name 'analyze.txt'. We download the 'WC10.exe' program from http://www.leydesdorff.net/overlay toolkit. This file 'analyze.txt' transformed by the mini-program 'WC10.exe' to 'WC10.vec' for upload into Pajek as a vector, and generate files like 'vos4.csv', 'vos6.csv', and 'vos19.csv' for use in VOSviewer (with 4, 6 or 19 base colors for the clusters, respectively). We keep 'analyze.txt' and 'WC10.exe' in a folder and run the exe file. The program 'WC10.exe' generates three map files: 'vos4.csv', 'vos6.csv', and 'vos19.csv'. We open the 'vos19.cs'v in VOSviewer. For example, Figure 11 shows PHYS-B research group's publications overlay map in WoS SCs.

The 'vos4.csv', 'vos6.csv', and 'vos19.csv' map files contain the VOS map as mentioned in the previous section. We save the map data to an Excel file, and save as 'WoS SCs_VOS map.xlsx' (Figure 10).

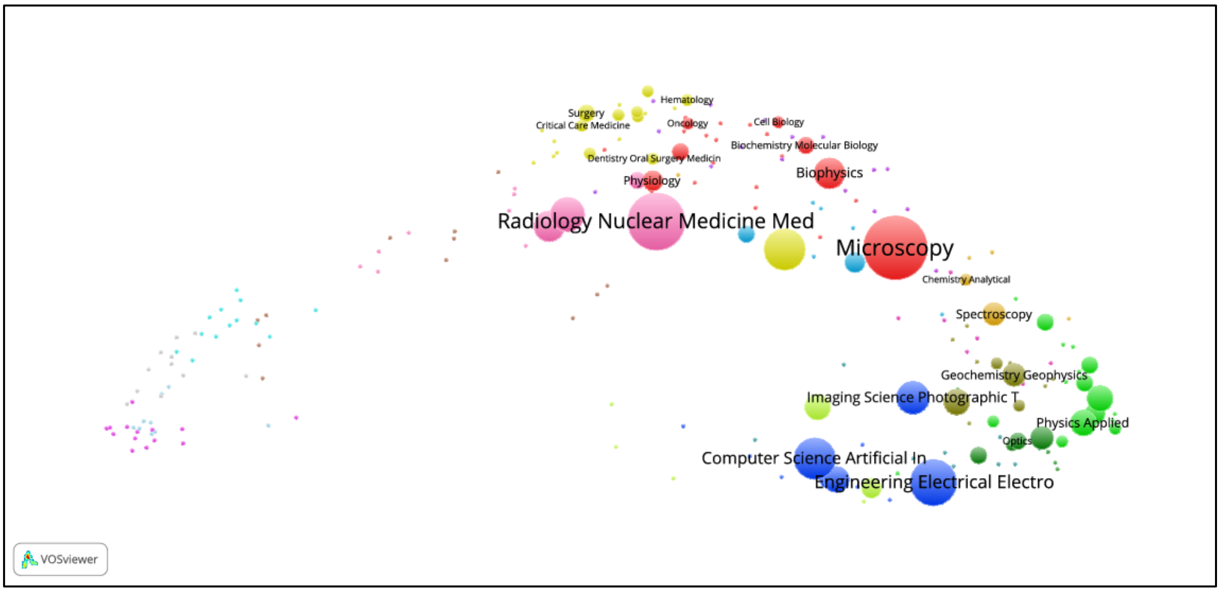


Figure 11. PHYS-B research group's publication overlay map in WoS SCs

We prepare overlay maps for each research group, each panel member, research groups together and panel (see the Appendix A).

2.5 Bootstrapping and confidence intervals

The barycenter (discussed in section 2.6 and 3.5) and Similarity-adapted vector (SAPV) methods (discussed in section 2.7 and 3.6) determine cognitive distance, on the basis of the WoS SCs/journals in which the groups and panel members have published. In the same way, Weighted cosine similarity method (discussed in section 2.8 and 3.7) determine similarity n the basis of the WoS SCs/journals in which the groups and panel members have published. However, such information is not entirely deterministic; it is, for instance, dependent on the database used as well as environmental factors like the speed with which a journal processes a submission. It logically follows that small differences in Euclidean distances or similarity bear little meaning.

To study this problem in a more systematic way, we employ a bootstrapping approach in order to determine 95 % confidence intervals (CIs) to each Euclidean distance (both between barycenters and SAPVs) and similarity. If two CIs do not overlap, the difference between the distances is statistically significant at the 0.05 level. Although it is possible for overlapping CIs to have a statistically significant difference between the corresponding distances, the difference between the distances is less likely to have practical meaning.

Bootstrapping (Efron & Tibshirani, 1998) is a simulation-based method for estimating standard error and confidence intervals. Bootstrapping depends on the notion of a bootstrap sample. To determine a bootstrap sample for a panel member or research group with N publications, we randomly sample with replacement N publications from its set of publications. In other words, the same publication can be chosen multiple times. Some publications in the original data set will not occur in the bootstrap data set, whereas others will occur once, twice or even more times. From the bootstrap sample, one can calculate a bootstrap replication, in our case a barycenter using formula (3), an SAPV using formula (5), and WCS using formula (7).

By generating a large amount of independent bootstrap samples (in our case 1000) and each time calculating the bootstrap replication, we can approximate the variability within the data set. Since we have a two-sample problem (distance between two entities; Efron & Tibshirani, 1998, Ch. 8), we calculate the distances between pairs of bootstrap replications, from which we obtain a CI using a bootstrap percentile approach (Efron & Tibshirani, 1998, Ch. 13). In the case of WCS, we generate 1000 independent bootstrap sample for both entities and calculate the similarity between them using formula 7. A more detailed explanation and implementation of our method is available on Github (http://nbviewer.jupyter.org/gist/rafguns/6fa3460677741e356538337003692389 and http://nbviewer.jupyter.org/gist/rafguns/faff8dc090b67a78 3b85d488f88952ba).

2.6 Barycenter method

a) Barycenter calculation

The barycenter of a set of points (here: WoS SCs) with associated weights (here: number of publications) is defined as the point $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2)$, where

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

Here, Lj,1 and Lj,2 are the horizontal and vertical coordinates of WoS SC j on the map, $\mathrm{m_j}$ is the number of publications in WoS SC j, and $\mathrm{T} = \sum_{\mathrm{j=1}}^{\mathrm{N}} \mathrm{m_j}$ is the total number of publications of the entity (panel member, research group). Note that T is larger than the total number of publications as we use full counting of WoS SCs: if a publication appears in a journal belonging to two categories, it will be counted twice. For further elaboration on the barycenter method, we refer to (Rousseau, 1989; Jin & Rousseau, 2001; Verleysen & Engels, 2013, 2014).

Formula (3) is implemented in a Python script 'barycenter-categories.py' (the actual barycenter calculation is done in the barycenter function, see section 5) that takes as input the map file ('WoS SC_VOS_map.xlsx',Figure 10) and the weights (number of publications) per WoS SC ('Physics research groups and panel_WoS SCs.xlsx', Figure 1), and calculates a barycenter for each entity (Figure 12). We run the program as:

```
python barycenter-categories.py "WoS SC_VOS_map.xlsx" "Physics research groups and panel_WoS SCs.xlsx"
```

This program calculates the barycenter and generates an output file 'Physics research groups and panel_WoS SCs-barycenter.xlsx'. Figure 12 shows the barycenter coordinates of the Physics individual research groups, panel members, research groups together and panel.

	A	B	C
1		**x**	**y**
2	**Groups**	0.853168	-0.18649
3	**PHYS-A**	-0.03281	0.235005
4	**PHYS-B**	0.987935	-0.27128
5	**PHYS-C**	0.983647	-0.17754
6	**PHYS-D**	0.94844	-0.37747
7	**PHYS-E**	0.937342	-0.24696
8	**PHYS-F**	0.753423	-0.3563
9	**PHYS-G**	0.7431	0.054318
10	**PHYS-H**	1.001275	-0.27801
11	**PHYS-I**	0.412145	-0.01438
12	**PM1**	0.958864	-0.39068
13	**PM2**	1.054279	-0.26053
14	**PM3**	0.849417	-0.31748
15	**PM4**	0.924517	-0.11581
16	**PM5**	1.003994	-0.22824
17	**PM6**	1.014333	-0.25591
18	**Panel**	0.969629	-0.25405

Figure 12. Barycenter coordinates of the Physics individual research groups, panel members, research groups together and panel using the WoS SCs VOS map

b) Euclidean distance between barycenters

Subsequently, we determine the Euclidean distance between the barycenters of different entities: individual research groups, research groups together, panel members and panel. The Euclidean distance between points $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2)$ and $\mathrm{D} = (\mathrm{D}_1, \mathrm{D}_2)$ is calculated as follows:

$$d = \sqrt{(C_1 - D_1)^2 + (C_2 - D_2)^2}. \tag{4}$$

We use the implementation of Euclidean distance in scipy.spatial.dist. We note that the Python script 'barycenter-categories.py' executes both formula (3) and (4). The distances thus obtained should be interpreted as having arbitrary units on a ratio scale (Egghe & Rousseau, 1990). This means that there is a fixed meaningful zero (distance zero in the map), and distances can be compared in terms of percentage or fraction (e.g. the distance between A and B is 1.5 times larger than the distance between C and D).

	A	B	C	D	E	F	G	H	I	J	K	L
1		**Groups**	**PHYS-A**	**PHYS-B**	**PHYS-C**	**PHYS-D**	**PHYS-E**	**PHYS-F**	**PHYS-G**	**PHYS-H**	**PHYS-I**	**PM1**
2	**Groups**	0	0.98113	0.159217	0.130786	0.213423	0.103641	0.196936	0.264774	0.174099	0.473418	0.229919
3	**PHYS-A**	0.981128	0	1.1394	1.09698	1.156707	1.083272	0.983767	0.796666	1.154342	0.510073	1.172556
4	**PHYS-B**	0.159217	1.1394	0	0.093838	0.113303	0.056133	0.24945	0.407377	0.014943	0.6305	0.12289
5	**PHYS-C**	0.130786	1.09698	0.093838	0	0.203012	0.08345	0.291479	0.334095	0.102008	0.594335	0.214577
6	**PHYS-D**	0.213423	1.15671	0.113303	0.203012	0	0.130982	0.196163	0.478129	0.112625	0.647648	0.016824
7	**PHYS-E**	0.103641	1.08327	0.056133	0.08345	0.130982	0	0.213966	0.358468	0.071073	0.574392	0.145319
8	**PHYS-F**	0.196936	0.98377	0.24945	0.291479	0.196163	0.213966	0	0.410749	0.259923	0.483095	0.208298
9	**PHYS-G**	0.264774	0.79667	0.407377	0.334095	0.478129	0.358468	0.410749	0	0.420828	0.33801	0.494546
10	**PHYS-H**	0.174099	1.15434	0.014943	0.102008	0.112625	0.071073	0.259923	0.420828	0	0.645426	0.120386
11	**PHYS-I**	0.473418	0.51007	0.6305	0.594335	0.647648	0.574392	0.483095	0.33801	0.645426	0	0.663703
12	**PM1**	0.229919	1.17256	0.12289	0.214577	0.016824	0.145319	0.208298	0.494546	0.120386	0.663703	0
13	**PM2**	0.214306	1.1947	0.067208	0.108982	0.157725	0.117722	0.315731	0.442677	0.055812	0.687697	0.161375
14	**PM3**	0.13104	1.04094	0.146021	0.193912	0.115778	0.112711	0.103547	0.386701	0.156904	0.532049	0.131669
15	**PM4**	0.100435	1.01958	0.167906	0.08548	0.262756	0.131779	0.295145	0.248706	0.179447	0.522315	0.277008
16	**PM5**	0.156495	1.13558	0.045939	0.054629	0.159242	0.069232	0.281401	0.38458	0.049849	0.6293	0.168595
17	**PM6**	0.175477	1.1565	0.030545	0.084165	0.138274	0.077509	0.279558	0.412077	0.025671	0.648819	0.145738
18	**Panel**	0.134635	1.11537	0.025139	0.077784	0.125231	0.033056	0.239168	0.382628	0.039694	0.606819	0.137054

Figure 13. Excerpt of Euclidean distances matrix of barycenter of the Physics individual research groups, panel members, research groups together and panel using WoS SCs VOS map

From the matrix of Euclidean distances, which includes distances between all entity pairs (Figure 13), we extract Table 3, containing only the distances between the research groups and research groups together on the one hand and the panel and panel members on the other, for the convenience of analysis.

Table 3. Euclidean distances between barycenter of Physics individual research groups, panel members, research groups together and panel using WoS SCs VOS map

	Group	PHYS-A	PHYS- B	PHYS-C	PHYS- D	PHYS-E	PHYS- F	PHYS- G	PHYS- H	PHYS-I
Panel	0.135	1.115	0.025	0.078	0.125	0.033	0.239	0.383	0.040	0.607
PM 1	0.230	1.173	0.123	0.215	**<u>0.017</u>**	0.145	**0.208**	0.495	0.120	0.664
PM 2	0.214	1.195	0.067	0.109	0.158	0.118	0.316	0.443	0.056	0.688
PM 3	0.131	**1.041**	0.146	0.194	0.116	0.113	**<u>0.104</u>**	**0.387**	0.157	**0.532**
PM 4	0.100	**<u>1.020</u>**	0.168	**0.085**	0.263	0.132	0.295	**<u>0.249</u>**	0.179	**<u>0.522</u>**
PM 5	0.156	**1.136**	**0.046**	**<u>0.055</u>**	0.159	**<u>0.069</u>**	0.281	**0.385**	**0.050**	**0.629**
PM 6	0.175	**1.157**	**<u>0.031</u>**	0.084	0.138	**0.078**	0.280	**0.412**	**<u>0.026</u>**	0.649

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.232 (SD 0.337). The number in the row of 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).

In Table 3, for each research group we find the shortest distance to one of the panel members, and underline and bold it. In addition, the average and standard deviation of the shortest distances are calculated. The confidence intervals (discussed in section 2.5) are included through the typography of the values.

c) Barycenter overlay map

We take the 'WoS SC_VOS map.xlsx' (Figure 10) file and manually input the Physics individual groups, panel members, research groups together and panel's coordinates (shown in Figure 12) after the 224 WoS SCs. We fill up the 'weight' column with 20 (we can put other numbers too) to highlight the size of the bubble.

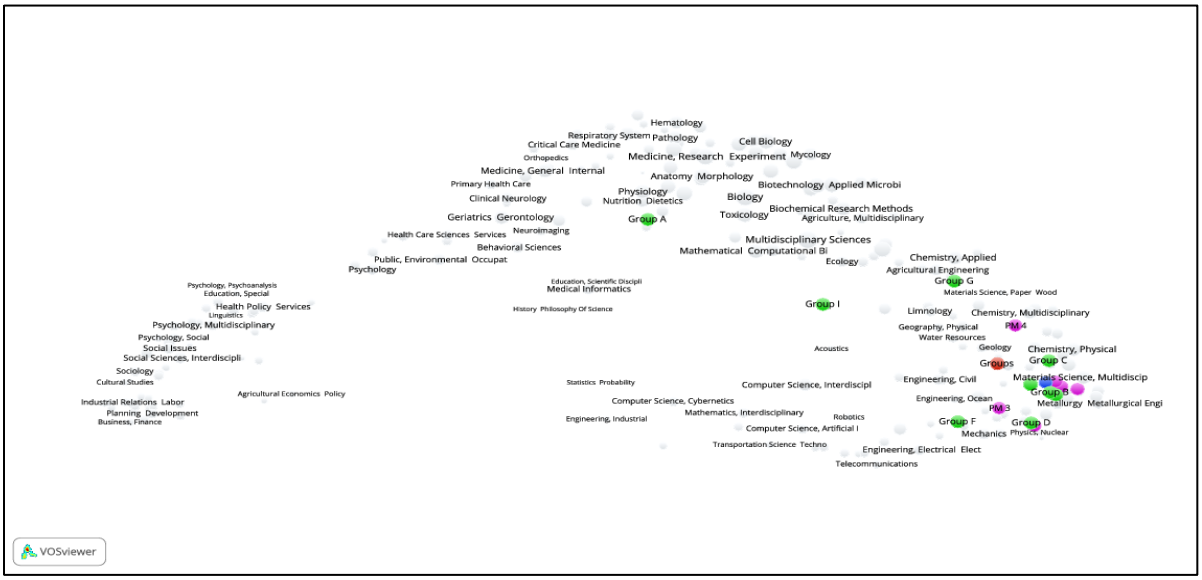


Figure 14. Barycenter overlay map of Physics individual research groups, panel members (PM), research groups together and panel in WoS SCs

Figure 15. Barycenter map of Physics individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed)

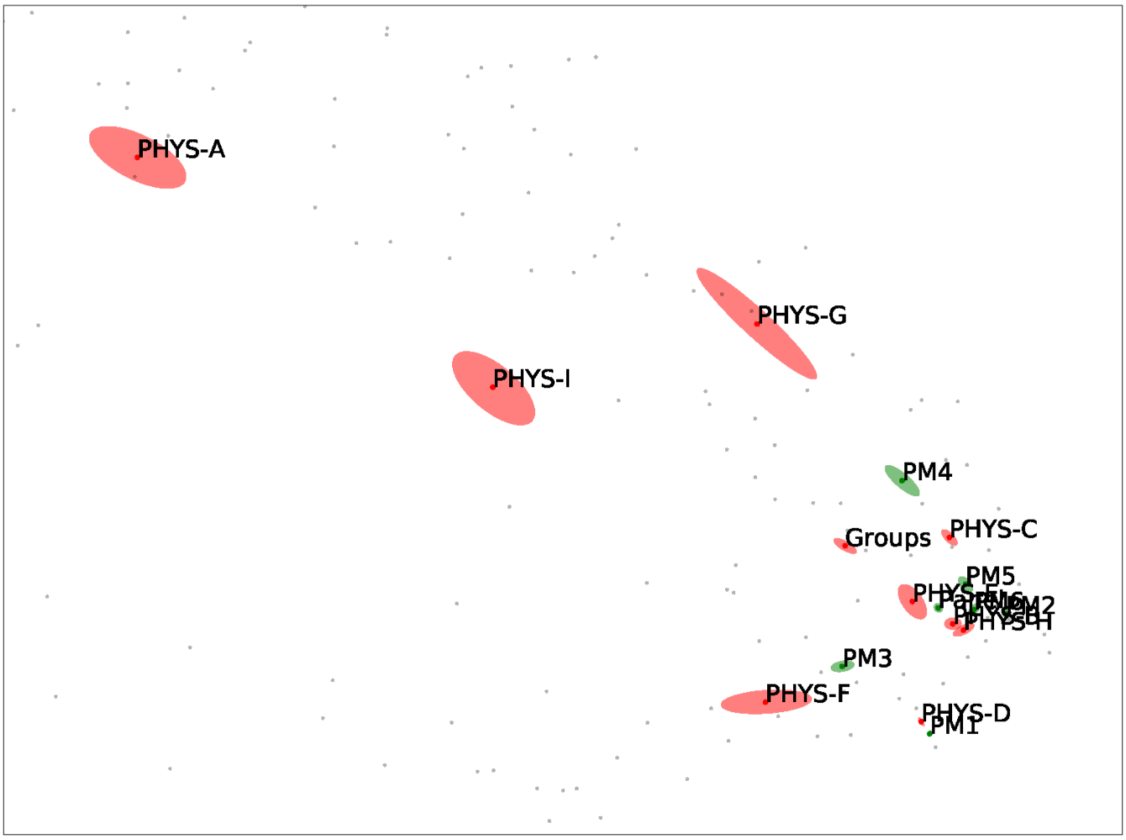


Figure 16. Barycenter overlay map of Physics panel, panel members (PM), research groups and research groups together (groups) with their confidence regions

In the ‘cluster’ column, we assign 1 to all the 224 WoS SCS, 2 to the research groups together, 3 to all research groups, 4 to the panel, and 5 to individual panel members. We save the file as ‘Barycenter overlap map of Physics department.csv’. After that, we open the file with VOSviewer to visualize the barycenters (Figure 14). Figure 15 shows a zoomed in version of Figure 14.

We also create the barycenter overlap map of Physics department and include the confidence regions of the respective barycenter of panel, panel members (PM), research groups and research groups together using the WoS SCs VOS map (Figure 16). The bootstrap replications of barycenters are also used to add a 95% confidence region for each barycenter to the maps. For each barycenter, we have a cloud of 1000 points (bootstrapped barycenters) surrounding it. The confidence region is an ellipse that covers 95% of the bootstrapped barycenters. The larger the confidence region, the less stable the barycenter is. Although the CI of the distance between two barycenters and their confidence regions are related, the two should not be conflated. In particular, we stress that overlapping confidence regions as seen in Figure 16 (figure with overlapping regions in it) does not correspond to overlap between CIs for distances.

The maps were plotted using Matplotlib (http://matplotlib.org). First, the base map was plotted using the pre-existing coordinates. Next, the barycenters were added as slightly larger red or green points. Finally, a partially translucent confidence region (ellipse) was calculated and superimposed on the map. Calculation of the confidence region was done using an implementation by Kington (2014). We briefly outline what elements determine the location and placement of such a confidence ellipse. The center of the ellipse is simply the mean of all bootstrapped barycenters. The width and height of the ellipse (or its axes) depend on the variance in the cloud of points. Finally, the orientation of the ellipse is obtained from the largest eigenvector.

2.7 Similarity-adapted publication vector method

a) Similarity-adapted publication vector calculation

A similarity-adapted publication vector (SAPV) is determined as the vector $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_\mathrm{N})$, where:

$$C_k = \frac{\sum_{j=1}^{N} s_{kj} m_j}{\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} m_j} \tag{5}$$

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. The numerator of formula (5) is equal to the k-th element of $\mathrm{S} * \mathrm{M}$, the multiplication of the similarity matrix S and the column matrix of publications $\mathrm{M} = \left(\mathrm{m}_\mathrm{j}\right)_\mathrm{j}$. The denominator is the L1-norm of the unnormalized vector.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Acoustics	al Econom	tural Engir	Dairy Ani	re, Multidi	Agronomy	Allergy	my Morpł	Andrology	esthesiolo	nthropolog	rea Studie	omy Astro
2	Acoustics	1	0	0	0	0	0	0	0	0	0	0	0	0
3	Agricultural Economics Policy	0	1	0	0	0	0	0	0	0	0	0	0.185	0
4	Agricultural Engineering	0	0	1	0.16	0.445	0.319	0	0	0	0	0	0	0
5	Agriculture, Dairy Animal Science	0	0	0.16	1	0.413	0	0	0.184	0.247	0	0	0	0
6	Agriculture, Multidisciplinary	0	0	0.445	0.413	1	0.615	0	0.159	0	0	0	0	0
7	Agronomy	0	0	0.319	0	0.615	1	0	0	0	0	0	0	0
8	Allergy	0	0	0	0	0	0	1	0	0	0	0	0	0
9	Anatomy Morphology	0	0	0	0.184	0.159	0	0	1	0.478	0.23	0	0	0
10	Andrology	0	0	0	0.247	0	0	0	0.478	1	0	0	0	0
11	Anesthesiology	0	0	0	0	0	0	0	0.23	0	1	0	0	0
12	Anthropology	0	0	0	0	0	0	0	0	0	0	1	0.277	0
13	Area Studies	0	0.185	0	0	0	0	0	0	0	0	0.277	1	0
14	Astronomy Astrophysics	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 17. Excerpt of WoS SCs similarity matrix

We take the 'map10.net' file (see section 2.3) and with a Python script, we transform the network back into the adjacency matrix and save it as 'WoS SCs similarity matrix.xlsx' (Figure 17).

A python script 'sa-vector-categories.py' is used that takes as input the WoS SCs similarity matrix (Figure 17) and the number of publications of Physics individual research groups and panel members per WoS SC (weights) (Figure 1), and calculates SAPVs for all entities. The calculation of SAPVs is carried out by the sa_vector function, (see section 5). We run the program as:

```
python sa-vector-categories.py "WoS SC_similarity matrix.xlsx" "Physics research groups and Panel_WoS SCs.xlsx"
```

This program calculates the SAPV of each entity and stores the result in an output file named ‘Physics research groups and panel_WoS SCs-sa-vectors.xlsx’ (Figure 18).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Acoustics	Agricultur	Agricultur	Agricultur	Agricultur	Agronomy	Allergy	Anatomy	Andrology	Anesthesi	Anthropol	Area Studi
2	Groups	0.001073	1.81E-05	0.002113	0.000276	0.001042	0.000364	0.000176	0.003418	0.001503	0.001156	0	0
3	PHYS-A	0.002697	0	0.002257	0.000969	0.002151	0.001298	0.001369	0.019731	0.005706	0.013565	0	0
4	PHYS-B	0.000863	1.9E-05	0.000447	3.1E-05	8.25E-05	1.46E-05	0	0.000345	0.00016	4.91E-05	0	0
5	PHYS-C	0.000306	1.09E-05	0.003196	0.00019	0.001241	0.000239	2.91E-05	0.002353	0.001432	3.68E-05	0	0
6	PHYS-D	0.001217	0	0.000353	0	0.000374	0	0	0.0003	0	0	0	0
7	PHYS-E	0.001197	0	0.002397	0.000343	0.00135	0.000637	0.00013	0.001675	0.000903	0.000112	0	0
8	PHYS-F	0.002828	0.00051	0.00075	0	0	0	0	0.000392	0	0	0	0
9	PHYS-G	0.000235	0	0.007149	0.003044	0.005268	0.003431	0.00147	0.010962	0.006823	0.000121	0	0
10	PHYS-H	0.000644	0	0	0	0	0	0	0	0	0	0	0
11	PHYS-I	0.004632	0	0.00276	0.000642	0.002862	0.000803	0.00046	0.014142	0.006444	0.004144	0	0
12	PM1	0	0	0	0	0	0	0	0	0	0	0	0
13	PM2	0	0	0.000428	0	0	0	0	0	0	0	0	0
14	PM3	0.003283	0	0.000125	0	3.31E-05	0	0	0.000202	7.54E-05	0	0	0
15	PM4	0.000339	0	0.005858	0.000579	0.002471	0.000521	0.000124	0.002883	0.001332	0.000158	0	0
16	PM5	0.00032	0	0.001823	2.04E-05	0.00047	1.83E-05	0	0.00129	0.000815	0	0	0
17	PM6	0.000363	0	0.001865	0	0	0	0	0	0	0	0	0
18	Panel	0.000728	0	0.00197	9.18E-05	0.000533	8.25E-05	1.81E-05	0.000911	0.000496	2.3E-05	0	0

Figure 18. Excerpt of SAPV of the Physics individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix

b) Euclidean distance between similarity-adapted publication vectors

Subsequently, we determine the Euclidean distances between different entities SAPV: individual research groups, research groups together, panel members, and panel. The Euclidean distance between vectors a and b in $\mathbf{R}^N$ is:

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + \cdots + (a_N - b_N)^2} \qquad (6)$$

Again, we use the implementation of Euclidean distance in scipy.spatial.dist. We note that the python script ‘sa-vector-categories.py’ executes both formula (5) and (6), and calculates Euclidean distances between the SAPV in an output file ‘Physics research groups and panel_WoS SCs-sa-vectors.xlsx’ (Figure 19).

From the calculated matrix of pairwise Euclidean distances between SAPVs of Physics groups, panel members, groups together, and panel together (Figure 18) we extract Table 4 containing only the distances between the research groups and groups together on the one hand and the panel and panel members on the other, for the convenience of analysis. In Table 4, for each research group we find the shortest distance to one of the panel members, and underline and bold it. In addition, the average and standard deviation of the shortest distances are calculated. The confidence intervals (discussed in section 2.5) are included through the typography of the values.

	A	B	C	D	E	F	G	H	I	J	K	L
1		**Groups**	**PHYS-A**	**PHYS-B**	**PHYS-C**	**PHYS-D**	**PHYS-E**	**PHYS-F**	**PHYS-G**	**PHYS-H**	**PHYS-I**	**PM1**
2	**Groups**	0	0.135329	0.03191	0.027238	0.255711	0.028371	0.108465	0.077858	0.036881	0.093634	0.35271
3	**PHYS-A**	0.13533	0	0.162943	0.149541	0.289291	0.147343	0.172859	0.101063	0.166924	0.06485	0.376014
4	**PHYS-B**	0.03191	0.162943	0	0.036832	0.260513	0.039741	0.10938	0.10666	0.007931	0.12202	0.357612
5	**PHYS-C**	0.02724	0.149541	0.036832	0	0.276113	0.047625	0.131094	0.083228	0.041152	0.107765	0.372638
6	**PHYS-D**	0.25571	0.289291	0.260513	0.276113	0	0.230414	0.206556	0.27827	0.260627	0.275469	0.098003
7	**PHYS-E**	0.02837	0.147343	0.039741	0.047625	0.230414	0	0.096422	0.08915	0.04262	0.106878	0.327559
8	**PHYS-F**	0.10847	0.172859	0.10938	0.131094	0.206556	0.096422	0	0.14604	0.111309	0.138997	0.300876
9	**PHYS-G**	0.07786	0.101063	0.10666	0.083228	0.27827	0.08915	0.14604	0	0.110979	0.061791	0.371345
10	**PHYS-H**	0.03688	0.166924	0.007931	0.041152	0.260627	0.04262	0.111309	0.110979	0	0.12624	0.357684
11	**PHYS-I**	0.09363	0.06485	0.12202	0.107765	0.275469	0.106878	0.138997	0.061791	0.12624	0	0.36742
12	**PM1**	0.35271	0.376014	0.357612	0.372638	0.098003	0.327559	0.300876	0.371345	0.357684	0.36742	0
13	**PM2**	0.04383	0.172467	0.019347	0.037744	0.272327	0.053698	0.126851	0.114513	0.019078	0.13322	0.368779
14	**PM3**	0.06572	0.156281	0.065015	0.080346	0.25611	0.069077	0.099936	0.116267	0.06299	0.110929	0.352828
15	**PM4**	0.04226	0.143667	0.059887	0.038917	0.270561	0.051294	0.128743	0.066449	0.062785	0.103049	0.366666
16	**PM5**	0.02712	0.156852	0.022943	0.015968	0.271472	0.044208	0.125088	0.09524	0.026859	0.115292	0.368259
17	**PM6**	0.03239	0.164976	0.011569	0.03492	0.258225	0.036654	0.11141	0.105507	0.014582	0.125042	0.355148
18	**Panel**	0.02053	0.154099	0.017909	0.029688	0.254839	0.027826	0.108848	0.094183	0.020516	0.111795	0.352044

Figure 19. Excerpt of pairwise Euclidean distance matrix between SAPVs of Physics individual research groups, panel members, research groups together and panel together using WoS SCs similarity matrix

Table 4. Euclidean distances between SAPVs of Physics individual groups, panel members, research groups together and panel in WoS SCs similarity matrix

	Groups	PHYS-A	PHYS- B	PHYS-C	PHYS- D	PHYS-E	PHYS- F	PHYS- G	PHYS- H	PHYS-I
Panel	0.021	0.154	0.018	0.030	0.255	0.028	0.109	0.094	0.021	0.112
PM 1	0.353	0.376	0.358	0.373	<u>**0.098**</u>	0.328	0.301	0.371	0.358	0.367
PM 2	0.044	0.172	0.019	0.038	0.272	0.054	0.127	0.115	**0.019**	0.133
PM 3	0.066	**0.156**	0.065	0.080	0.256	0.069	<u>**0.100**</u>	0.116	0.063	**0.111**
PM 4	0.042	<u>**0.144**</u>	0.060	0.039	0.271	0.051	0.129	<u>**0.066**</u>	0.063	<u>**0.103**</u>
PM 5	0.027	**0.157**	0.023	<u>**0.016**</u>	0.271	**0.044**	0.125	**0.095**	0.027	**0.115**
PM 6	0.032	**0.165**	<u>**0.012**</u>	0.035	0.258	<u>**0.037**</u>	**0.111**	**0.106**	<u>**0.015**</u>	0.125

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.07 (SD 0.05). The number in the row of 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).

c) Similarity-adapted publication vector overlay map

Results of the SAPV approach cannot be visualized easily since an SAPV has N coordinates itself. However, visualization is possible if one expands the similarity matrix with one extra row and column, containing the SAPV's coordinates. The expanded $(N + 1) \times (N + 1)$ matrix can then be visualized using, for instance, VOSviewer. Note that this approach works well for visualizing the location of one SAPV but cannot be used for multiple SAPVs at the same time, for two reasons:

- Adding extra rows/columns affects the layout algorithm and may distort the original base map. The effect of one extra point turns out to be negligible.
- It is unclear what similarity score should be assigned to two SAPVs.

A	B	C	D	E	F
id	label	x	y	weight	cluster
205	Engineering, Petroleum	-0.981	0.0182	9.6043	1
206	Computer Science, Theory M	-0.6484	-0.253	10.6667	1
207	Physics, Multidisciplinary	-1.0551	-0.2236	12.1876	1
208	Law	1.4546	-0.1087	0.7	1
209	Materials Science, Ceramics	-1.0136	-0.1327	16.3173	1
210	PHYS-B	-0.9847	-0.1641	20	3
211	International Relations	1.4859	-0.1487	6.709	1
212	Mathematics, Interdisciplinai	-0.5099	-0.2145	13.1301	1
213	Agriculture, Multidisciplinary	-0.4731	0.1654	15.4221	1
214	Engineering, Chemical	-0.9481	-0.0612	16.8237	1
215	Optics	-0.9555	-0.2139	10.7744	1
216	Rehabilitation	0.6987	0.0786	11.2431	1
217	Microscopy	-0.4965	0.0472	30.5151	1
218	Cell Biology	-0.1134	0.1974	28.0914	1
219	Meteorology Atmospheric S	-0.9213	0.0643	7.582	1
220	Engineering, Marine	-0.8787	-0.1243	7.3078	1
221	Allergy	0.1116	0.2355	8.034	1

Figure 20. Excerpt of PHYS-B.csv file

We determine SAPVs of all entities (Figure 18). We take the WoS SCs similarity matrix Excel file (Figure 17). We copy PHYS-B's SAPV and paste at the bottom row and last column of the matrix file, thereby expanding the matrix to dimensions $(N+1) \times (N+1)$. We save the file as 'PHYS-B_similarity matrix.xlsx'. A python script 'excel2network.py' is used to convert '[Research group code]_similarity matrix.xlsx' files to Pajek network files (which can then be used in Pajek or VOSviewer). We run the program as:

```
python excel2network.py "PHYS-B_similarity matrix.xlsx" Sheet1
```

This program yields an output file named 'PHYS-B_similarity matrix.net'. We create a map based on the network file using VOSviewer. It is not possible to easily locate PHYS-B in the map due to many different cluster colors. Therefore, we save the map data as 'PHYS-B.txt' file. In the text file, we can identify PHYS-B, but cannot easily change cluster number of all the WoS SCs in the file that is necessary to highlight the PHYS-B's location in the overlay map. Therefore, we import the data from the 'PHYS-B.txt' file to 'PHYS-B.xlsx' file.

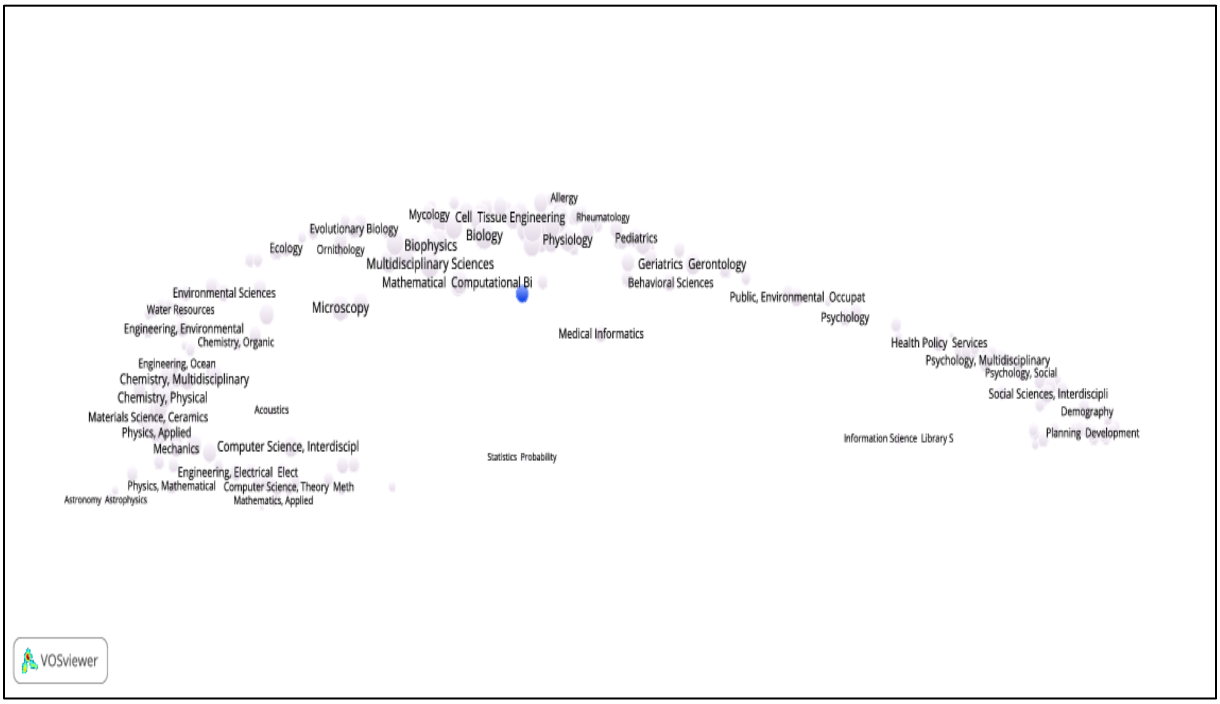


Figure 21. Location of the SAPV of PHYS-B in the WoS SCs similarity matrix

In the 'PHYS-B.xlsx' file we first identify the PHYS-B label and assigned 20 (we can put other numbers too) for weight. In the cluster column, we assign 1 to all the WoS SCs and 2 to PHYS-B and save as CSV file (Figure 20). We open the file with VOSviewer to visualize the SAPV 'location' of PHYS-B (Figure 21).

We repeat the above-mentioned process to create separate maps for each research group, each panel member, research groups together and panel (see Appendix B).

2.8 Weighted cosine similarity method

We consider a weighted similarity method (generalized cosine similarity). The weighted similarity between panel member (PM) k and research group m, according to Zhou et al. (2012) is:

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

$$= \frac{\left(M^k\right)^t * S * R^m}{\sqrt{\left(M^k\right)^t * S * M^k} \cdot \sqrt{\left(R^m\right)^t * S * R^m}} \tag{7}$$

The numerator is nothing but the matrix multiplication: $\left(M^k\right)^t * S * R^m$, where ^{t} denotes matrix transposition, S is the WoS SCs 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 similarity between panel member k and research group m.

This value is calculated for each panel member and each research group. Weighted cosine similarity (WCS) is implemented in Python as a fairly straightforward set of matrix operations (see section 5, weighted_cosine). A python script 'cosine-categories.py' is used that takes as input the similarity matrix ('WoS SCs_similarity matrix.xlsx', see Figure 17) and the weights (number of publications) per WoS SC ('Physics research groups and panel_WoS SCs.xlsx', see Figure 1), and calculates the weighted cosine similarity between all entities. We run the program as:

```
python cosine-categories.py "WoS SC_similarity matrix.xlsx" "Physics research groups and panel_WoS SCs.xlsx"
```

This program calculates the WCS value in an output file as 'Physics research groups and panel_WoS SCs-cosine.xlsx' (Figure 22). From the calculated WCS value matrix (Figure 22), we extract Table 5 containing only the WCS value of the research groups and groups on the one hand and the panel and panel members on the other, for the convenience of analysis. The confidence intervals (discussed in section 2.5) are included through the typography of the values.

Since our barycenter method (see section 2.6) and and SAPV method (see section 2.7) are distance-based rather than similarity-based, we use 1 – WCS as values to obtain dissimilarity values: weighted cosine dissimilarity (WCD) in Table 6, which can more easily be compared with the other two approaches. For the sake of simplicity, the results are shown under the WCS method.

	A	B	C	D	E	F	G	H	I	J	K	L
1		**Groups**	**PHYS-A**	**PHYS-B**	**PHYS-C**	**PHYS-D**	**PHYS-E**	**PHYS-F**	**PHYS-G**	**PHYS-H**	**PHYS-I**	**PM1**
2	**Groups**	1	0.263602	0.971727	0.954617	0.317642	0.93174	0.618074	0.687645	0.965325	0.62309	0.249702
3	**PHYS-A**	0.263602	1	0.173253	0.185799	0.045866	0.183934	0.116029	0.277396	0.176108	0.551504	0.029549
4	**PHYS-B**	0.971727	0.173253	1	0.921883	0.224681	0.871926	0.628511	0.57281	0.994947	0.539098	0.155487
5	**PHYS-C**	0.954617	0.185799	0.921883	1	0.106107	0.82342	0.405778	0.732521	0.911419	0.574115	0.042705
6	**PHYS-D**	0.317642	0.045866	0.224681	0.106107	1	0.620935	0.562135	0.07742	0.22355	0.099295	0.996124
7	**PHYS-E**	0.93174	0.183934	0.871926	0.82342	0.620935	1	0.713424	0.630793	0.869488	0.517694	0.56127
8	**PHYS-F**	0.618074	0.116029	0.628511	0.405778	0.562135	0.713424	1	0.304448	0.617536	0.359369	0.507608
9	**PHYS-G**	0.687645	0.277396	0.57281	0.732521	0.07742	0.630793	0.304448	1	0.573189	0.603912	0.028129
10	**PHYS-H**	0.965325	0.176108	0.994947	0.911419	0.22355	0.869488	0.617536	0.573189	1	0.532573	0.153552
11	**PHYS-I**	0.62309	0.551504	0.539098	0.574115	0.099295	0.517694	0.359369	0.603912	0.532573	1	0.051584
12	**PM1**	0.249702	0.029549	0.155487	0.042705	0.996124	0.56127	0.507608	0.028129	0.153552	0.051584	1
13	**PM2**	0.939156	0.150703	0.982228	0.919711	0.127364	0.806204	0.513078	0.542599	0.976502	0.497294	0.062642
14	**PM3**	0.711592	0.220492	0.713912	0.625419	0.21089	0.667682	0.5263	0.440036	0.761591	0.544413	0.146249
15	**PM4**	0.804006	0.182447	0.728717	0.828958	0.128813	0.756601	0.436357	0.894987	0.740867	0.478563	0.066501
16	**PM5**	0.973918	0.182461	0.964627	0.98598	0.157677	0.851858	0.4754	0.655897	0.957439	0.567214	0.091726
17	**PM6**	0.979405	0.163983	0.989279	0.92974	0.272087	0.902815	0.64274	0.630886	0.984688	0.516068	0.203139
18	**Panel**	0.987941	0.195962	0.969848	0.938837	0.324254	0.933496	0.618027	0.673416	0.975915	0.569795	0.255016

Figure 22. Excerpt of WCS value matrix of the Physics individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix

Table 5. WCS value of Physics individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.988	0.196	0.970	0.939	0.324	0.933	0.618	0.673	0.976	0.570
PM1	0.250	0.030	0.155	0.043	**0.996**	0.561	0.508	0.028	0.154	0.052
PM2	0.939	**0.151**	**0.982**	0.920	0.127	0.806	0.513	0.543	**0.977**	**0.497**
PM3	0.712	**0.220**	0.714	0.625	0.211	0.668	0.526	0.440	0.762	**0.544**
PM4	0.804	**0.182**	0.729	0.829	0.129	0.757	0.436	**0.895**	0.741	**0.479**
PM5	0.974	**0.182**	0.965	**0.986**	0.158	**0.852**	0.475	0.656	0.957	**0.567**
PM6	0.979	**0.164**	**0.989**	0.930	0.272	**0.903**	**0.643**	0.631	**0.985**	**0.516**

For each research group we determine the panel member at the highest similarity. The number in the row corresponding to this panel member is indicated in bold and underlined. Similarities whose confidence intervals overlap with that of the highest similarities are in bold (same column).

Table 6. WCD value between Physics individual research groups, panel members, groups and panel using WoS SCs similarity matrix

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.012	0.804	0.030	0.061	0.676	0.067	0.382	0.327	0.024	0.430
PM1	0.750	0.970	0.845	0.957	**0.004**	0.439	0.492	0.972	0.846	0.948
PM2	0.061	**0.849**	**0.018**	0.080	0.873	0.194	0.487	0.457	**0.023**	**0.503**
PM3	0.288	**0.780**	0.286	0.375	0.789	0.332	0.474	0.56	0.238	**0.456**
PM4	0.196	**0.818**	0.271	0.171	0.871	0.243	0.564	**0.105**	0.259	**0.521**
PM5	0.026	**0.818**	0.035	**0.014**	0.842	**0.148**	0.525	0.344	0.043	**0.433**
PM6	0.021	**0.836**	**0.011**	0.070	0.728	**0.097**	**0.357**	0.369	**0.015**	**0.484**

The lowest similarity between a group and a panel member is underlined and printed in bold.

Table 7. Pearson and Spearman correlation between three methods using data from Physics individual research groups and panel members

Pearson / Spearman	Barycenter	SAPV	WCS
Barycenter	1.00	0.29 (0.87)	0.60 (0.89)
SAPV	0.64 (0.94)	1.00	0.86 (0.97)
WCS	0.71 (0.91)	0.94 (0.97)	1.00

In Table 7, the upper triangle refers to Pearson's correlations while the lower triangle refers to Spearman's correlation values. Values between brackets are correlations calculated after removal of PHYS-D and PM1. Table 7 and Figure 23 shows that all correlations are high or moderately high except the Pearson's correlation between the barycenter method and SAPV in the case of Physics. For the sake of simplicity, the results are shown under the WCS method.

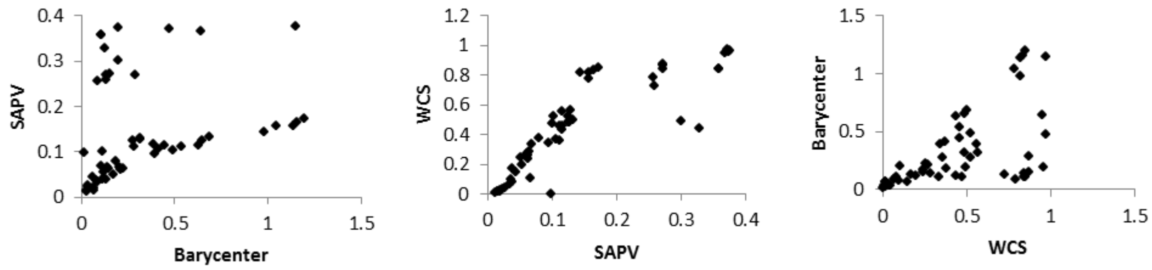


Figure 23. Scatter plots of the Euclidean distances between barycenter and SAPV of individual research groups and panel members, and WCS between individual research groups in the Physics department

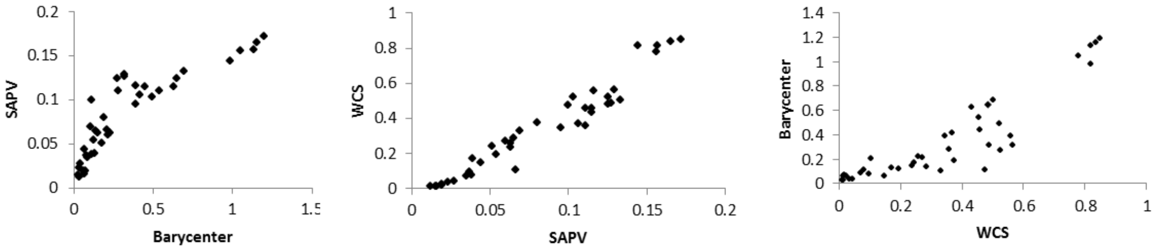


Figure 24. Scatter plots of the of correlation between barycenter, SAPV and WCS methods in WoS SCs in Physics department excluding PHYS-D and PM1

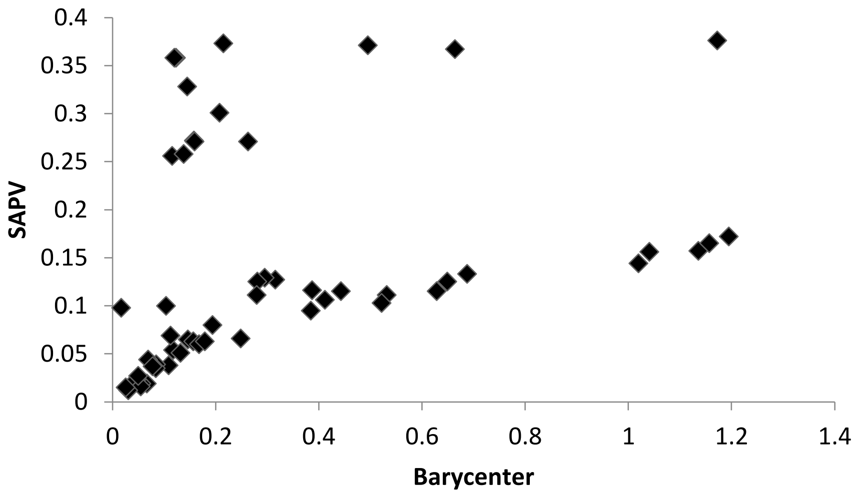


Figure 25. Scatter plot of the cognitive distances between individual research groups and panel members for the barycenter and SAPV methods in the Physics department

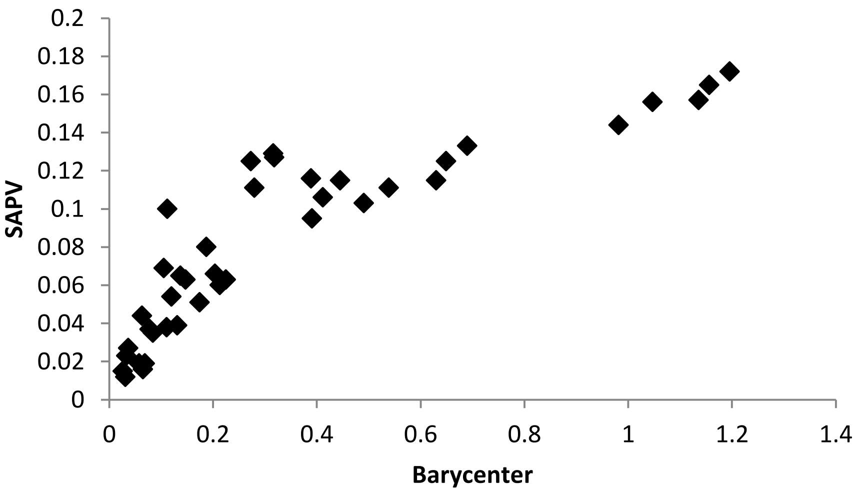


Figure 26. Scatter plot of the cognitive distances between individual research groups and panel members for the barycenter and SAPV methods in the Physics department excluding PHYS-D and PM1

Figure 25 illustrates what happened. This low Pearson's correlation is due to the 13 points (including two times two points that overlap) in the upper half of Figure 25. All these points correspond to distances involving either research group PHYS-D or PM1 (but not both). This group and 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 WoS SCs.

Figure 26 presents the same data as Figure 25, but leaves out distances involving PHYS-D and PM1. In this case, $r = 0.87$ and $\rho = 0.94$. These values can also be seen in Table 7 and Figure 26, where all values between brackets refer to correlations calculated without PHYS-D and PM1.

3 Cognitive distance based on journals

3.1 Data collection process

For collecting journal data, after the search result (see section 2.1) we use the 'Analyze Results' option in the WoS, and rank the record by Source title (hereafter journal title) and set the minimum record count (threshold) to one. We repeat this procedure for each of the research groups and panel members. We save the record as 'analyze.txt' and subsequently rename the file to '[Research group code]_ journals title.txt', for example 'PHYS-B_journals title.txt'. For panel members we rename to '[Panel member code]_ journals title.txt', for example 'PM2_ journals title.txt'.

We combine the search sets for each research group and panel member from the search history of the WoS, and get the result for the research groups as a whole and the panel. In this way, any publication that has been co-authored by members of two or more research groups or by two or more panel members is counted only once. We save the resulting list as 'analyze.txt' and save a copy of the file as 'Groups together_journals titles.txt' for the groups as a whole, and as 'Panel_journals title.txt' for the panel.

All downloaded data files are exported to an MS Excel file. The downloaded data files, ‘[Research group code]_journals title.txt’, ‘[PM code]_journals title.txt’, ‘Groups together_ journals title.txt’ and ‘Panel_journals title.txt’ have been exported to an MS Excel file. The sheets in the Excel file contain data on and are named after the research groups’ code names (PHYS-A, PHYS-B, PHYS-C, etc.), the panel members’ code names, (PM1, PM2, PM3, etc.), Panel members together and Groups together The Excel file is saved as ‘Physics research groups and panel_ journals title.xlsx’ (Figure 27). Publication statistics for each research groups and panel members have shown in the Table 1 and Table 2 respectively.

	A	B	C
1	**Source Titles**	**records**	**% of 1732**
2	PHYSICAL REVIEW B	291	16.801
3	EUROPEAN PHYSICAL JOURNAL C	108	6.236
4	PHYSICS LETTERS B	72	4.157
5	PHYSICAL REVIEW LETTERS	60	3.464
6	PHYSICA C SUPERCONDUCTIVITY AND ITS APPL	43	2.483
7	ULTRAMICROSCOPY	42	2.425
8	APPLIED PHYSICS LETTERS	37	2.136
9	CHEMISTRY OF MATERIALS	31	1.79
10	PHYSICAL REVIEW D	30	1.732
11	PHYSICA E LOW DIMENSIONAL SYSTEMS NANO	30	1.732
12	PHYSICAL REVIEW E	29	1.674
13	JOURNAL OF SOLID STATE CHEMISTRY	28	1.617
14	JOURNAL OF APPLIED PHYSICS	26	1.501
15	JOURNAL OF PHYSICS CONDENSED MATTER	22	1.27
16	SOLID STATE COMMUNICATIONS	20	1.155

Groups PHYS-A PHYS-B PHYS-C PHYS-D PHYS-E PI

Figure 27. Excerpt of Physics research groups and panel_journal title.xlsx file

3.2 Correlation between publication profiles of research groups together and panel

a) Pearson’s correlation coefficient and Spearman’s rank-order correlation coefficient

We determine the correlation between the publication output of groups and panel, using Pearson’s correlation coefficient and Spearman’s rank-order correlation coefficient for the numbers of publications per journal. We make an Excel file ‘Physics panel and groups together_ journals title.xlsx’ and export data from ‘Panel_ journals titles. txt’ and ‘Groups together_ journals title.txt’ in two different sheets (Figure 28). We reuse the Python script ‘join-sheets.py’ (see section 2.2) to take the data of the two sheets and join it into one. We run the program as:

```
python join-sheets.py "Physics panel and research groups together_journals
title.xlsx"
```

	A	B	C
1	**Source Titles**	**records**	**% of 1732**
2	PHYSICAL REVIEW B	291	16.801
3	EUROPEAN PHYSICAL JOURNAL C	108	6.236
4	PHYSICS LETTERS B	72	4.157
5	PHYSICAL REVIEW LETTERS	60	3.464
6	PHYSICA C SUPERCONDUCTIVITY AND ITS APPL	43	2.483
7	ULTRAMICROSCOPY	42	2.425
8	APPLIED PHYSICS LETTERS	37	2.136
9	CHEMISTRY OF MATERIALS	31	1.79
10	PHYSICAL REVIEW D	30	1.732
11	PHYSICA E LOW DIMENSIONAL SYSTEMS NANO	30	1.732
12	PHYSICAL REVIEW E	29	1.674
13	JOURNAL OF SOLID STATE CHEMISTRY	28	1.617
14	JOURNAL OF APPLIED PHYSICS	26	1.501
15	JOURNAL OF PHYSICS CONDENSED MATTER	22	1.27
16	SOLID STATE COMMUNICATIONS	20	1.155
17	JOURNAL OF MATERIALS CHEMISTRY	15	0.866
18	PHYSICAL REVIEW A	14	0.808
19	NANOTECHNOLOGY	14	0.808
20	OTOLOGY NEUROTOLOGY	13	0.751
21	EUROPHYSICS LETTERS	13	0.751
22	SOLID STATE SCIENCES	12	0.693
23	PHYSICAL CHEMISTRY CHEMICAL PHYSICS	12	0.693
24	JOURNAL OF PHYSICAL CHEMISTRY B	11	0.635
25	PHYSICA STATUS SOLIDI A APPLIED RESEARCH	10	0.577
26	PHYSICA A STATISTICAL MECHANICS AND ITS AI	10	0.577
27	JOURNAL OF THE AMERICAN CHEMICAL SOCIET	10	0.577
28	JOURNAL OF LOW TEMPERATURE PHYSICS	10	0.577
29	EPL	10	0.577
30	CARBON	10	0.577
31	MICROPOROUS AND MESOPOROUS MATERIAL!	9	0.52
32	MATERIALS SCIENCE AND ENGINEERING A STRI	9	0.52

GroupsTogether PanelTogether

	A	B	C
1	**Source Titles**	**records**	**1104**
2	PHYSICAL REVIEW B	246	18.65
3	PHYSICAL REVIEW LETTERS	98	7.43
4	SURFACE SCIENCE	56	4.246
5	APPLIED PHYSICS LETTERS	54	4.094
6	PHYSICS LETTERS B	51	3.867
7	PROCEEDINGS OF THE SOCIETY OF PHOTO OPTICA	49	3.715
8	PROCEEDINGS OF SPIE	39	2.957
9	EUROPEAN PHYSICAL JOURNAL C	34	2.578
10	APPLIED OPTICS	33	2.502
11	SOLID STATE COMMUNICATIONS	30	2.274
12	APPLIED MAGNETIC RESONANCE	28	2.123
13	CHEMICAL PHYSICS LETTERS	27	2.047
14	ZEITSCHRIFT FUR PHYSIK C PARTICLES AND FIELDS	22	1.668
15	OPTICS LETTERS	18	1.365
16	JOURNAL OF APPLIED PHYSICS	18	1.365
17	INSTITUTE OF PHYSICS CONFERENCE SERIES	16	1.213
18	OPTICS EXPRESS	15	1.137
19	ADVANCED MATERIALS	15	1.137
20	JOURNAL OF PHYSICS CONDENSED MATTER	14	1.061
21	JOURNAL OF PHYSICAL CHEMISTRY B	14	1.061
22	MICROSCOPY AND MICROANALYSIS	13	0.986
23	JOURNAL OF CHEMICAL PHYSICS	13	0.986
24	OPTICAL ENGINEERING	12	0.91
25	NANO LETTERS	12	0.91
26	JOURNAL OF PHYSICS C SOLID STATE PHYSICS	11	0.834
27	JOURNAL OF PHYSICAL CHEMISTRY	11	0.834
28	CHEMICAL PHYSICS	11	0.834
29	MOLECULAR PHYSICS	10	0.758
30	JOURNAL OF MAGNETIC RESONANCE	10	0.758
31	NATURE MATERIALS	9	0.682
32	JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	9	0.682

GroupsTogether **PanelTogether**

Figure 28. Excerpt of the Physics panel and research groups together_journals title.xlsx file

It produces a new Excel file named ‘Physics panel and research groups together_journals title-joined.xlsx’ (Figure 29). To calculate correlation, the value zero was kept on the corresponding journals in which either the panel or the groups had no publications (but not both). We calculate correlation coefficient using SPSS and obtain value ($r = 0.85$, $\rho = -0.24$). A log-log plot of the number of publications per journal for the Physics panel and research groups together are shown in Figure 30.

	A	B	C	D
1		**Source Titles**	**records_x**	**records_y**
2	**0**	PHYSICAL REVIEW B	291	246
3	**1**	EUROPEAN PHYSICAL JOURNAL C	108	34
4	**2**	PHYSICS LETTERS B	72	51
5	**3**	PHYSICAL REVIEW LETTERS	60	98
6	**4**	PHYSICA C SUPERCONDUCTIVITY AND ITS APPLICATIONS	43	1
7	**5**	ULTRAMICROSCOPY	42	0
8	**6**	APPLIED PHYSICS LETTERS	37	54
9	**7**	CHEMISTRY OF MATERIALS	31	0
10	**8**	PHYSICAL REVIEW D	30	4
11	**9**	PHYSICA E LOW DIMENSIONAL SYSTEMS NANOSTRUCTURES	30	1
12	**10**	PHYSICAL REVIEW E	29	5
13	**11**	JOURNAL OF SOLID STATE CHEMISTRY	28	0
14	**12**	JOURNAL OF APPLIED PHYSICS	26	18
15	**13**	JOURNAL OF PHYSICS CONDENSED MATTER	22	14
16	**14**	SOLID STATE COMMUNICATIONS	20	30
17	**15**	JOURNAL OF MATERIALS CHEMISTRY	15	0
18	**16**	PHYSICAL REVIEW A	14	1
19	**17**	NANOTECHNOLOGY	14	5
20	**18**	OTOLOGY NEUROTOLOGY	13	0

Figure 29. Excerpt of the Physics panel and research groups together_journals title-joined.xlsx file

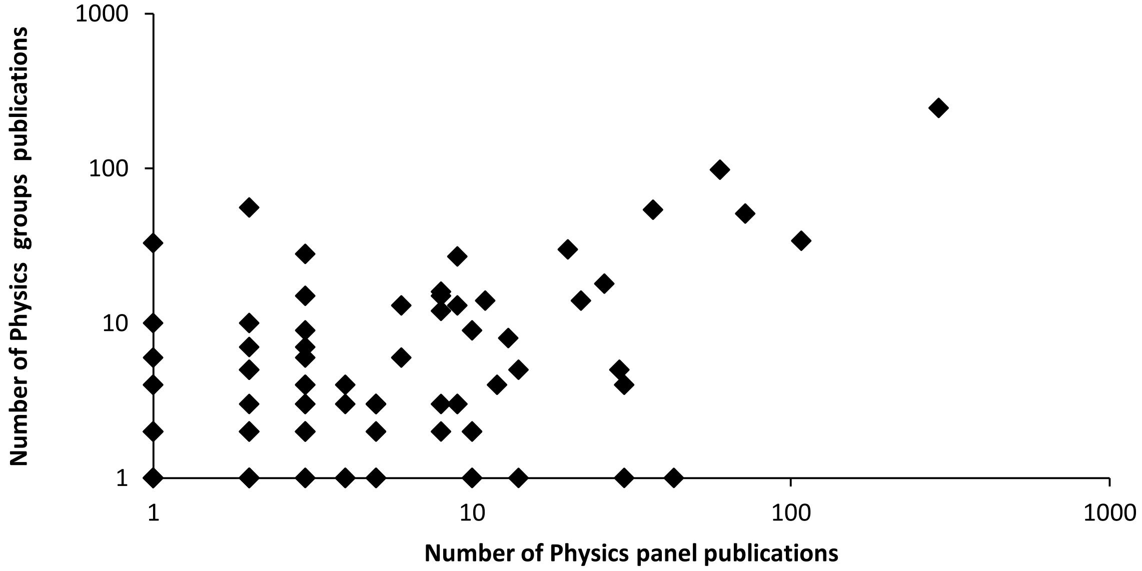


Figure 30. Log-log plot of the number of publications (log-log scale) per journals for the panel (horizontal axis) and research groups together (vertical axis) of the Physics department

b) Top-Down correlation coefficient

In some cases, the panel has published in a journal where the research groups have not or vice versa , i.e. there are many zeroes on both sides. Since traditional correlation coefficients like Pearson's and Spearman's are not well-suited to zero-inflated data (i.e., data with a large amounts of zeroes), we adopt the Top-down correlation coefficient (Iman & Conover, 1987). This correlation coefficient was found to be an adequate rank correlation coefficient for zero-inflated data (Huson, 2007). For a full description of the Top-down correlation coefficient we refer to Iman and Conover (1987). This coefficient places emphasis on the higher ranked data by computing the correlation using Savage scores derived from the ranked data.

We reuse the formula 1 and 2 (details in section 2.2b) and the python script "calc_topdowncorr.py" (all core logic is in topdowncorr.py, see section 5). We reuse the 'Physics research groups and panel_journal title- joined.xlsx' (Figure 29) file, but keep the zeros in the WoS SCs where neither the panel nor the research groups have publications. We run the program as:

```
python calc_topdowncorr.py "Physics research groups and panel_ journals title-
joined.xlsx"
```

The outcome shows that the Top-down correlation between Physics research groups together and the panel based on the journals in which they publish is low (0.16).

In our opinion, the correlations are an insufficient measure in this case, as the similarity of journals is not taken into account here. This is reminiscent of the way diversity is sometimes studied using only the dimensions of variety and balance. As discussed by Stirling (2007), the additional dimension of disparity – the opposite concept of similarity – is needed to provide a complete picture. Likewise, a comparison of publication profiles based on journals that does not consider journal similarity might yield distorted results.

3.3 Journal similarity matrix

Journal similarity data were received as a NET file (file name cosine.net) from Loet Leydesdorff in the context of the joint paper (Rahman, Guns, Leydesdorff, & Engels, 2016). While we did not construct this similarity matrix ourselves, we briefly outline the main steps that were taken to create it. The data was harvested from Clarivate Analytics's (formerly Thomson Reuters') Journal Citation Reports (JCR) of the Science and Social Science Editions 2011. An aggregated journal-journal citation matrix of 10,675 journals1 was

[1] The Science and Social Science Editions 2011 contain 8281 and 2943 journals respectively. Of these journals, 549 are contained in both databases.

constructed with a grand total of 35,295,459 citations over the entire matrix, which was subsequently normalized in the citing direction. The similarities between journals are calculated using the cosine similarity between their citing distributions respectively (see Leydesdorff, Rafols, & Chen (2013) for details). The resulting journal similarity matrix can be considered as an adjacency matrix, and thus is equivalent to a weighted network where similar journals are linked and link weights increase with similarity strength.

The size of the file ‘cosine.net’ is around 1 gigabyte. First, we compress the file using gzip to ‘cosine.net.gz’. After compression, the file is 291 megabytes. Next, we use a Python script ‘load_ndim_data.py’ to produce a file ‘matrix.h5’, which contains the network’s adjacency matrix and is used further on. We use the gzipped network file ‘cosine.net.gz’ as input and run:

```
python load_ndim_data.py cosine.net.gz
```

This way, we store the adjacency matrix in HDF5 (Hierarchical Data Format version 5), which was found to be the most efficient way of storing the data in terms of speed and memory requirements. From http://www.leydesdorff.net/journals11/citing_all.txt, we download the journal VOS map and save it to a file named ‘Journal_VOS_map.xlsx’ (Figure 31).

	A	B	C	D	E	F
1	id	label description	x	y	normalized weight	cluster
2	1	4OR-A Quarterly Journal of Operations Research	0.6434	0.2171	0.01	1000
3	2	AAOHN JOURNAL	-0.0632	-0.2391	0.01	1000
4	3	AAPG BULLETIN	0.0281	0.6925	0.01	1000
5	4	AAPS Journal	-0.6656	0.1736	0.01	1000
6	5	AAPS PHARMSCITECH	-0.4438	0.3727	0.01	1000
7	6	AATCC REVIEW	0.1818	0.6869	0.01	1000
8	7	ABDOMINAL IMAGING	-0.6488	-0.3543	0.01	1000
9	8	ABHANDLUNGEN AUS DEM MATHEMATISCHEN SEMINAR DER UNIVER	0.9848	0.617	0.01	1000
10	9	Abstract and Applied Analysis	0.8847	0.595	0.01	1000
11	10	ACADEMIC EMERGENCY MEDICINE	-0.6345	-0.5654	0.01	1000

Figure 31. Excerpt of the journal VOS map data

3.4 Journal overlay map creation

During data collection, the resulting files were downloaded using the default name ‘analyze.txt’ (see section 3.1). We downloaded the ‘Analyze.exe’ program, as well as the file ‘citing.dbf’ from http://www.leydesdorff.net/journals11.

For each entity (Individual research groups, panel members, research groups together and panel), we save the corresponding ‘analyze.txt’ file in the concerned folder and run the program ‘Analyze.exe’. ‘Analyze.exe’ reads ‘analyze.txt’, and generates an output file ‘citing.txt’. We open the latter in VOSviewer to obtain an overlay map. For example, Figure 32 shows the journal overlay map of the PHYS-B research group.

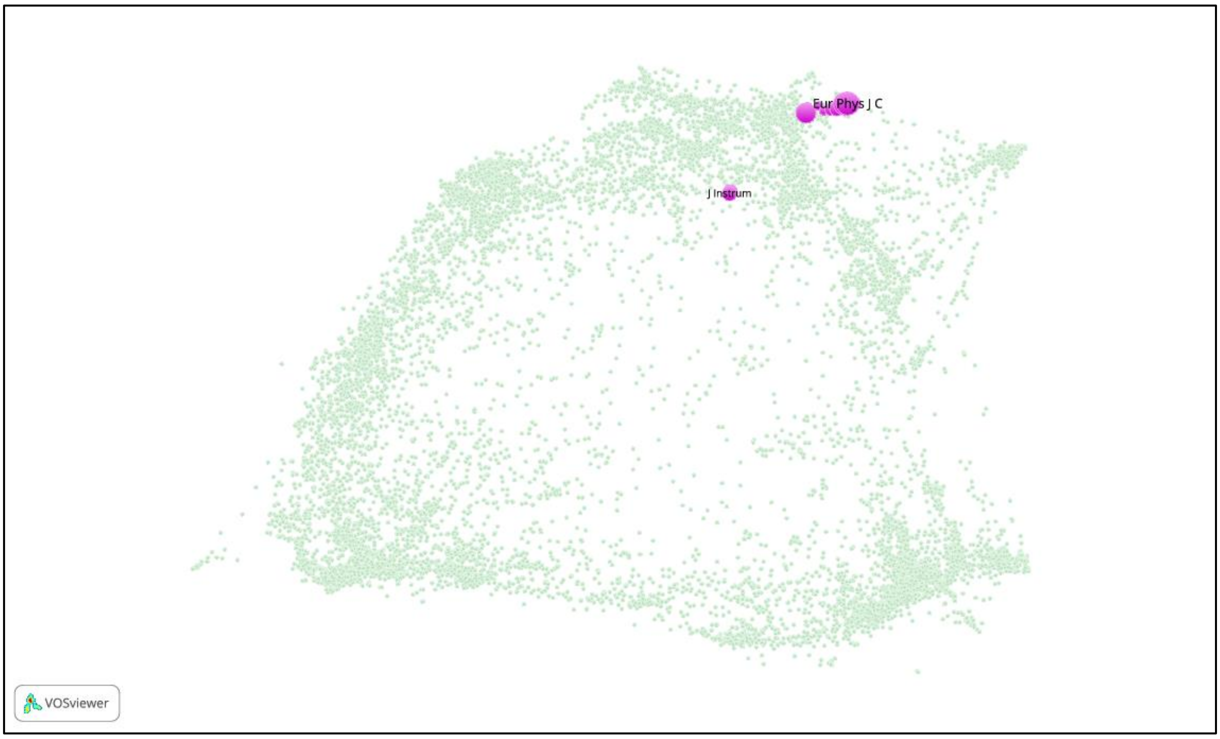


Figure 32. Journal overlay map of the PHYS-B research group

We prepare separate journal overlay maps for each research group, each panel member, research groups together and panel (see Appendix C).

3.5 Barycenter method

a) Barycenter calculation

We recall the formula 3. The barycenter is defined as the point $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2)$, where

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

Here, Lj,1 and Lj,2 are the horizontal and vertical coordinates of journal j on the map, m_j is the number of publications in journal j, and $T = \sum_{j=1}^{N} m_j$ is the total number of publications of the entity.

Based on formula 3, Python script 'journals-barycenter.py' is used. This script takes 'Journal_VOS_map.xlsx' (Figure 31) and 'Physics research groups and panel_journals title.xlsx' (Figure 27) as input.

We run the program as follows

```
python journals-barycenter.py "Journal_VOS_map.xlsx" "Physics research groups and panel journals title.xlsx"
```

At this point, we notice that our program indicates that the journal titles in our Physics data do not match with the journal titles of the VOS map. We find that in the journal similarity matrix, the journal titles are written in short form while our downloaded data from WoS contains the full titles. In 'citing.dbf' (available at http://www.leydesdorff.net/journals11) both shortened and full titles are available. In addition, we have received 487 records from Loet Leydesdorff that were not included in the 'citing.dbf' file. Based on 'citing.dbf' and the additional data, we make a separate file 'translation table.xlsx' (Figure 33). We use the full title of the journals for matching.

	A	B
1	**CITEDJ**	**TITLE**
2	INDOGER FORSCH	INDOGERMANISCHE FORSCHUNGEN
3	SOCIETY	SOCIETY
4	ECOHEALTH	ECOHEALTH
5	GLASS PHYS CHEM	GLASS PHYSICS AND CHEMISTRY
6	ANN UROL	ANNALES D UROLOGIE
7	POPULATION	POPULATION
8	ADULT EDUC QUART	ADULT EDUCATION QUARTERLY
9	DEV DISABIL RES REV	DEVELOPMENTAL DISABILITIES RESEARCH REVIEWS
10	ANTIBIOTIQUES	ANTIBIOTIQUES
11	FUJITSU SCI TECH J	FUJITSU SCIENTIFIC & TECHNICAL JOURNAL
12	AGRARFORSCHUNG	AGRARFORSCHUNG

Figure 33. Excerpt of short form to full journal titles

We modify the program to accommodate the translation table. We rerun the program. This time our program indicates that there are some journals that do not match with any journal in the VOS map. This turns out to be due to name or organizational changes over time; indeed, journals are not static entities. More specifically, possible reasons are:

- The journal title is changed, shortened or extended;
- Two or more journals merge into a new journal;
- One journal splits into two or more new journals;
- A journal is excluded from the WoS, discontinued, or not listed during the construction of the aggregated journal-journal citation matrix.

We have developed the following guidelines to handle these uniformly:

- If journal A is renamed to B then treat both as equivalent.
- If journals A1 and A2 are merged into journal B, we treat both A1 and A2 as equivalent to B.
- If journal X splits into multiple journals, we look up which research groups or panel members have publications in journal X and determine which of the new journals best corresponds to the specialty of the authors, then change all occurrences of the journals in the WoS exported data with the best fitting latter journals.
- If a journal is discontinued or excluded from WoS, or not included in the aggregated journal-journal citation matrix and there is no equivalent for some other reason, then it is removed from the sample.

	A	B
1	OLD TITLES	CORRECT TITLES
159	JOURNAL OF PHYSICS E SCIENTIFIC INSTRUMENTS	MEASUREMENT SCIENCE AND TECHNOLOGY
160	ZEITSCHRIFT FUR PHYSIK C PARTICLES AND FIELDS	EUROPEAN PHYSICAL JOURNAL C
161	ZEITSCHRIFT FUR PHYSIK B CONDENSED MATTER	EUROPEAN PHYSICAL JOURNAL B
162	JOURNAL OF PHYSICS C SOLID STATE PHYSICS	JOURNAL OF PHYSICS: CONDENSED MATTER
163	JOURNAL OF PHYSICAL CHEMISTRY	JOURNAL OF PHYSICAL CHEMISTRY A
164	PHILOSOPHICAL MAGAZINE B PHYSICS OF CONDENSED MATTER STAT	PHILOSOPHICAL MAGAZINE
165	EUROPEAN PHYSICAL JOURNAL	EUROPEAN PHYSICAL JOURNAL A
166	CHEMISCHE BERICHTE RECUEIL	EUROPEAN JOURNAL OF ORGANIC CHEMISTRY
167	CHEMISCHE BERICHTE	EUROPEAN JOURNAL OF INORGANIC CHEMISTRY
168	CANADIAN JOURNAL OF APPLIED SPECTROSCOPY	CANADIAN JOURNAL OF ANALYTICAL SCIENCES AND SPECTROSCOPY
169	INORGANICA CHIMICA ACTA ARTICLES AND LETTERS	INORGANICA CHIMICA ACTA
170	AMERICAN JOURNAL OF MEDICAL GENETICS	AMERICAN JOURNAL OF MEDICAL GENETICS PART A
171	ANALYTICAL LETTERS PART A CHEMICAL ANALYSIS	ANALYTICAL LETTERS
172	JOURNAL OF THE SOUTH AFRICAN CHEMICAL INSTITUTE	SOUTH AFRICAN JOURNAL OF CHEMISTRY-SUID-AFRIKAANSE TYDSKRIF VIR CHEMIE
173	GEMS GEMOLOGY	GEMS & GEMOLOGY
174	BULLETIN DE LA SOCIETE CHIMIQUE DE FRANCE	EUROPEAN JOURNAL OF ORGANIC CHEMISTRY
175	INORGANICA CHIMICA ACTA LETTERS	INORGANICA CHIMICA ACTA
176	JOURNAL OF GEOPHYSICAL RESEARCH ATMOSPHERES	JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES
177	PHILOSOPHICAL MAGAZINE B PHYSICS OF CONDENSED MATTER STAT	PHILOSOPHICAL MAGAZINE
178	INORGANICA CHIMICA ACTA ARTICLES	INORGANICA CHIMICA ACTA

Figure 34. Excerpt of journal name change.xlsx file

For each journal that is not found in the map, we search the title in the WoS and Journal Citation Reports, and consult its website as well as the ISSN database (www.issn.org) to identify the reasons behind the title change. Subsequently, based on the abovementioned guidelines we make a separate MS Excel file 'Journal name change.xlsx' (Figure 34) to translate 'old' titles to 'correct' titles.

We keep the 'Physics research groups and panel_journals title.xlsx' (Figure 27), 'Journals_VOS_map.xlsx' (Figure 31), 'translate.xlsx' (Figure 33), and 'Journal name change.xlsx' (Figure 34) files in a folder. A modified Python script 'journals-barycenter.py' is used that takes the 'Journal name change.xlsx' file into account.

We run the program as follows:

```
python journals-barycenter.py "Journals_VOS_map.xlsx" "Physics research groups and
panel journals title.xlsx"
```

	A	B	C
1		x	y
2	Groups	0.26546	0.64079
3	PHYS-A	-0.18082	-0.12955
4	PHYS-B	0.369299	0.734285
5	PHYS-C	0.212416	0.729048
6	PHYS-D	0.494506	0.747526
7	PHYS-E	0.321577	0.717452
8	PHYS-F	0.532545	0.674528
9	PHYS-G	-0.25341	0.514983
10	PHYS-H	0.375508	0.746617
11	PHYS-I	-0.00977	0.294102
12	PM1	0.52123	0.760538
13	PM2	0.368131	0.759631
14	PM3	0.379768	0.649184
15	PM4	0.071586	0.689779
16	PM5	0.285512	0.72069
17	PM6	0.352572	0.749822
18	Panel	0.316137	0.726916

Figure 35. Barycenter coordinates of the Physics individual research groups, panel members, research groups together, and panel using journal VOS map

This program calculates the barycenter coordinates of Physics individual research groups, panel members, groups and panel in the journals VOS map in an output file 'Physics research groups and panel_ journals title -barycenters.xlsx' (Figure 35).

b) Euclidean distance calculation between barycenters

Subsequently, we determine the Euclidean distances between the barycenters of different entities: individual research groups, panel members, research groups together and panel. We reuse the formula 4 and determine the Euclidean distance between barycenters. We note that the python script 'journals-categories.py' executes both formula 3 and 4.

From the matrix of Euclidean distances between all entity pairs (Figure 36), we extract Table 8 containing only distances between the research groups and panel on the one hand and the panel and panel members on the other, for the convenience of analysis.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I	PM1
2	Groups	0	0.890282	0.139727	0.102972	0.252695	0.095006	0.269207	0.533904	0.152676	0.442657	0.282414
3	PHYS-A	0.890282	0	1.024136	0.944371	1.106952	0.984799	1.074916	0.648612	1.037875	0.456885	1.133643
4	PHYS-B	0.139727	1.024136	0	0.15697	0.125906	0.050603	0.17384	0.660196	0.013807	0.580909	0.154183
5	PHYS-C	0.102972	0.944371	0.15697	0	0.282695	0.109776	0.324739	0.512657	0.164036	0.488411	0.310416
6	PHYS-D	0.252695	1.106952	0.125906	0.282695	0	0.175525	0.082315	0.783233	0.119002	0.678151	0.029724
7	PHYS-E	0.095006	0.984799	0.050603	0.109776	0.175525	0	0.21529	0.609593	0.061312	0.537603	0.204249
8	PHYS-F	0.269207	1.074916	0.17384	0.324739	0.082315	0.21529	0	0.801984	0.172794	0.662442	0.086752
9	PHYS-G	0.533904	0.648612	0.660196	0.512657	0.783233	0.609593	0.801984	0	0.670218	0.32886	0.812628
10	PHYS-H	0.152676	1.037875	0.013807	0.164036	0.119002	0.061312	0.172794	0.670218	0	0.594315	0.146386
11	PHYS-I	0.442657	0.456885	0.580909	0.488411	0.678151	0.537603	0.662442	0.32886	0.594315	0	0.706771
12	PM1	0.282414	1.133643	0.154183	0.310416	0.029724	0.204249	0.086752	0.812628	0.146386	0.706771	0
13	PM2	0.157049	1.04499	0.025373	0.15869	0.126954	0.06282	0.185134	0.667956	0.014959	0.599606	0.153102
14	PM3	0.114615	0.95953	0.085743	0.185432	0.151116	0.089703	0.154865	0.647243	0.097526	0.52709	0.180031
15	PM4	0.199968	0.857333	0.301021	0.146202	0.426844	0.251518	0.461211	0.36902	0.309191	0.403955	0.455177
16	PM5	0.082377	0.969734	0.084883	0.073572	0.210711	0.036211	0.251309	0.576846	0.093657	0.518815	0.239063
17	PM6	0.139557	1.028501	0.022829	0.141687	0.141953	0.044815	0.195089	0.649894	0.023159	0.582213	0.168999
18	Panel	0.099929	0.990209	0.05367	0.103743	0.179556	0.010917	0.222659	0.607699	0.062555	0.541797	0.207831

Figure 36. Excerpt of Euclidean distances matrix of the barycenter of the Physics groups, panel members, research groups together and panel using the journal VOS map

Table 8. Euclidean distances between barycenter of Physics individual research groups, panel members, research groups together and panel using the journal VOS map

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.100	0.990	0.054	0.104	0.180	0.011	0.223	0.608	0.063	0.542
PM1	0.282	1.134	0.154	0.310	**0.030**	0.204	**0.087**	0.813	0.146	0.707
PM2	0.157	1.045	**0.025**	0.159	0.127	**0.063**	0.185	0.668	**0.015**	**0.600**
PM3	0.115	**0.960**	0.086	0.185	0.151	**0.090**	**0.155**	**0.647**	0.098	**0.527**
PM4	0.200	**0.857**	0.301	0.146	0.427	0.252	0.461	**0.369**	0.309	**0.404**
PM5	0.082	**0.970**	0.085	**0.074**	0.211	**0.036**	0.251	**0.577**	0.094	**0.519**
PM6	0.140	**1.029**	**0.023**	0.142	0.142	**0.045**	0.195	**0.650**	**0.023**	**0.582**

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.210 (SD. 0.285). The number in the row of 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).

In Table 8, for each research group we find the shortest distances to one of the panel members, and underline and bold it. In addition, the average and standard deviation of the shortest distances are calculated. The confidence intervals (discussed in section 2.5) are included through the typography of the values.

c) Barycenter overlay map

We take the Journal level_VOS map (Figure 31) file and manually add the Physics individual groups, panel members, research groups together and panel's coordinates (Figure 35) after the 10,673 journals title. We fill up the 'weight' column with 20 (we can put other numbers too) to highlight the size of the bubble of the added entities. In the 'cluster' column, we assign 1 to all the 10,673 journals, 2 to the research groups together, 3 to all research groups, 4 to the panel, and 5 to individual panel members. We save the map file as 'Barycenter map of Physics department in the journal level.csv'. After that, we open the file with VOSviewer to visualize the barycenters (Figure 37). Figure 38 shows a zoomed in version of Figure 37.

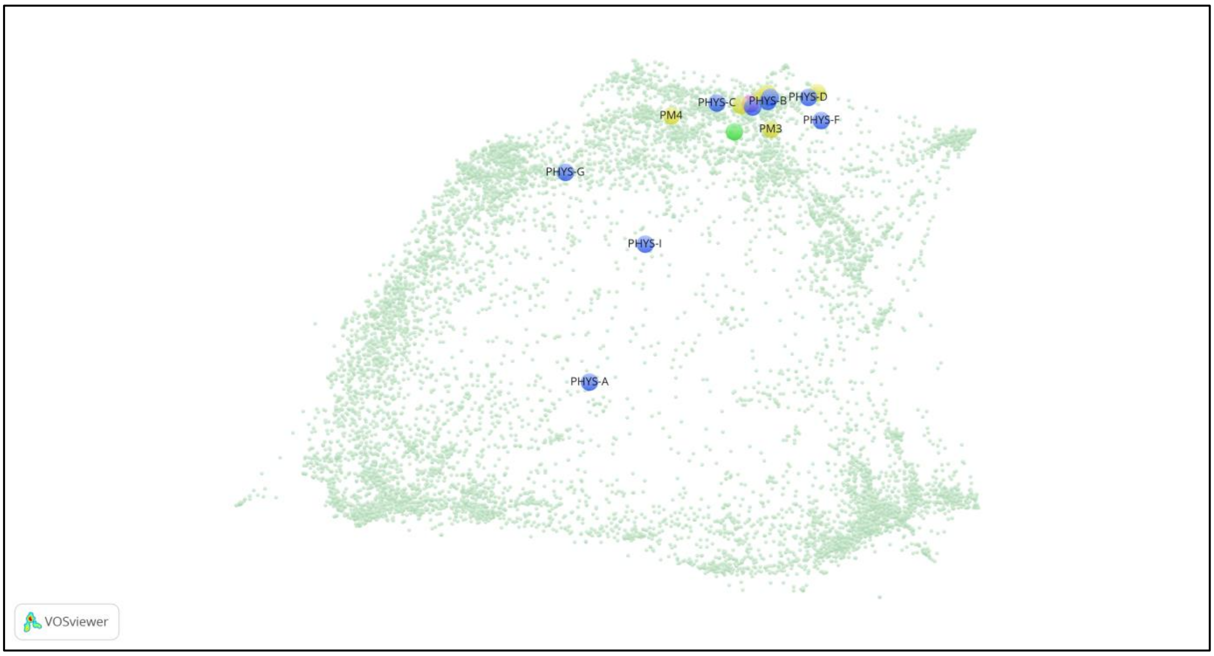


Figure 37. Barycenter overlay map of Physics panel, panel members (PM), research groups and research groups together)

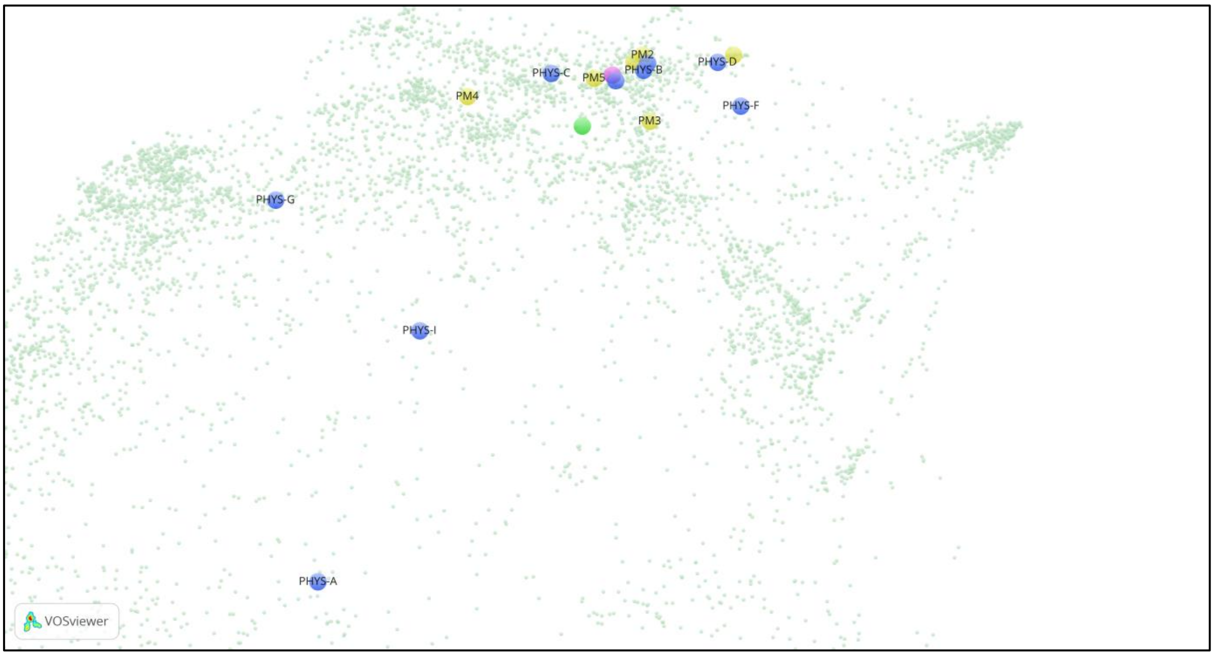


Figure 38. Barycenter overlay map of Physics panel, panel members (PM), research groups and research groups together (zoomed)

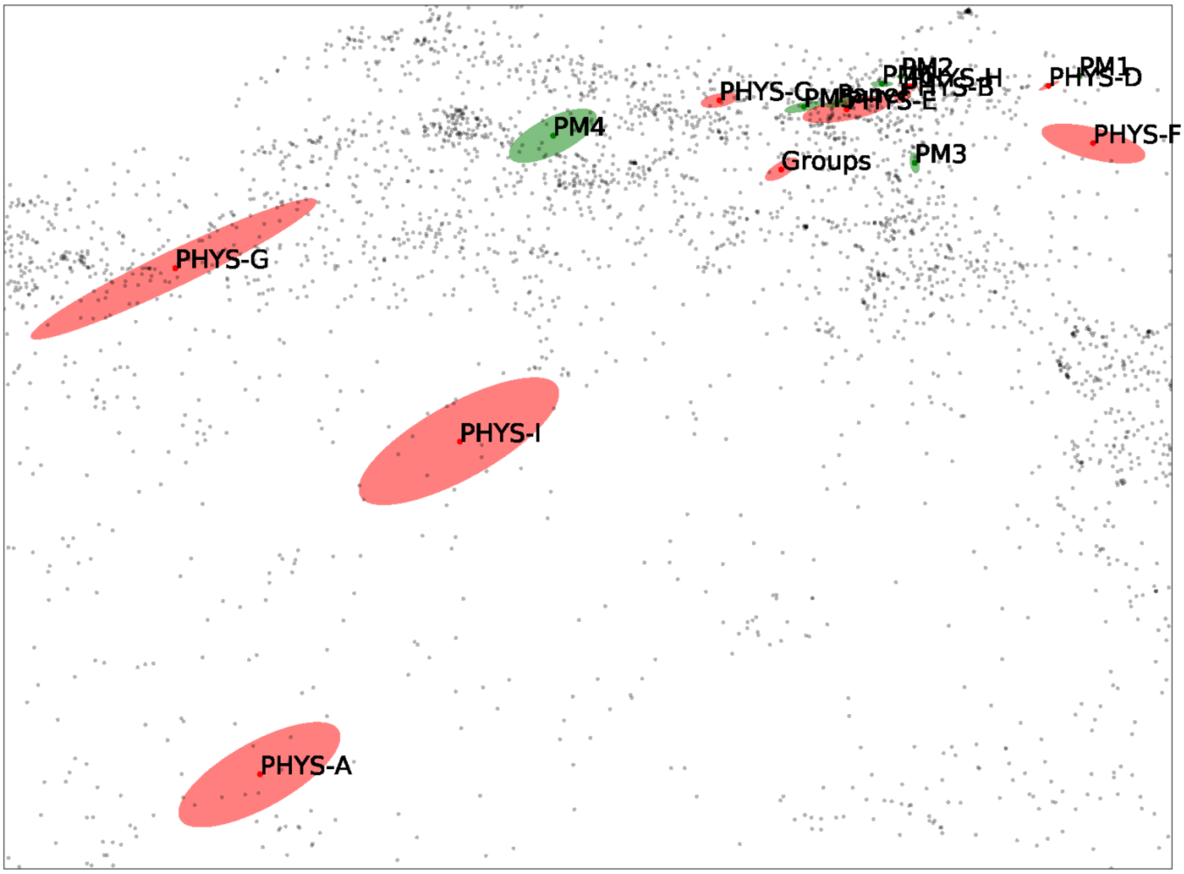


Figure 39. Barycenter overlay map of Physics panel, panel members (PM), research groups and research groups together with their confidence regions

We also create the barycenter overlap map of Physics department and include the confidence regions of the respective barycenter of panel, panel members (PM), research groups and research groups together using the journal VOS map (Figure 39). The bootstrap replications of barycenters are also used to add a 95% confidence region for each barycenter to the maps. In particular, we stress that overlapping confidence regions as seen in Figure 39 (figure with overlapping regions in it) does not correspond to overlap between CIs for distances. For detail process about confidence regions see section 2.6c.

3.6 Similarity-adapted publication vector method

a) Similarity-adapted publication vector calculation

Recall formula 5. A similarity-adapted publication vector is determined as the vector $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_\mathrm{N})$, where:

$$C_k = \frac{\sum_{j=1}^{N} s_{kj} m_j}{\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} m_j}$$

Here, s_{kj} denotes the similarity value between the k-th and the j-th journal, and m_j is the number of publications in journal j. The numerator of the formula 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		4OR A QU.	AAOHN JC	AAPG BUL	AAPS JOU	AAPS PHA	AATCC RE'	ABDOMIN	ABHANDL	ABSTRACT	ACADEMI	ACADEMI	ACADEMI
2	Groups	1.77E-06	2.04E-05	2.5E-05	8.75E-05	2.94E-05	8.19E-05	1.33E-05	5.12E-06	2.84E-05	9.43E-06	1.38E-05	8.54E-06
3	PHYS-A	2.49E-07	7.88E-05	1.75E-05	0.000136	2.15E-05	4.12E-05	9.23E-05	4.66E-06	1.15E-05	0.000104	8.86E-05	0.000105
4	PHYS-B	4.51E-07	1.52E-05	2.55E-05	5.86E-05	1.49E-05	4.97E-05	1E-06	2.41E-06	2.74E-05	1.97E-06	7.79E-06	1.54E-06
5	PHYS-C	1.32E-07	1.81E-05	2.71E-05	9.61E-05	4.76E-05	0.000125	1.2E-06	7.44E-07	1.27E-05	2.72E-06	8.95E-06	2.14E-06
6	PHYS-D	1.73E-07	8.55E-06	1.52E-05	3.2E-05	5.41E-06	2.33E-05	1.68E-06	1.41E-05	3.67E-05	7.48E-07	4.73E-06	6.07E-07
7	PHYS-E	1.93E-07	1.52E-05	2.35E-05	9.07E-05	3.29E-05	9.97E-05	1.17E-05	3.76E-06	2.16E-05	3.35E-06	8.19E-06	2.11E-06
8	PHYS-F	4.03E-05	1.19E-05	3.13E-05	6.72E-05	1.06E-05	4.31E-05	2.49E-07	0.000108	0.000356	2.02E-06	6.97E-06	1.66E-06
9	PHYS-G	1.17E-07	1.82E-05	2.33E-05	0.000222	4.63E-05	0.000116	8.81E-06	7.34E-07	1.84E-05	5.93E-06	1.09E-05	4.95E-06
10	PHYS-H	2.22E-07	1.49E-05	2.49E-05	5.52E-05	1.34E-05	4.89E-05	3.9E-07	1.67E-06	2.59E-05	1.56E-06	7.63E-06	1.27E-06
11	PHYS-I	1.14E-05	3.51E-05	3.24E-05	0.000143	2.52E-05	8.54E-05	0.000107	6.91E-06	2.77E-05	2.46E-05	2.84E-05	1.7E-05
12	PM1	0	2.98E-06	5.46E-06	1.14E-05	1.5E-06	7.85E-06	0	2.34E-05	3.82E-05	0	1.95E-06	0
13	PM2	5.45E-09	1.36E-05	2.13E-05	4.7E-05	1.2E-05	4.17E-05	1.22E-07	7.97E-07	1.94E-05	1.36E-06	6.67E-06	1.03E-06
14	PM3	6.89E-07	1.49E-05	2.7E-05	5.82E-05	1.56E-05	4.96E-05	2.75E-06	7.1E-07	2.46E-05	1.88E-06	7.75E-06	1.44E-06
15	PM4	1.81E-07	1.57E-05	2.44E-05	0.000136	6.21E-05	0.000112	4.67E-05	1.32E-06	3.63E-05	6.01E-06	9.72E-06	4.23E-06
16	PM5	8.12E-08	1.97E-05	2.88E-05	8.95E-05	3.45E-05	9.36E-05	1.05E-06	1.03E-06	1.58E-05	2.5E-06	9.56E-06	1.95E-06
17	PM6	3.59E-07	1.55E-05	2.65E-05	6.42E-05	1.97E-05	6.12E-05	2.46E-07	2.86E-06	2.63E-05	1.69E-06	7.8E-06	1.42E-06
18	Panel	2.01E-07	1.62E-05	2.55E-05	7.99E-05	2.97E-05	7.51E-05	8.4E-06	2.2E-06	2.4E-05	2.59E-06	8.34E-06	1.97E-06

Figure 40. Excerpt of SAPV of the Physics research groups, research groups together, panel members and panel using journal similarity matrix

A Python script ‘sim_adapted_pub_vectors_journals.py’ (the calculation of SAPVs is carried out by the sa_vector function, see section 5) is used that takes as input the similarity matrix and the weights (number of publications) per journals (‘Physics groups and panel_ journals title.xlsx’, Figure 27). This script calculates SAPVs for all entities. We keep ‘matrix.h5’ (see section 3.3), ‘cosine.net.gz’ (see section 3.3), ‘translate.xlsx’ (Figure 33), ‘Journal name change.xlsx’ (Figure 34) and ‘Physics research groups and panel_ journals title.xlsx’ in a folder. We run the program as:

```
python sim_adapted_pub_vectors_journals.py matrix.h5 "Physics research groups and panel_journals title.xlsx"
```

This program calculates the SAPV (Figure 40) of each entity and stores the results in an output file named ‘Physics research groups and panel_ journals title-SA-Vector.xlsx’ (Figure 41).

b) Euclidean distance between similarity-adapted publication vectors

Subsequently, we determine the Euclidean distances between the SAPV of different entities: individual research groups, panel members, research groups together and panel. We reuse formula 6. Again, we use the implementation of Euclidean distance in scipy.spatial.dist.

	A	B	C	D	E	F	G	H	I	J	K	L
1		**Groups**	**PHYS-A**	**PHYS-B**	**PHYS-C**	**PHYS-D**	**PHYS-E**	**PHYS-F**	**PHYS-G**	**PHYS-H**	**PHYS-I**	**PM1**
2	**Groups**	0	0.020861	0.007866	0.006951	0.034287	0.004403	0.011906	0.017205	0.008748	0.013814	0.051938
3	**PHYS-A**	0.020861	0	0.026721	0.021303	0.044539	0.022563	0.023545	0.015424	0.02743	0.011451	0.060552
4	**PHYS-B**	0.007866	0.026721	0	0.012707	0.034477	0.009554	0.013418	0.024046	0.001311	0.020195	0.052217
5	**PHYS-C**	0.006951	0.021303	0.012707	0	0.039661	0.009709	0.017324	0.01557	0.013733	0.013851	0.056936
6	**PHYS-D**	0.034287	0.044539	0.034477	0.039661	0	0.030508	0.030595	0.043765	0.034036	0.041341	0.017998
7	**PHYS-E**	0.004403	0.022563	0.009554	0.009709	0.030508	0	0.011243	0.018351	0.010116	0.01581	0.048095
8	**PHYS-F**	0.011906	0.023545	0.013418	0.017324	0.030595	0.011243	0	0.02171	0.013455	0.018007	0.048122
9	**PHYS-G**	0.017205	0.015424	0.024046	0.01557	0.043765	0.018351	0.02171	0	0.024905	0.009925	0.06019
10	**PHYS-H**	0.008748	0.02743	0.001311	0.013733	0.034036	0.010116	0.013455	0.024905	0	0.021019	0.051774
11	**PHYS-I**	0.013814	0.011451	0.020195	0.013851	0.041341	0.01581	0.018007	0.009925	0.021019	0	0.058206
12	**PM1**	0.051938	0.060552	0.052217	0.056936	0.017998	0.048095	0.048122	0.06019	0.051774	0.058206	0
13	**PM2**	0.011757	0.030176	0.004387	0.015807	0.035484	0.013188	0.01687	0.027682	0.00401	0.023963	0.052927
14	**PM3**	0.018385	0.026914	0.019375	0.019653	0.040656	0.019431	0.021806	0.026663	0.019659	0.021961	0.057176
15	**PM4**	0.014047	0.021439	0.020085	0.011258	0.041873	0.01471	0.020142	0.011749	0.020958	0.015489	0.058538
16	**PM5**	0.004614	0.02169	0.008831	0.00533	0.037611	0.008135	0.015031	0.017873	0.009858	0.014226	0.055157
17	**PM6**	0.006366	0.025679	0.002436	0.01148	0.033997	0.007919	0.012299	0.022425	0.00305	0.019	0.051794
18	**Panel**	0.002796	0.023032	0.006109	0.007679	0.034544	0.005229	0.012512	0.019021	0.007018	0.015928	0.052236

Figure 41. Excerpt of pairwise Euclidean distances matrix between the SAPV of the Physics individual research groups, panel members, panel and research groups together using the journal similarity matrix

It is mentionable that the Python script ‘sim_adapted_pub_vectors_journals.py’ executes both formulas 5 and 6. Although the matrix and vectors are large, the calculation of SAPV and distances is relatively fast, due to the use of efficient matrix procedures implemented in NumPy (http://www.numpy.org) and SciPy (http://www.scipy.org).

From the calculated matrix of pairwise Euclidean distances between SAPVs of Physics individual research groups, panel members, research groups together and panel (Figure 41), we extract Table 9 containing only the distances between the research groups and research groups together on the one hand and the panel and panel members on the other, for the convenience of analysis.

Table 9. Euclidean distances between SAPV of Physics individual research groups, panel members, research groups together and panel using the journal similarity matrix

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.003	0.023	0.006	0.008	0.035	0.005	0.013	0.019	0.007	0.016
PM1	0.052	0.061	0.052	0.057	<u>**0.018**</u>	0.048	0.048	0.060	0.052	0.058
PM2	0.012	0.030	0.004	0.016	0.035	0.013	0.017	0.028	**0.004**	0.024
PM3	0.018	0.027	0.019	0.020	0.041	0.019	0.022	0.027	0.020	0.022
PM4	0.014	<u>**0.021**</u>	0.020	0.011	0.042	0.015	0.020	<u>**0.012**</u>	0.021	**0.015**
PM5	0.005	**0.022**	0.009	<u>**0.005**</u>	0.038	**0.008**	**0.015**	0.018	0.010	<u>**0.014**</u>
PM6	0.006	0.026	<u>**0.002**</u>	0.011	0.034	<u>**0.008**</u>	<u>**0.012**</u>	0.022	<u>**0.003**</u>	0.019

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.011 (SD. 0.007). The number in the row of 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).

In Table 9, for each research group we find the shortest distances to one of the panel members, and underline and bold those. In addition, the average and standard deviation of the shortest distances are calculated. We use the average and standard deviation of the shortest distances as a comparative measure. The confidence intervals (discussed in section 2.5) are included through the typography of the values.

3.7 Weighted cosine similarity method

Recall the formula 7. We consider a weighted similarity method (generalized cosine similarity). The weighted similarity between panel member (PM) k and research group m, according to Zhou et al. (2012) is:

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

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

The numerator is nothing but the matrix multiplication: $\left(M^{k}\right)^{t} * S * R^{m}$, where t denotes matrix transposition, S is the journal similarity matrix, Mk denotes the column matrix of publications of panel member PMk and Rm 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 similarity between panel member PMk and research group m.

This value is calculated for each panel member and each research group. Similarity-weighted cosine is implemented in Python as a fairly straightforward set of matrix operations (see section 5, weighted_cosine).

We keep 'matrix.h5' (see section 3.3), 'cosine.net.gz' (see section 3.3), 'translate.xlsx' (Figure 33), 'Journal name change.xlsx' (Figure 34) and 'Physics research groups and panel_ journals title.xlsx' in a folder. A python script 'cosine-journals.py' is used that takes as input the similarity matrix and the weights (number of publications) per journals ('Physics research groups and panel_ journals title.xlsx', Figure 27), and calculates SAPVs for all entities, as well as the pairwise distances between them. We run the program as:

```
python cosine-journals.py  matrix.h5  "Physics research groups and panel_journals
title.xlsx"
```

This program calculates the WCS value in an output file as 'Physics research groups and panel _ journals title-cosine.xlsx' (Figure 42).

From the calculated WCS value matrix of Physics individual research groups, panel members, research groups together and panel in journals (Figure 42), we extract Table 10 containing only the WCS value of the research groups and research groups together on the one hand and the panel and panel members on the other, for the convenience of analysis.

In Table 10, for each research group we find the highest similarity to one of the panel members, and underline and bold those. The confidence intervals (discussed in the section 2.5) are included through the typography of the values. We calculate similarity between two entities based on their publication vectors. We generated 1000 independent bootstrap samples and each time calculated the similarity.

	A	B	C	D	E	F	G	H	I	J	K	L
1		**Groups**	**PHYS-A**	**PHYS-B**	**PHYS-C**	**PHYS-D**	**PHYS-E**	**PHYS-F**	**PHYS-G**	**PHYS-H**	**PHYS-I**	**PM1**
2	**Groups**	1	0.188502	0.935792	0.869317	0.500414	0.915509	0.666615	0.418336	0.928094	0.50918	0.406794
3	**PHYS-A**	0.188502	1	0.108572	0.137968	0.034367	0.116957	0.095292	0.139613	0.10722	0.247714	0.019366
4	**PHYS-B**	0.935792	0.108572	1	0.779502	0.29328	0.769224	0.643255	0.281555	0.994777	0.412422	0.185796
5	**PHYS-C**	0.869317	0.137968	0.779502	1	0.170957	0.705078	0.416431	0.530775	0.758812	0.558565	0.094574
6	**PHYS-D**	0.500414	0.034367	0.29328	0.170957	1	0.764284	0.455264	0.068557	0.300692	0.111873	0.990188
7	**PHYS-E**	0.915509	0.116957	0.769224	0.705078	0.764284	1	0.656357	0.402177	0.767938	0.39288	0.688218
8	**PHYS-F**	0.666615	0.095292	0.643255	0.416431	0.455264	0.656357	1	0.230945	0.657687	0.306978	0.372169
9	**PHYS-G**	0.418336	0.139613	0.281555	0.530775	0.068557	0.402177	0.230945	1	0.263521	0.354696	0.034942
10	**PHYS-H**	0.928094	0.10722	0.994777	0.758812	0.300692	0.767938	0.657687	0.263521	1	0.401421	0.190579
11	**PHYS-I**	0.50918	0.247714	0.412422	0.558565	0.111873	0.39288	0.306978	0.354696	0.401421	1	0.06265
12	**PM1**	0.406794	0.019366	0.185796	0.094574	0.990188	0.688218	0.372169	0.034942	0.190579	0.06265	1
13	**PM2**	0.893076	0.086199	0.981868	0.74485	0.243828	0.708022	0.54244	0.234252	0.975011	0.368229	0.143338
14	**PM3**	0.341634	0.252004	0.306111	0.32021	0.102326	0.300871	0.245876	0.130528	0.310048	0.322659	0.062899
15	**PM4**	0.500557	0.082585	0.396566	0.58221	0.101953	0.481653	0.309712	0.821919	0.376505	0.297745	0.058433
16	**PM5**	0.915648	0.15153	0.879727	0.926649	0.246528	0.76378	0.517183	0.389038	0.864921	0.530819	0.154715
17	**PM6**	0.938586	0.106455	0.982565	0.778895	0.328653	0.794768	0.665437	0.337241	0.980482	0.420273	0.219728
18	**Panel**	0.979498	0.148253	0.948134	0.857371	0.422069	0.879404	0.641619	0.442726	0.939386	0.48139	0.324565

Figure 42. Excerpt of WCS value matrix of the Physics individual research groups, panel members, groups and panel using the journal similarity matrix

Table 10. WCS value of the Physics groups, panel members, panel and research groups together using the journal similarity matrix

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.979	0.148	0.948	0.857	0.422	0.879	0.642	0.443	0.939	0.481
PM1	0.407	0.019	0.186	0.095	**0.990**	**0.688**	0.372	0.035	0.191	0.063
PM2	0.893	0.086	**0.982**	0.745	0.244	**0.708**	**0.542**	0.234	**0.975**	0.368
PM3	0.342	**0.252**	0.306	0.320	0.102	0.301	0.246	0.131	0.310	0.323
PM4	0.501	0.083	0.397	0.582	0.102	0.482	0.310	**0.822**	0.377	0.298
PM5	0.916	**0.152**	0.880	**0.927**	0.247	**0.764**	**0.517**	0.389	0.865	**0.531**
PM6	0.939	**0.106**	**0.983**	0.779	0.329	**0.795**	**0.665**	0.337	**0.980**	**0.420**

For each research group we determine the panel member at the highest similarity. The number in the row corresponding to this panel member is indicated in bold and underlined. Similarities whose confidence intervals overlap with that of the highest similarities are in bold (same column).

Since the barycenter (see section 3.5) and SAPV (see section 3.6) approaches are distance-based rather than similarity-based, we use 1 – WCS as values to obtain dissimilarity values: weighted cosine dissimilarity (Table 11), denoted as WCD, which can more easily be compared with the other two approaches. For the sake of simplicity, the results are shown under the WCS method.

Table 11. WCD value of the Physics groups, panel members, panel and research groups together using the journal similarity matrix

	Groups	PHYS-A	PHYS-B	PHYS-C	PHYS-D	PHYS-E	PHYS-F	PHYS-G	PHYS-H	PHYS-I
Panel	0.021	0.852	0.052	0.143	0.578	0.121	0.358	0.557	0.061	0.519
PM1	0.593	0.981	0.814	0.905	**<u>0.010</u>**	**0.312**	0.628	0.965	0.809	0.937
PM2	0.107	0.914	**0.018**	0.255	0.756	**0.292**	**0.458**	0.766	**0.025**	0.632
PM3	0.658	**<u>0.748</u>**	0.694	0.680	0.898	0.699	0.754	0.869	0.690	0.677
PM4	0.499	0.917	0.603	0.418	0.898	0.518	0.690	**<u>0.178</u>**	0.623	0.702
PM5	0.084	**0.848**	0.120	**<u>0.073</u>**	0.753	**0.236**	**0.483**	0.611	0.135	**<u>0.469</u>**
PM6	0.061	**0.894**	**<u>0.017</u>**	0.221	0.671	**<u>0.205</u>**	**<u>0.335</u>**	0.663	**<u>0.020</u>**	**0.580**

The lowest similarity between a group and a panel member is underlined and printed in bold.

4. Heat map with hierarchical clustering

A heat map with hierarchical clustering is a two-dimensional representation of data where the values are represented by colors and arranging items in a hierarchy based on the similarity between them. It provides an immediate visual summary of information.

We have proposed three methods, each of which can be applied at two levels of aggregation - WoS SCs and journals. This leads to six approaches, as follows:

WoS SCs

i) Barycenter
ii) Similarity-adapted publication vector (SAPV)
iii) Weighted cosine similarity (WCS)

Journals

iv) Barycenter
v) Similarity-adapted publication vector (SAPV)
vi) Weighted cosine similarity (WCS)

We calculate Spearman's rank-order correlation coefficient between each pair of the six approaches. More specifically, we determine the correlation using the distances between barycenters and between SAPVs, and dissimilarity of individual research groups and panel members using 1 – WCS. We create an MS Excel file (Figure 43) containing:

i) Euclidean distances between barycenters of the Physics individual research groups and panel members at the level of WoS SCs,

ii) Euclidean distances between barycenters of the Physics individual research groups and panel members at the level of journals,

iii) Euclidean distances between SAPVs of the Physics individual research groups and panel members at the level of WoS SCs,

iv) Euclidean distances between SAPVs of the Physics individual research groups and panel members at the level of journals,

v) WCS value of Physics individual research groups and panel members at the level of WoS SCs,

vi) WCS value of Physics individual research groups and panel members at the level of journals.

	A	B	C	D	E	F
1	**Barycenter_ PHYS_WoS SCs**	**Barycenter_ PHYS_Journals**	**SAPV_ PHYS_WoS SCs**	**SAPV_ PHYS_Journals**	**WCS_ PHYS_WoS SCs**	**WCS_ PHYS_Journals**
2	1.173	1.134	0.376	0.061	0.970	0.981
3	0.123	0.154	0.358	0.052	0.845	0.814
4	0.215	0.310	0.373	0.057	0.957	0.905
5	0.017	0.030	0.098	0.018	0.004	0.010
6	0.145	0.204	0.328	0.048	0.439	0.312
7	0.208	0.087	0.301	0.048	0.492	0.628
8	0.495	0.813	0.371	0.060	0.972	0.965
9	0.120	0.146	0.358	0.052	0.846	0.809
10	0.664	0.707	0.367	0.058	0.948	0.937
11	1.195	1.045	0.172	0.030	0.849	0.914
12	0.067	0.025	0.019	0.004	0.018	0.018
13	0.109	0.159	0.038	0.016	0.080	0.255
14	0.158	0.127	0.272	0.035	0.873	0.756
15	0.118	0.063	0.054	0.013	0.194	0.292
16	0.316	0.185	0.127	0.017	0.487	0.458
17	0.443	0.668	0.115	0.028	0.457	0.766
18	0.056	0.015	0.019	0.004	0.023	0.025
19	0.688	0.600	0.133	0.024	0.503	0.632
20	1.041	0.960	0.156	0.027	0.780	0.748

Figure 43. Excerpt of the dissimilarities/distances between panel members and individual research groups according to each of the six methods.

We import data from the MS Excel file (Figure 43) to SPSS, and calculate the Spearman's rank-order correlation coefficient between the six approaches.

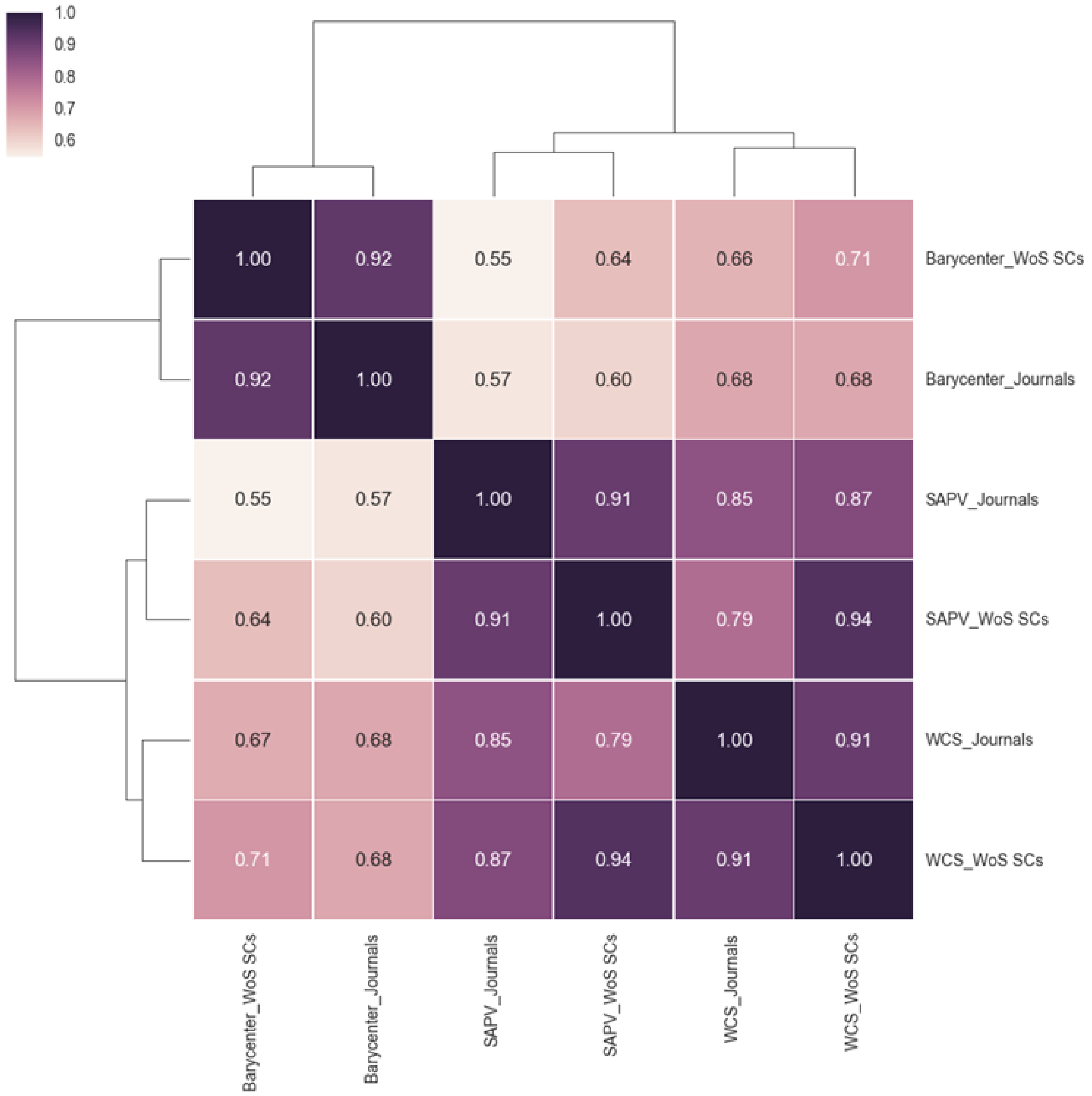


Figure 44. Heat map with hierarchical clustering based on correlation coefficient between six approaches in the Physics department

The heat map with hierarchical clustering (Figure 44) shows that correlations between the two level of analysis based on barycenter (ρ = 0.92), SAPV (ρ = 0.91) and WCS (ρ = 0.91) are strong. The correlations between the barycenter methods on the one hand and the SAPV and WCS methods on the other are moderate. In addition, correlation between SAPV and

WCS in both WoS SCs and journals are very strong. Overall, this suggests that the influence of the 2D reduction is substantial. Moreover, in general WoS SC and journal results correlate strongly. That suggests that the level of aggregation has minor influence for determining cognitive distances.

5. Programming code in Python

The essential code to calculate barycenters, similarity-adapted publication vectors, and similarity weighted cosine is as follows:

```
import numpy as np
import pandas as pd


def ensure_symmetric(M):
    m, n = M.shape
    if m != n:
        raise ValueError("M is not square!")


def barycenter(counts, coords):
    """Calculate the barycenter for the given counts and coordinates"""
    m, n = coords.shape

    if len(counts) != m:
        raise ValueError("'counts' should have the same number of items "
                         "(now: {}) as rows of 'coords' (now: {})".format(
                             len(counts), m))

    # Transposing twice because of broadcasting rules
    a = (coords.T * counts).T
    return a.sum(axis=0) / sum(counts)


def sa_vector(counts, S, normalize=True):
    """Calculate the similarity adapted vector for the given counts and
    similarity matrix S
    """
    ensure_symmetric(S)

    if len(counts) != len(S):
        raise ValueError("'counts' should have the same number of items "
                         "(now: {}) as rows of similarity matrix (now: {})"
                         .format(len(counts), len(S)))
```

```
    # Transposing twice because of broadcasting rules
    raw_sa_vector = (S.T * counts).T.sum(axis=0)
    return raw_sa_vector / raw_sa_vector.sum() if normalize else raw_sa_vector


def weighted_cosine(u, v, S):
    """Calculate cosine similarity between vectors u and v, weighted by
    similarity matrix S
    """
    ensure_symmetric(S)
    if len(u) != len(v) != len(S):
        raise ValueError("Vectors or similarity matrix of different length.")

    u = u / np.sum(u)
    v = v / np.sum(v)

    return u.dot(S).dot(v) / np.sqrt(u.dot(S).dot(u) * v.dot(S).dot(v))
```

Code to calculate top-down correlation, accounting for ties

```
from __future__ import division

import itertools
import numpy as np
from operator import itemgetter


def savage_score(rank, n, endrank=None):
    """Calculate savage score for given rank in list of n items

    If endrank is given, return array of savage scores for all items between
    rank and endrank.

    """
    if rank < 1 or rank > n:
        raise ValueError("rank should be between 1 and n")

    if not hasattr(savage_score, 'lookup') or n != savage_score.n:
        savage_score.n = n
        arr = np.cumsum([1 / i for i in xrange(n, 0, -1)])
        savage_score.lookup = arr[::-1]

    if endrank is not None:
        return savage_score.lookup[rank - 1:endrank - 1]
    else:
        return savage_score.lookup[rank - 1]


def avg_savage_score(start, length, n):
    return np.average(savage_score(start, n, start + length))


def _ties(values):
    """Find ties in list of values"""
```

```
    prev = None
    ties = []
    start = 0
    final = object()

    # We add an element 'final' at the end, to ensure that the last entry is
    # also properly handled.
    for rank, value in enumerate(itertools.chain(values, [final]), start=1):
        if value == prev:
            if start == 0:  # start of a tie
                start = rank - 1
        else:
            if start != 0:  # end of a tie
                ties.append((start, rank - start))
                start = 0
        prev = value

    return ties


def savage_scores_with_ties(values):

    def next_tie():
        try:
            return ties.pop(0)
        except IndexError:
            return -1, -1

    n = len(values)
    ties = _ties(values)
    tierank, tielength = next_tie()

    for rank in range(1, n + 1):
        value = values[rank - 1]
        if rank >= tierank and rank < tierank + tielength:
            yield avg_savage_score(tierank, tielength, n), value
        else:
            if rank == tierank + tielength:
                tierank, tielength = next_tie()
            yield savage_score(rank, n), value


def dict_with_savage_scores(d):
    # If d is a ranked list of items, convert it to a dict with decreasing
    # values.
    if isinstance(d, list):
        d = dict(zip(d, range(len(d), 0, -1)))

    d_sorted = sorted(d.iteritems(), reverse=True, key=itemgetter(1))
    items, values = zip(*d_sorted)
    return {item: rank for item, (rank, value)
            in zip(items, savage_scores_with_ties(values))}


def top_down_correlation(R, Q):
    n = len(R)
    assert len(Q) == n

    R_scores = dict_with_savage_scores(R)
```

```
    Q_scores = dict_with_savage_scores(Q)

    return (sum(R_scores[item] * Q_scores[item] for item in R_scores) - n) / \
        (n - savage_score(1, n))
```

References

Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling*. New York, NY: Springer New York.

de Nooy, W., Mrvar, A., & Batagelj, V. (2012). *Exploratory Social Network Analysis with Pajek* (2nd edition). England ; New York: Cambridge University Press.

Efron, B., & Tibshirani, R. J. (1998). *An Introduction to the Bootstrap*. London: Chapman & Hall/CRC.

Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics*. Elsevier Science Publishers. Retrieved from https://uhdspace.uhasselt.be/dspace/handle/1942/587

Huson, L. W. (2007). Performance of some correlation coefficients when applied to zero-clustered data. *Journal of Modern Applied Statistical Methods*, *6*(2), 530–536.

Iman, R. L., & Conover, W. J. (1987). A measure of Top-Down correlation. *Technometrics*, *29*(3), 351–357.

Jin, B., & Rousseau, R. (2001). An introduction to the barycentre method with an application to China's mean centre of publication. *Libri*, *51*(4), 225–233.

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

Kington, J. (2014). Balanced Cross Sections, Shortening Estimates, and the Magnitude of Out-of-Sequence Thrusting in the Nankai Trough Accretionary Prism, Japan. *Figshare*. https://doi.org/10.6084/m9.figshare.1015774.v1

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.

Leydesdorff, L., Rafols, I., & Chen, C. (2013). Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal–journal citations. *Journal of the American Society for Information Science and Technology*, *64*(12), 2573–2586.

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. (2016). Measuring the match between evaluators and evaluees: cognitive distances between panel members and research groups at the journal level. *Scientometrics*.

Rousseau, R. (1989). Kinematical statistics of scientific output. Part II: standardized polygonal approach. *Revue Française de Bibliométrie*, *4*, 65–77.

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

van Eck, N. J., & Waltman, L. (2007). VOS: A New Method for Visualizing Similarities Between Objects. In R. Decker & H.-J. Lenz (Eds.), *Advances in Data Analysis* (pp. 299–306). Springer Berlin Heidelberg.

van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS. *Journal of the American Society for Information Science and Technology*, *61*(12), 2405–2416.

Verleysen, F. T., & Engels, T. C. E. (2013). Measuring internationalisation of book publishing in the social sciences and humanities using the barycentre method. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Horlesberger, & H. Moed (Eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference (ISSI)*, *15 - 19 July 2013* (pp. 1170–1176). Vienna, Austria.

Verleysen, F. T., & Engels, T. C. E. (2014). Barycenter representation of book publishing internationalization in the Social Sciences and Humanities. *Journal of Informetrics*, *8*(1), 234–240.

VSNU. (2003). *Standard Evaluation Protocol 2003-2009 for Public Research Organisations*. Utrecht/den Haag/Amsterdam: VSNU, NWO and KNAW.

VSNU, KNAW, & NWO. (2014). *Standard evaluation Protocol 2015-2021: protocol for research assessment in The Netherlands*. Utrecht/den Haag/Amsterdam: VSNU, NWO and KNAW.

Appendix A

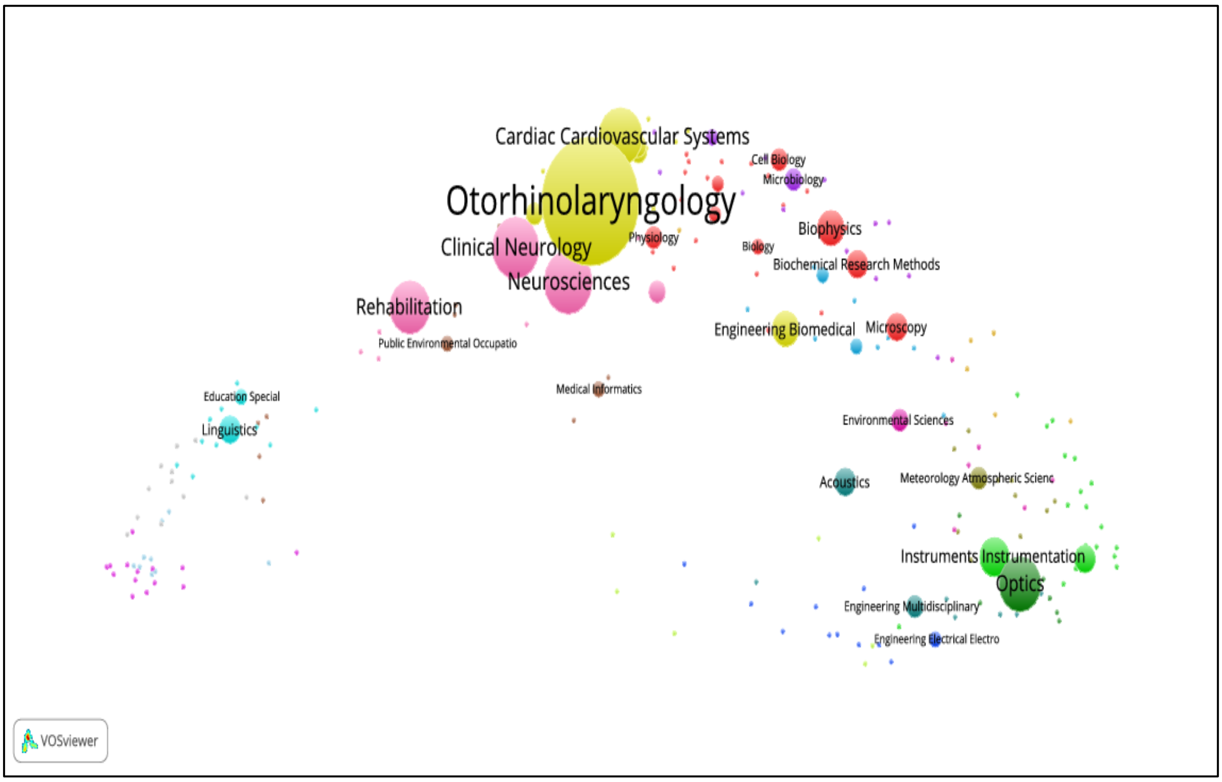


Figure 45. Figure 43. WoS SCs overlay map of PHYS-A research group's publications

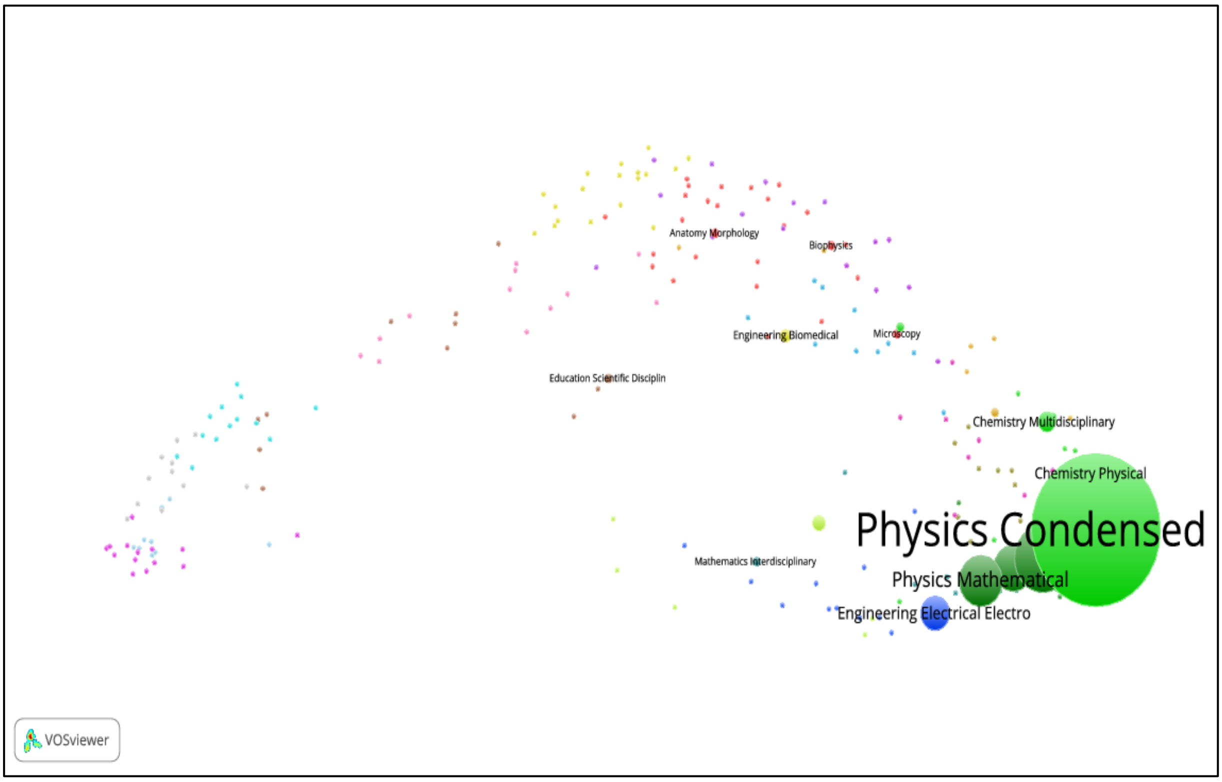


Figure 46. WoS SCs overlay map of PHYS-B research group's publications

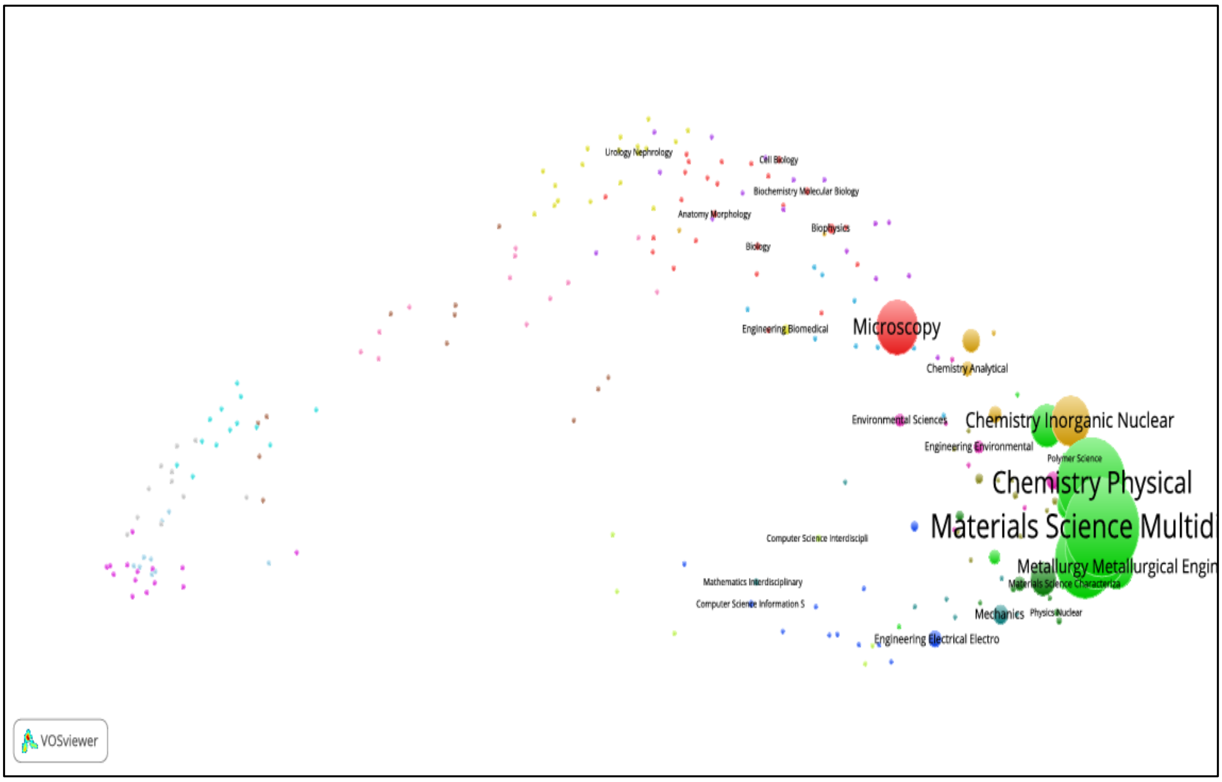


Figure 47.WoS SCs overlay map of PHYS-C research group's publications

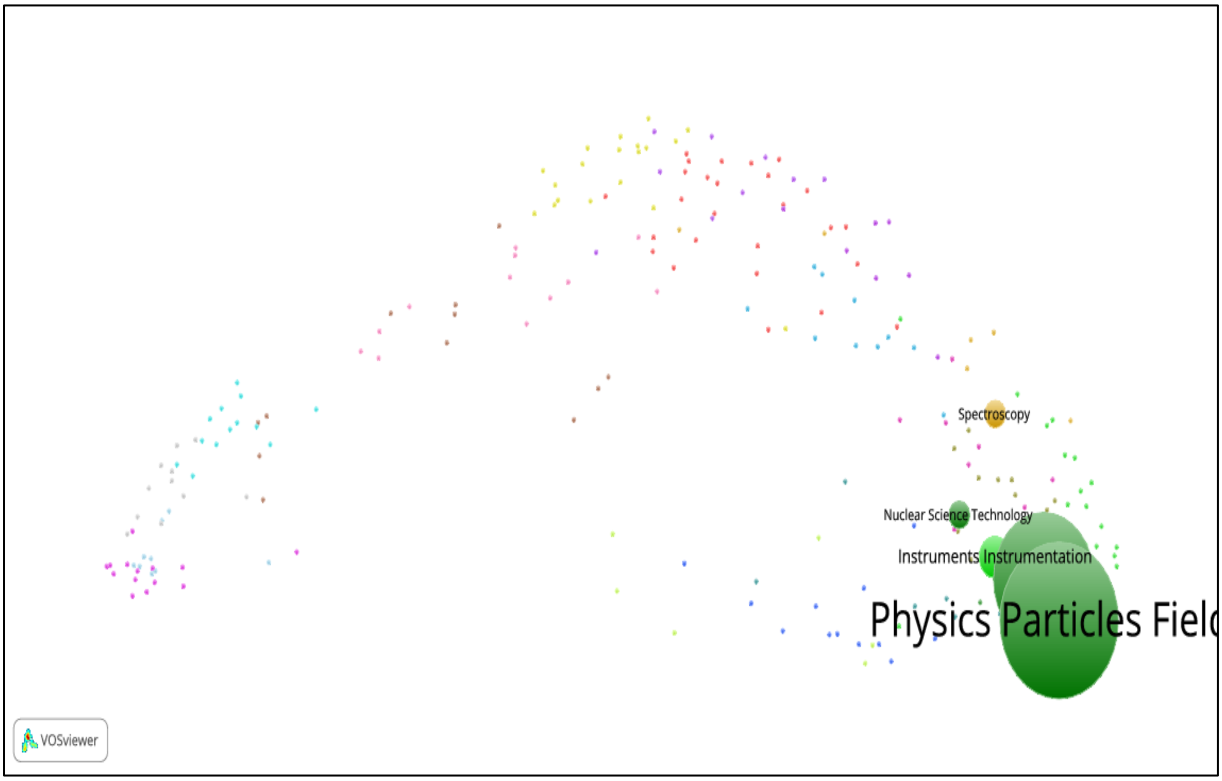


Figure 48. WoS SCs overlay map of PHYS-D research group's publications

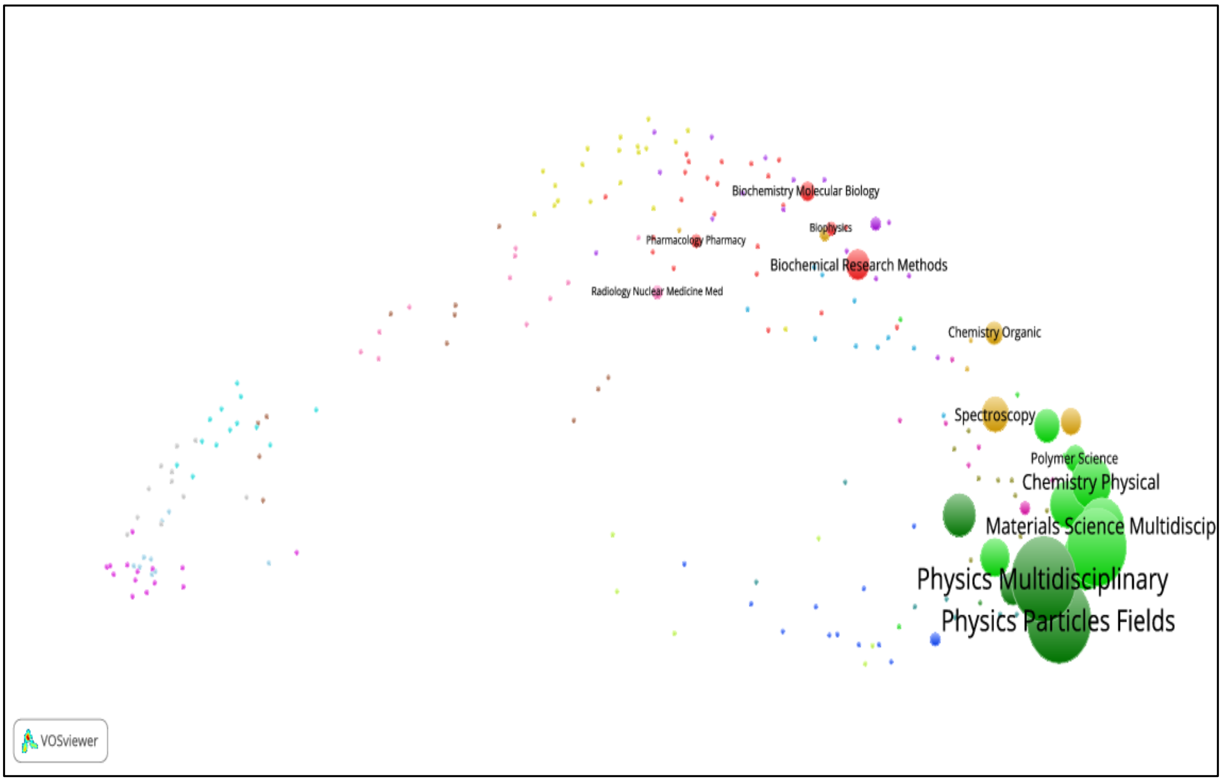


Figure 49. WoS SCs overlay map of PHYS-E research group's publications

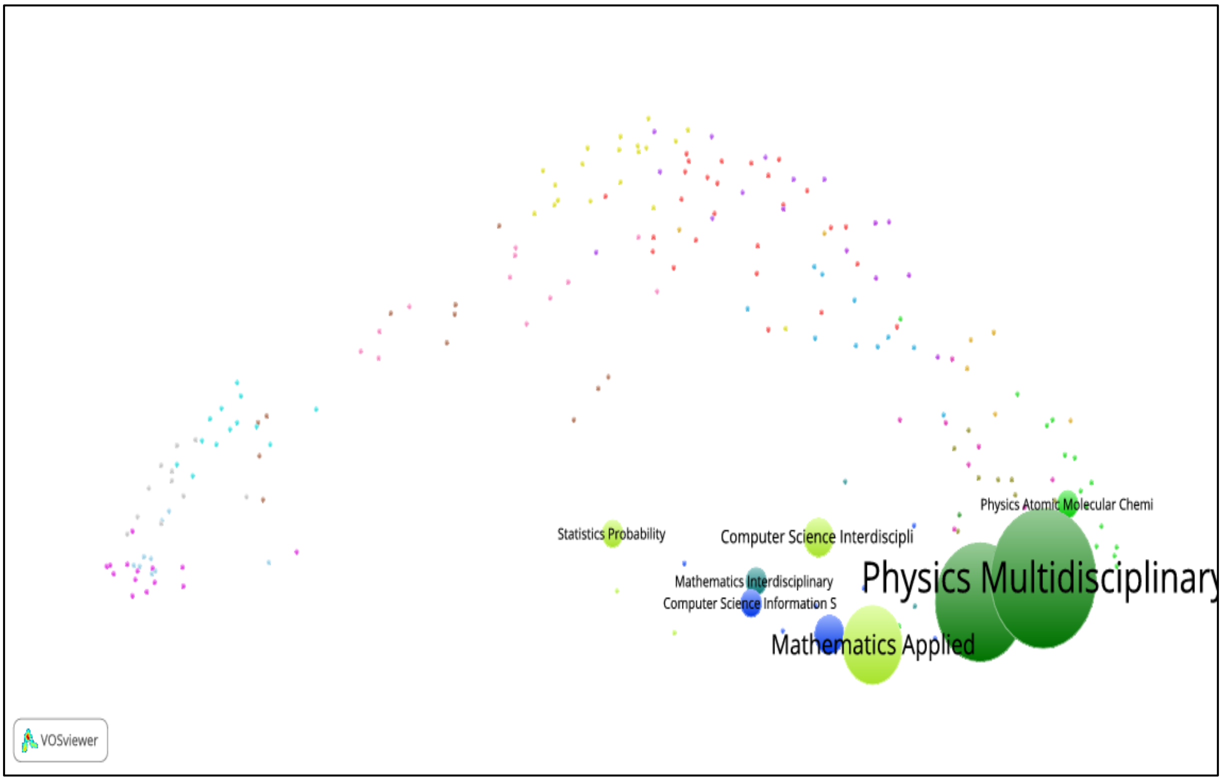


Figure 50. WoS SCs overlay map of PHYS-F research group's publications

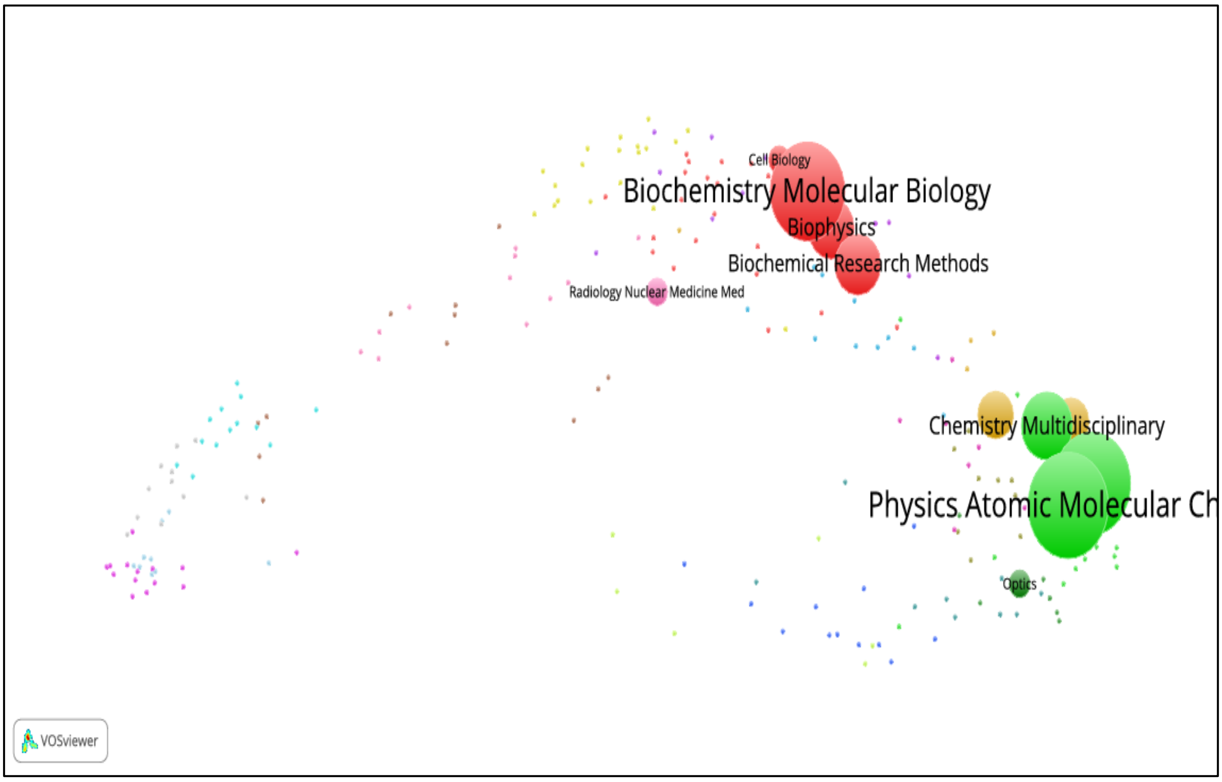


Figure 51. WoS SCs overlay map of PHYS-G research group's publications

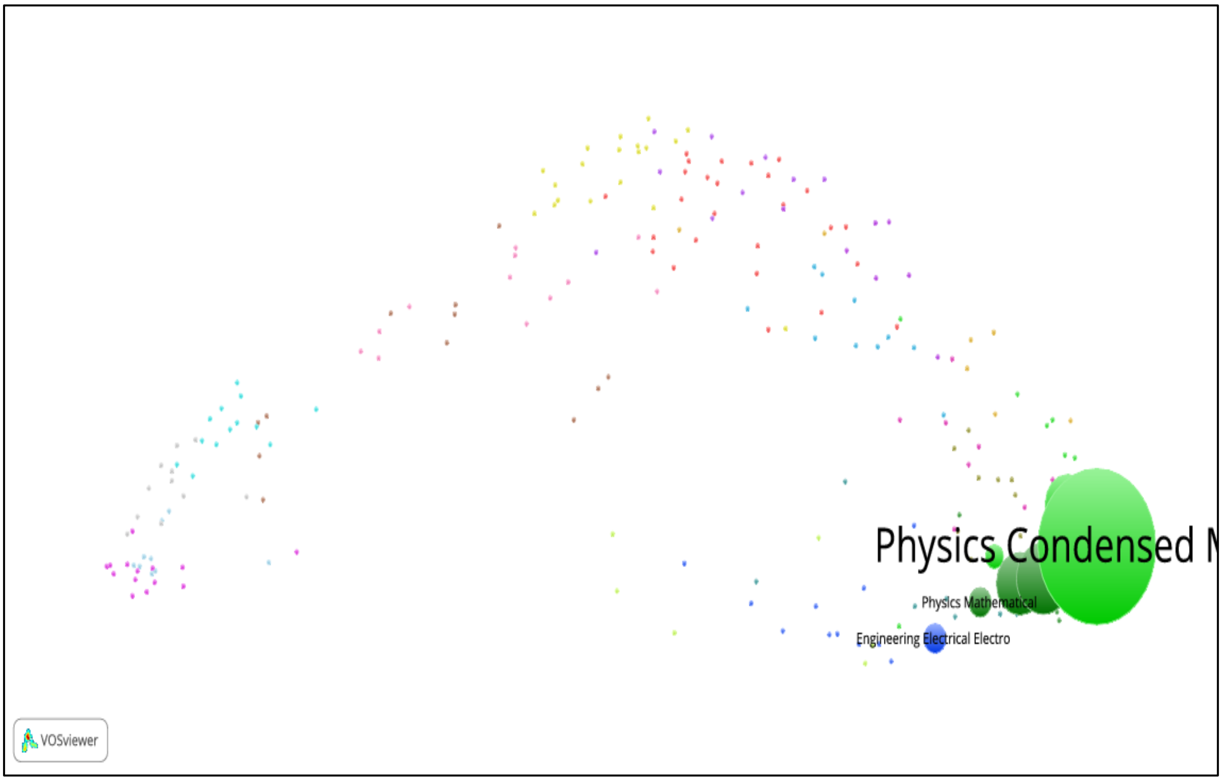


Figure 52. WoS SCs overlay map of PHYS-H research group's publications

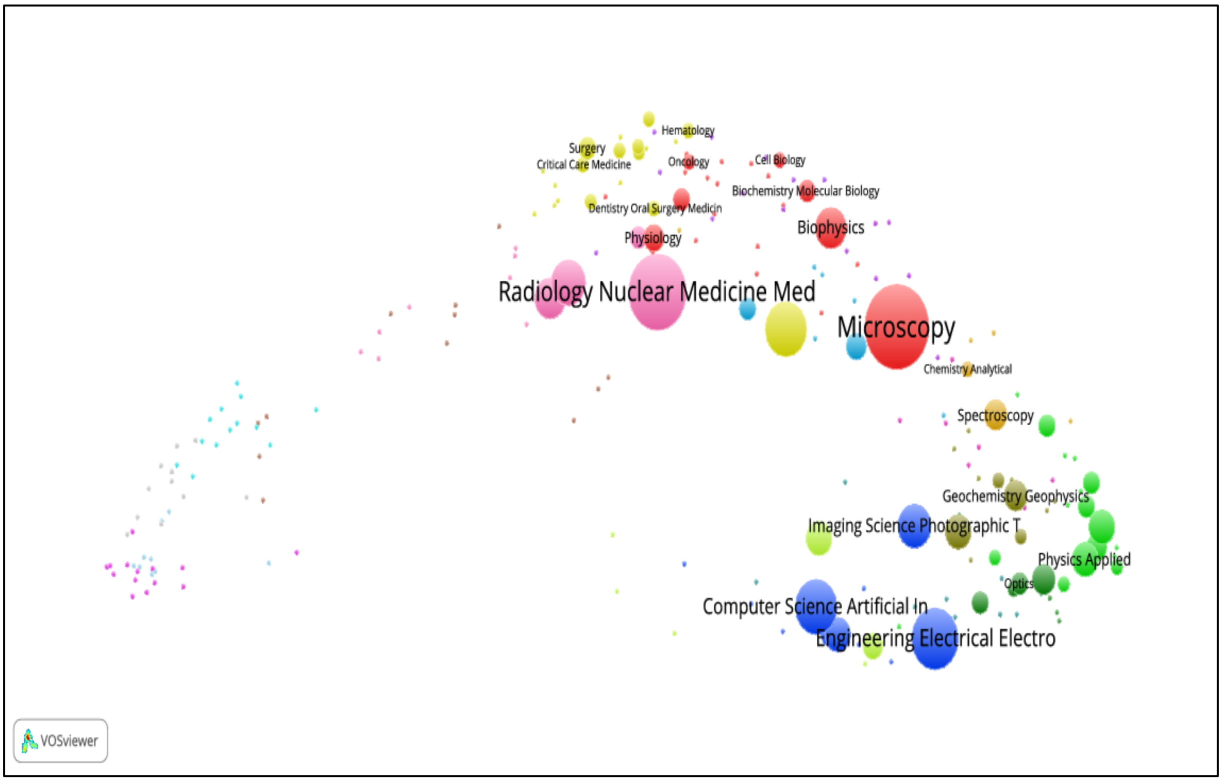


Figure 53. WoS SCs overlay map of PHYS-I research group's publications

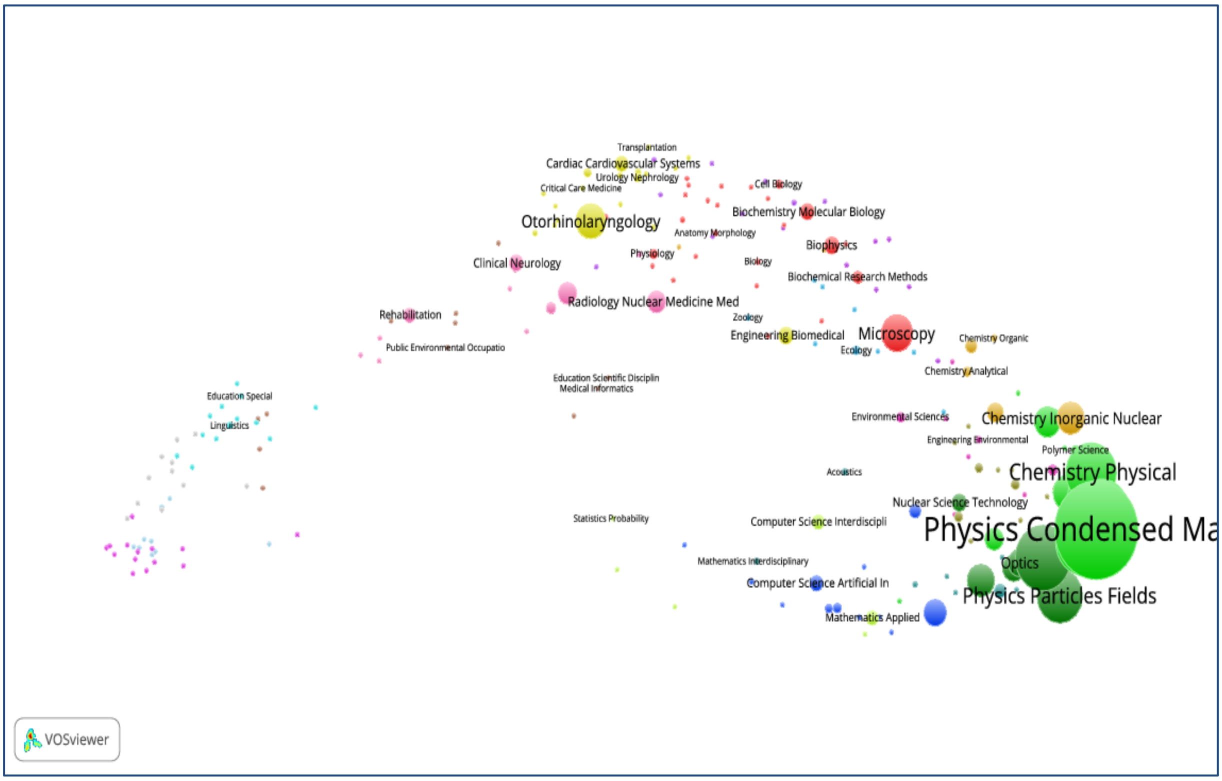


Figure 54. WoS SCs overlay map of Physics groups' publications

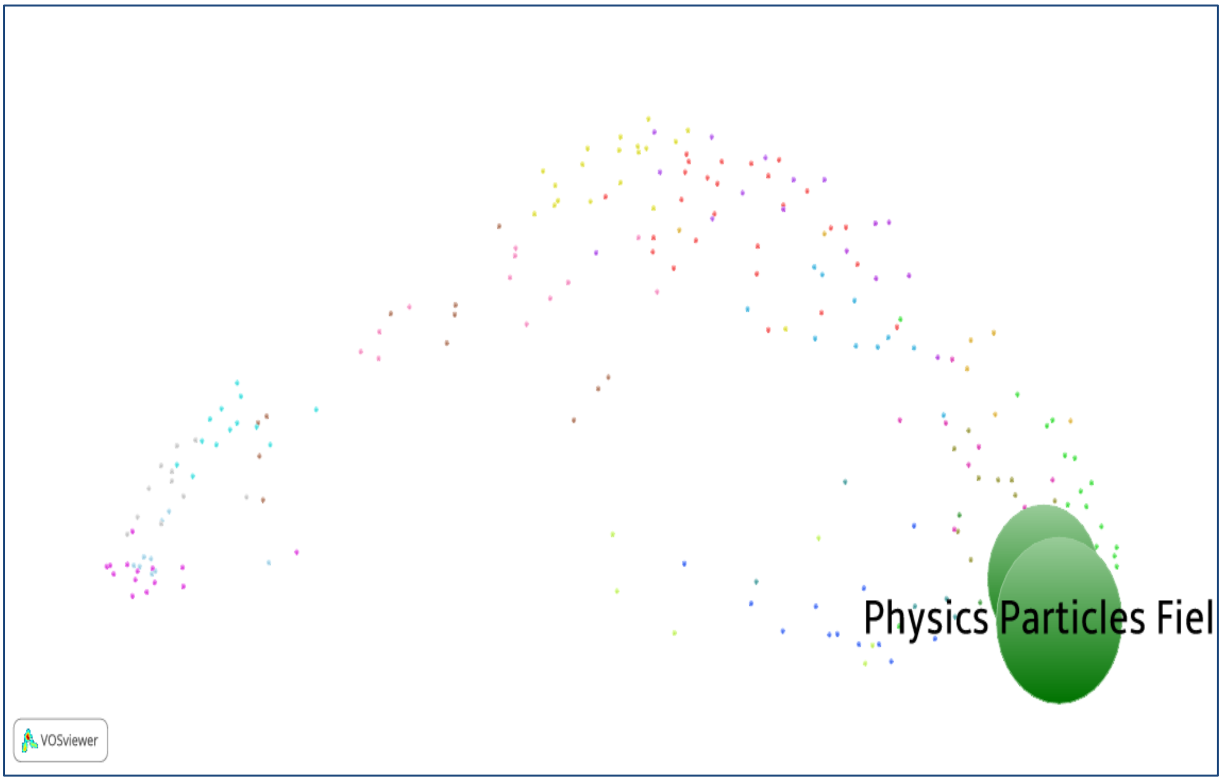


Figure 55. WoS SCs overlay map of PM1's publications

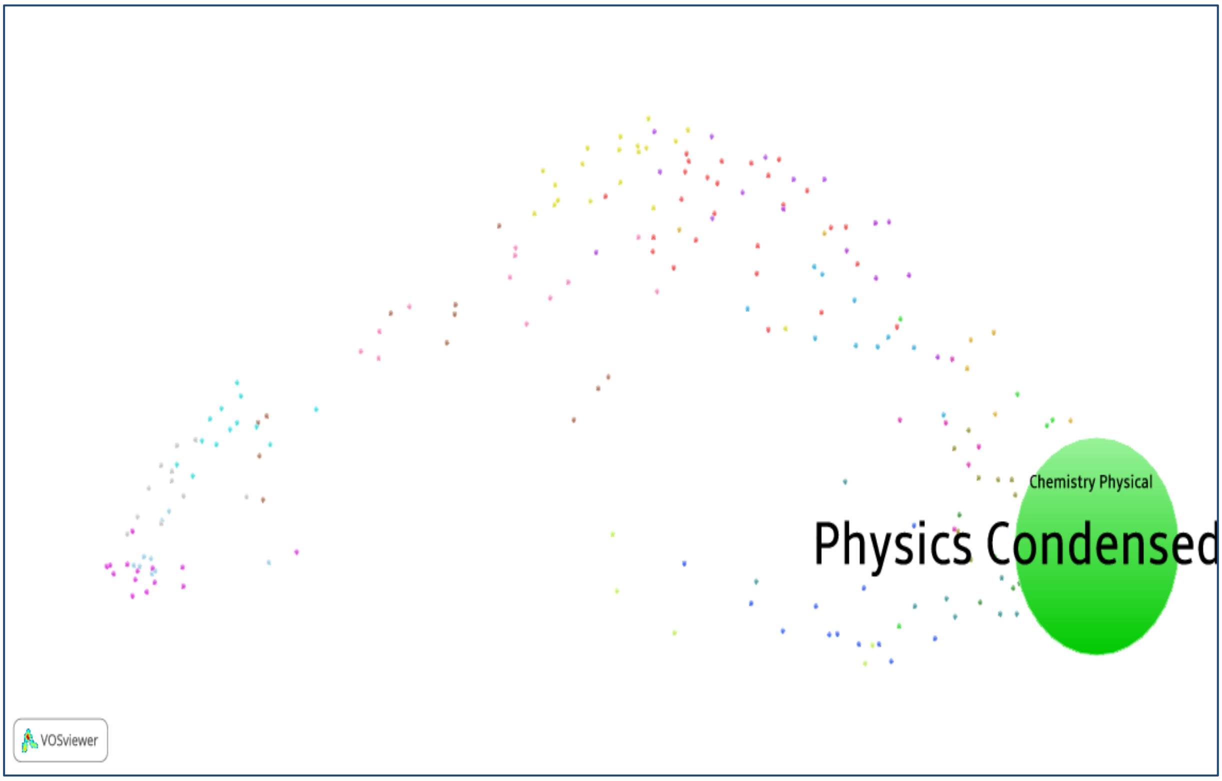


Figure 56. WoS SCs overlay map of PM2's publications

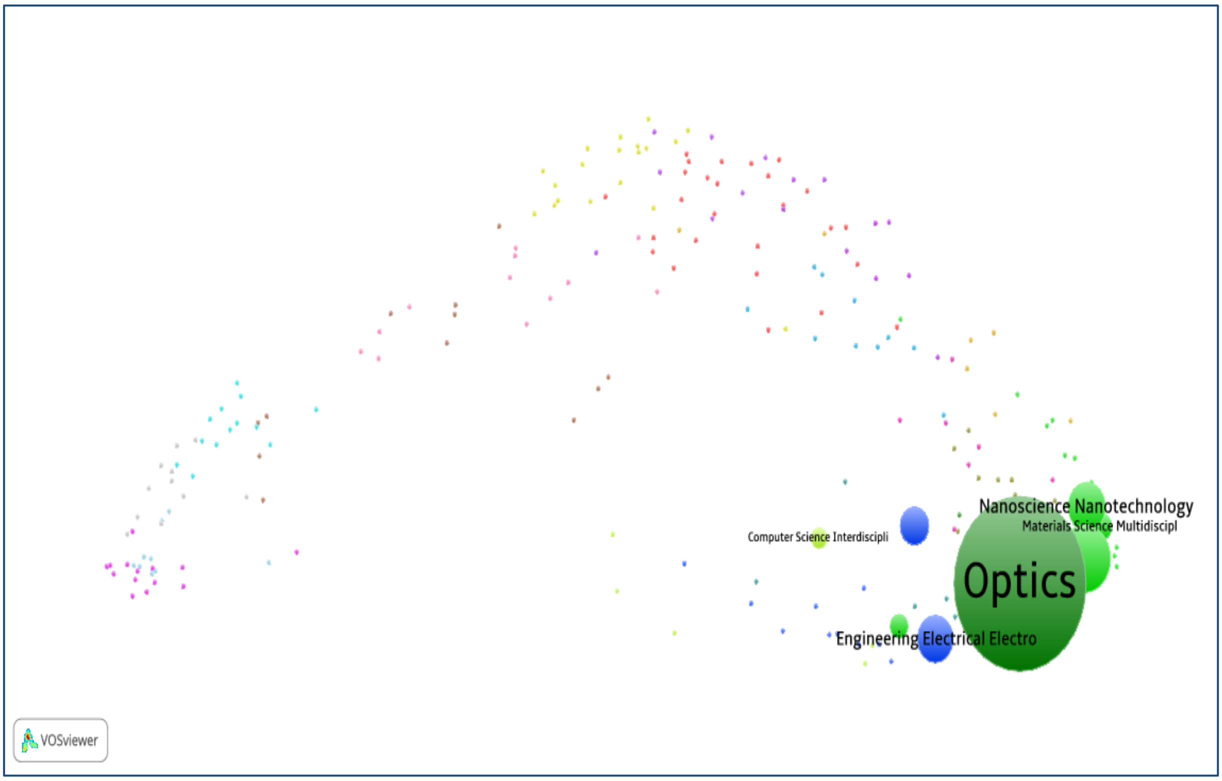


Figure 57. WoS SCs overlay map of PM3's publications

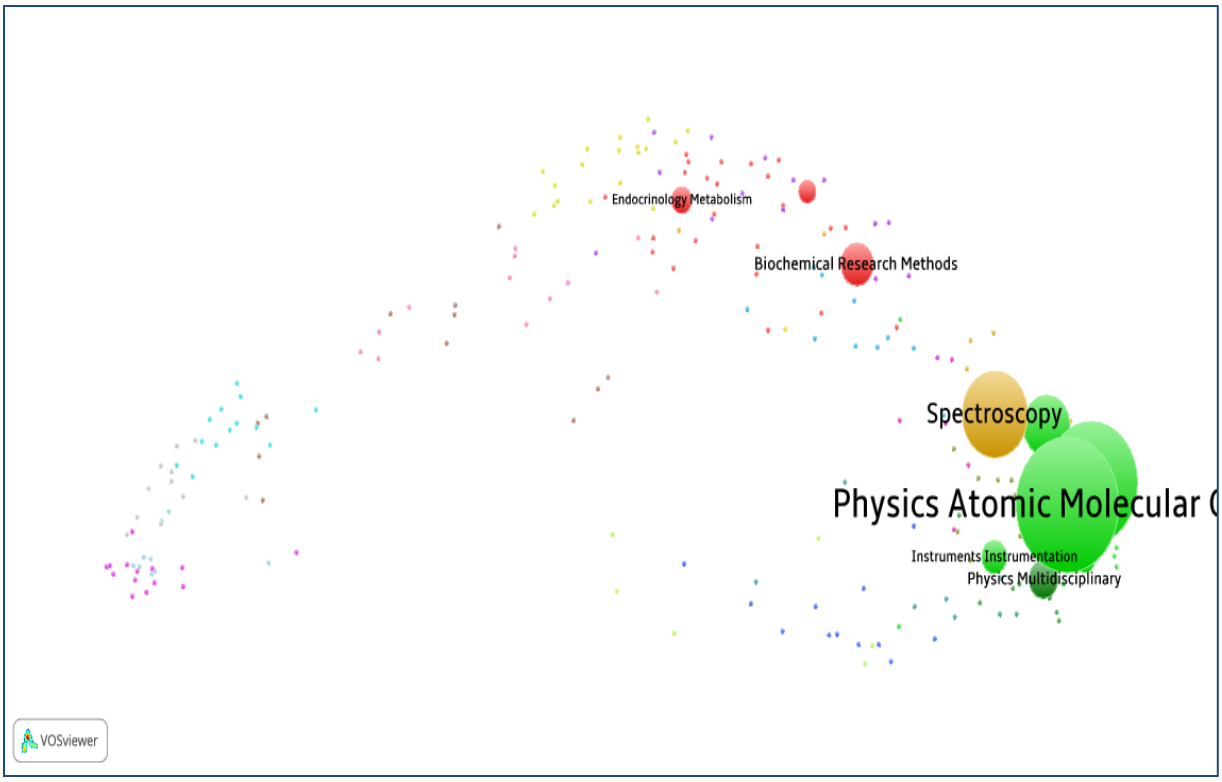


Figure 58. WoS SCs overlay map of PM4's publications

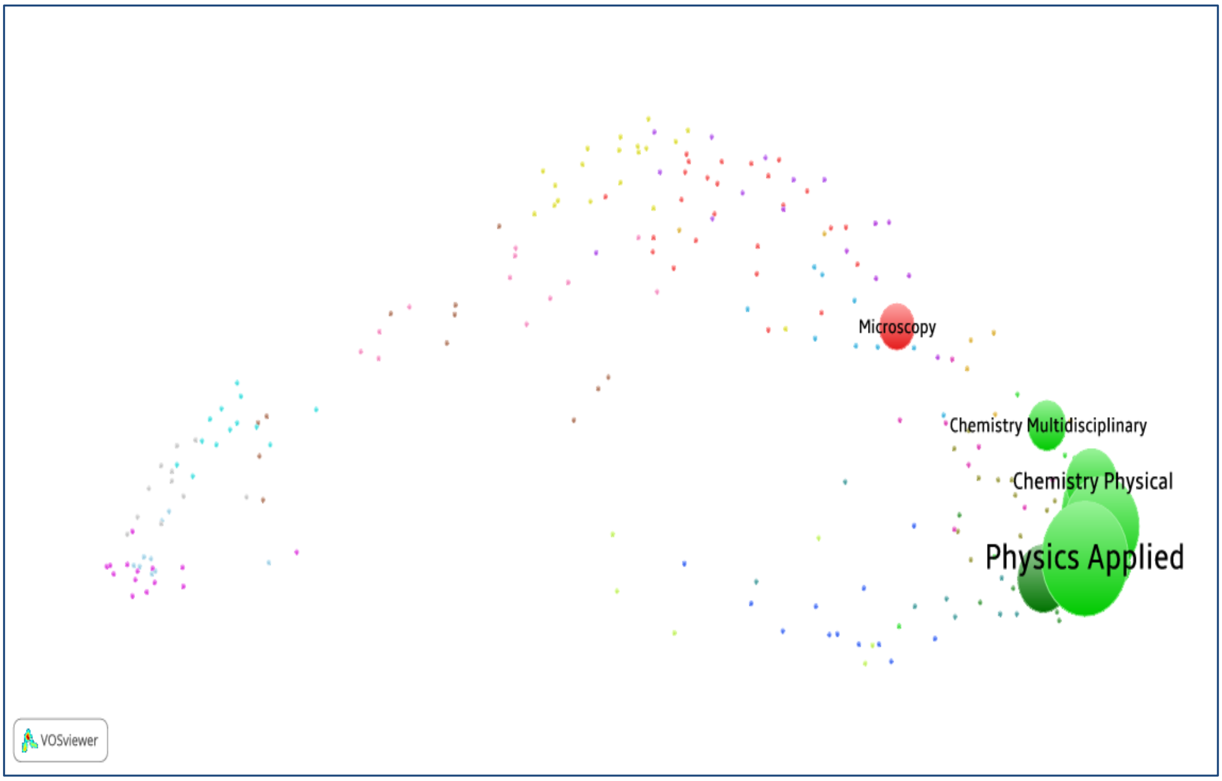


Figure 59. WoS SCs overlay map of PM5's publications

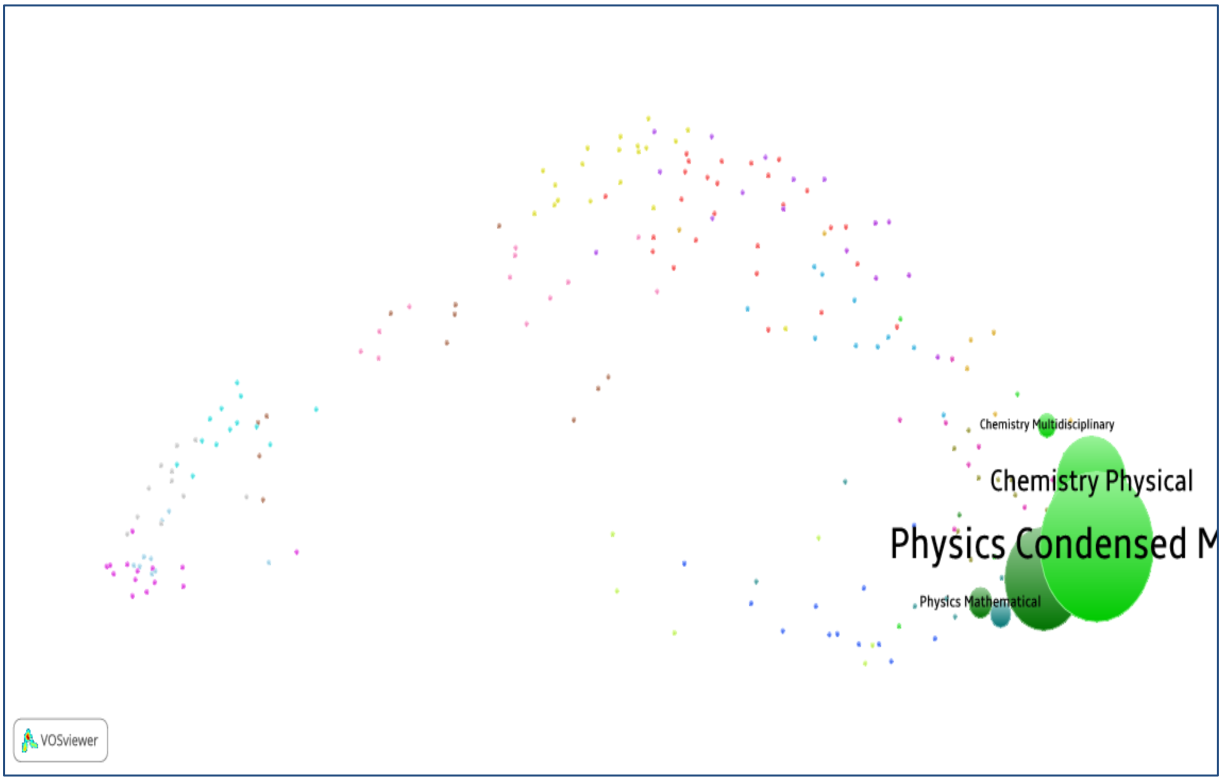


Figure 60. WoS SCs overlay map of PM6's publications

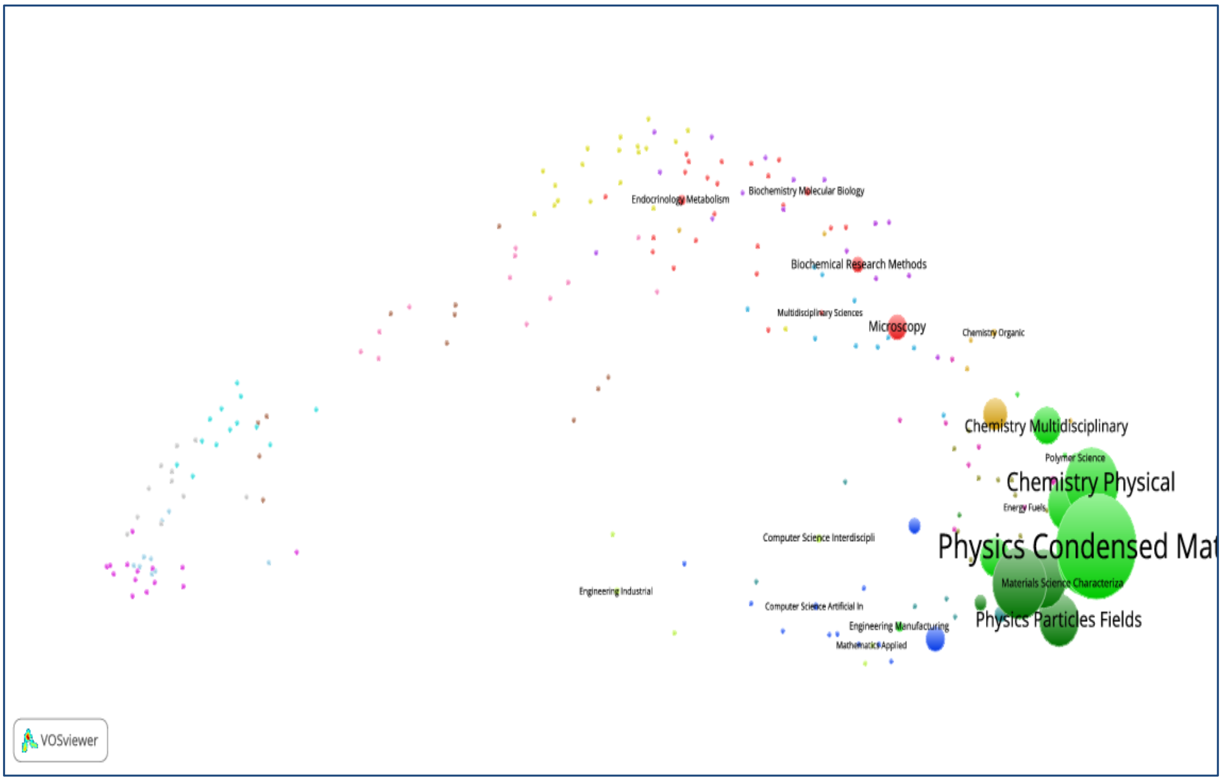


Figure 61. WoS SCs overlay map of panel's publications

Appendix B

Figure 62. SAPV of the PHYS-A research group's publications in WoS SCs similarity matrix

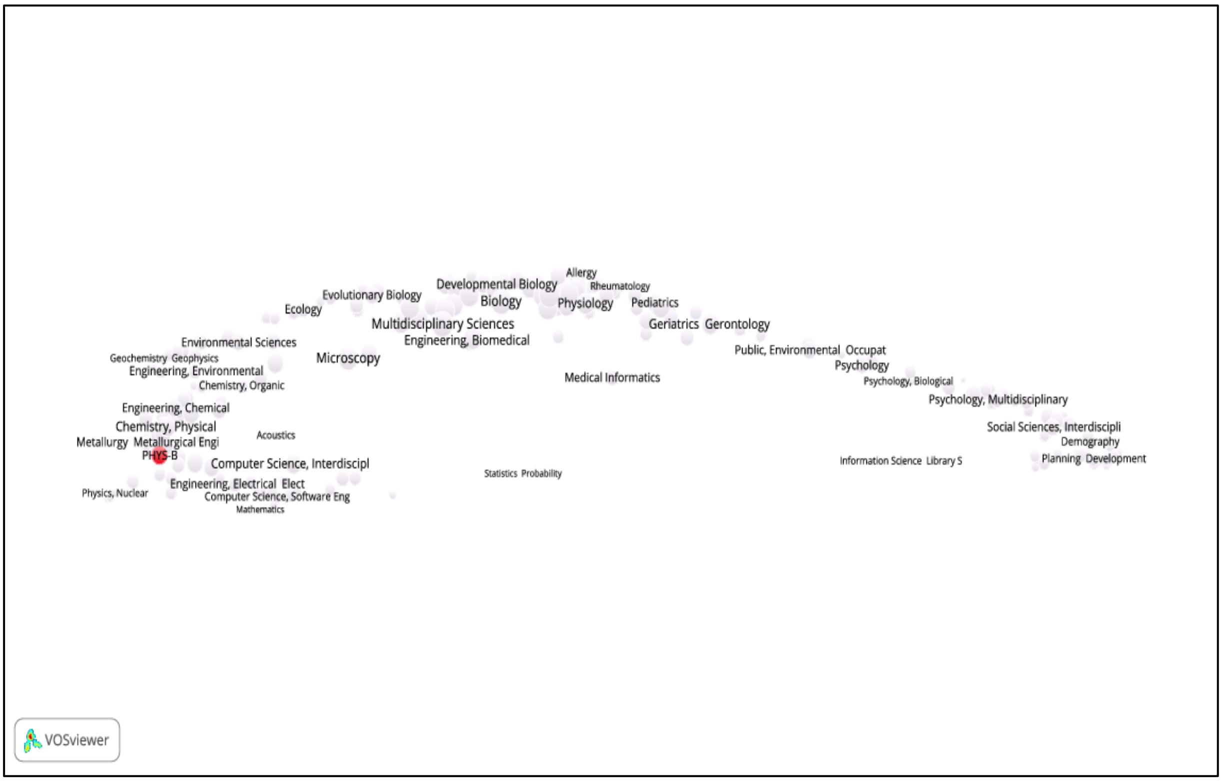


Figure 63. SAPV of the PHYS-B research group's publications in WoS SCs similarity matrix

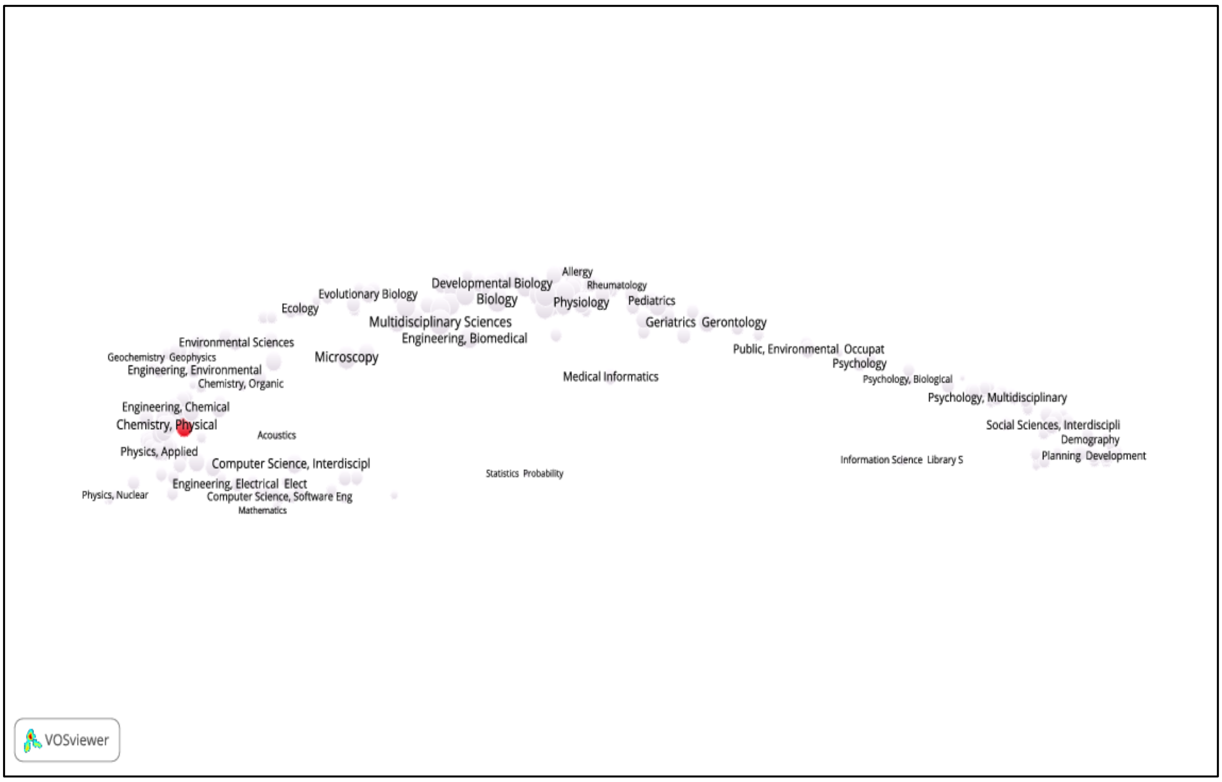


Figure 64. SAPV of the PHYS-C research group's publications in WoS SCs similarity matrix

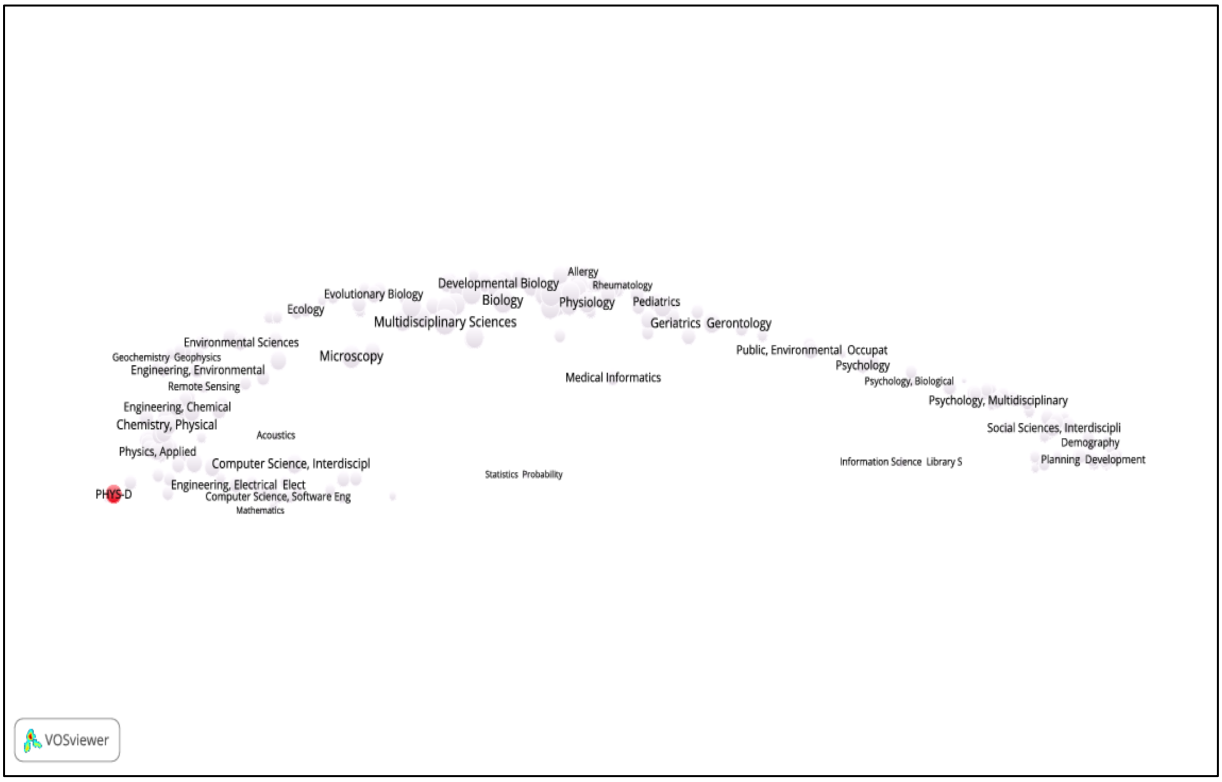


Figure 65. SAPV of the PHYS-D research group's publications in WoS SCs similarity matrix

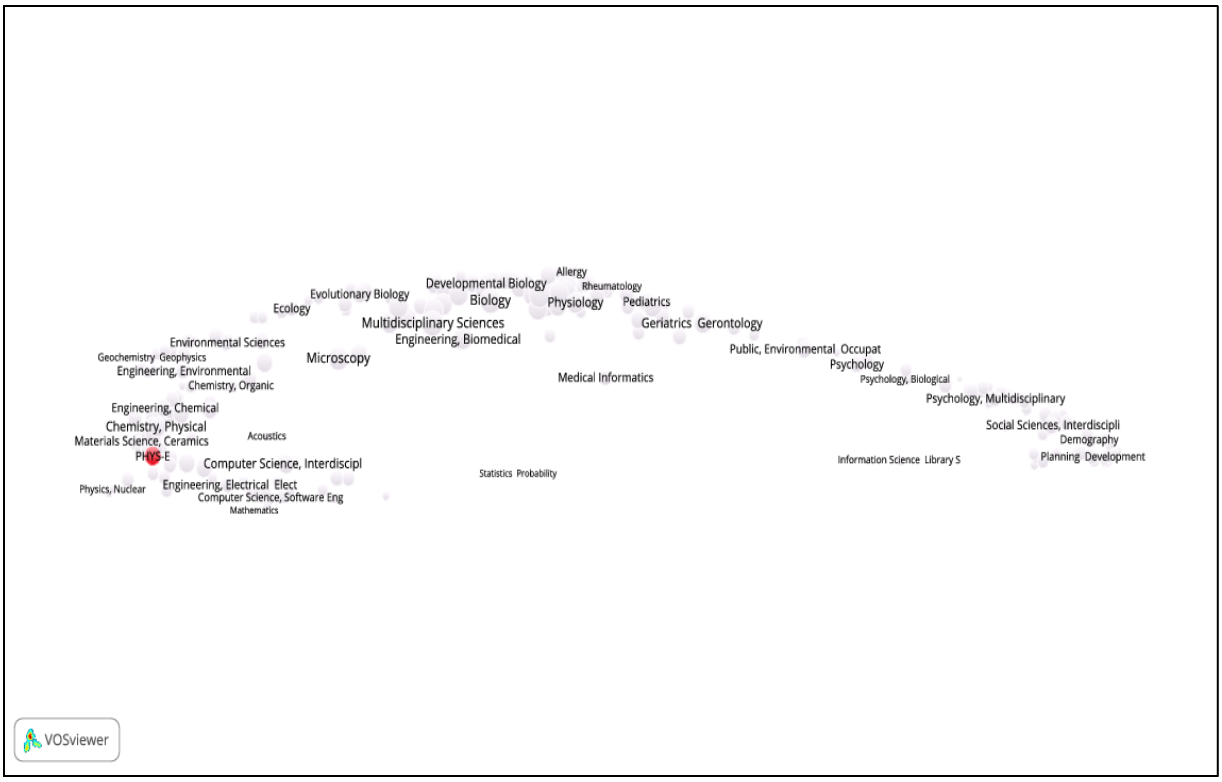


Figure 66. SAPV of the PHYS-E research group's publications in WoS SCs similarity matrix

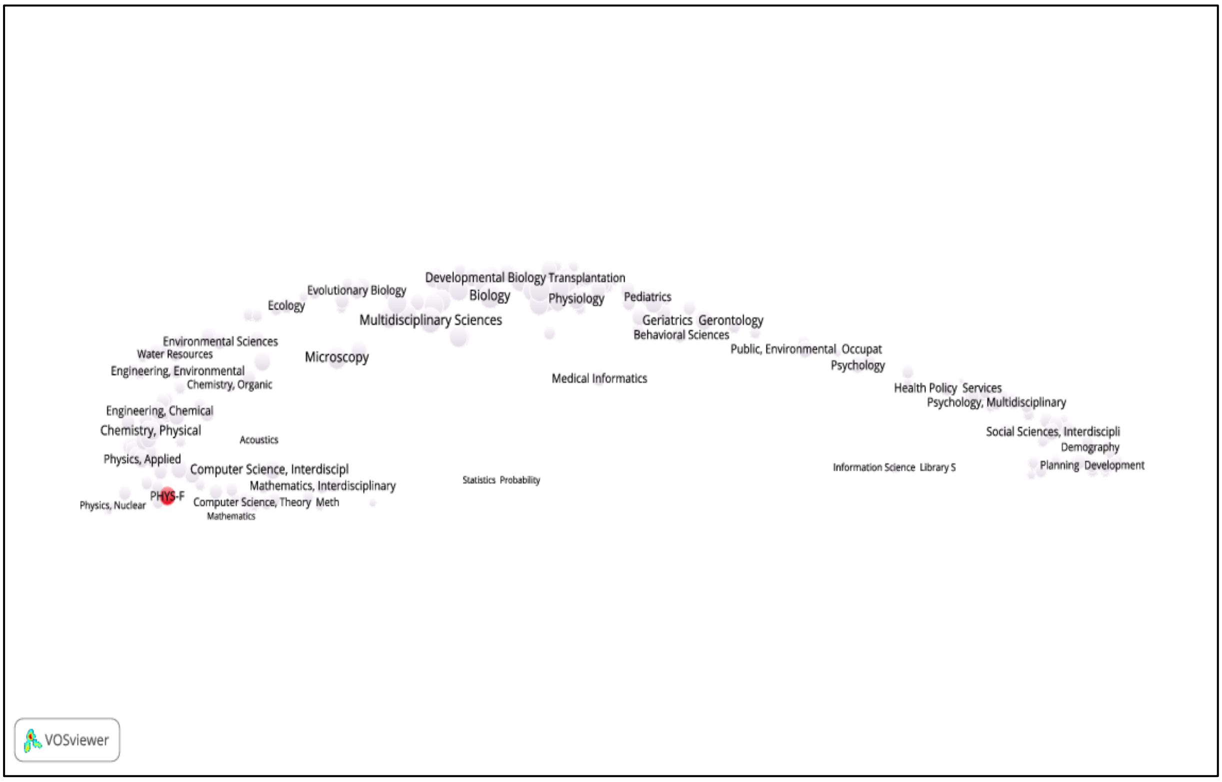


Figure 67. SAPV of the PHYS-F research group's publications in WoS SCs similarity matrix

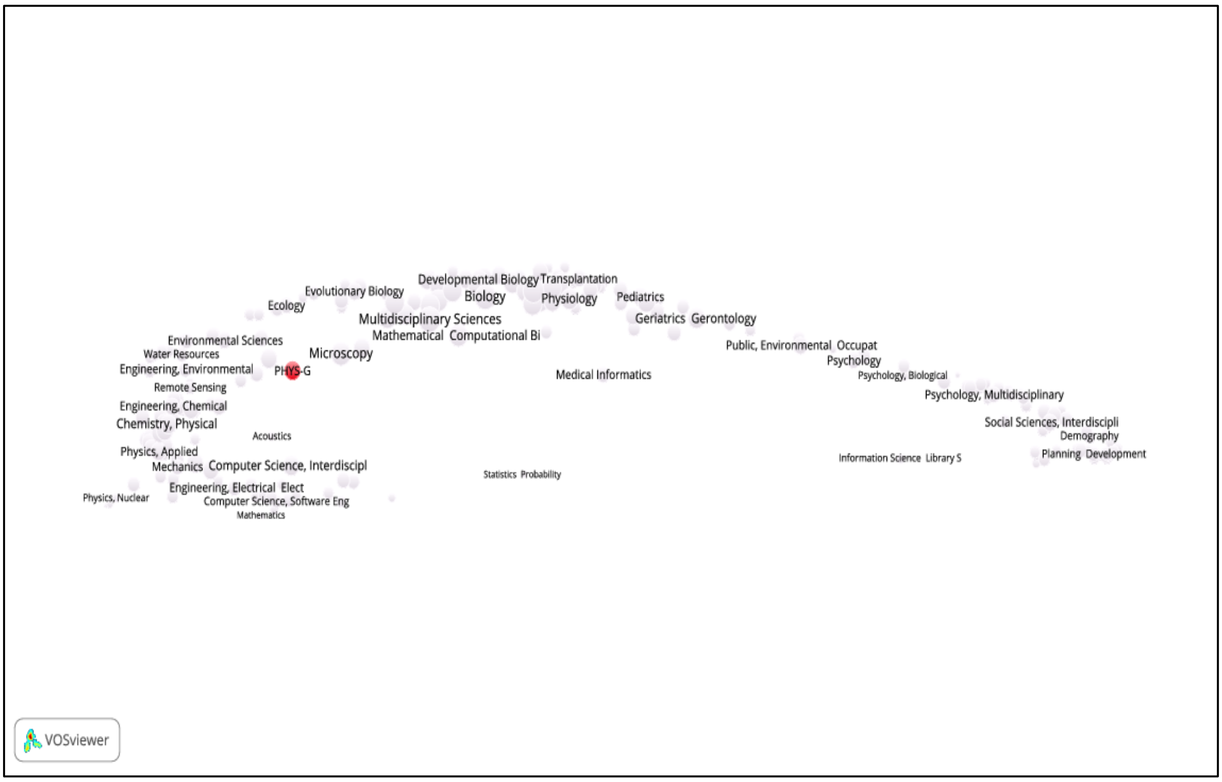


Figure 68. SAPV of the PHYS-G research group's publications in WoS SCs similarity matrix

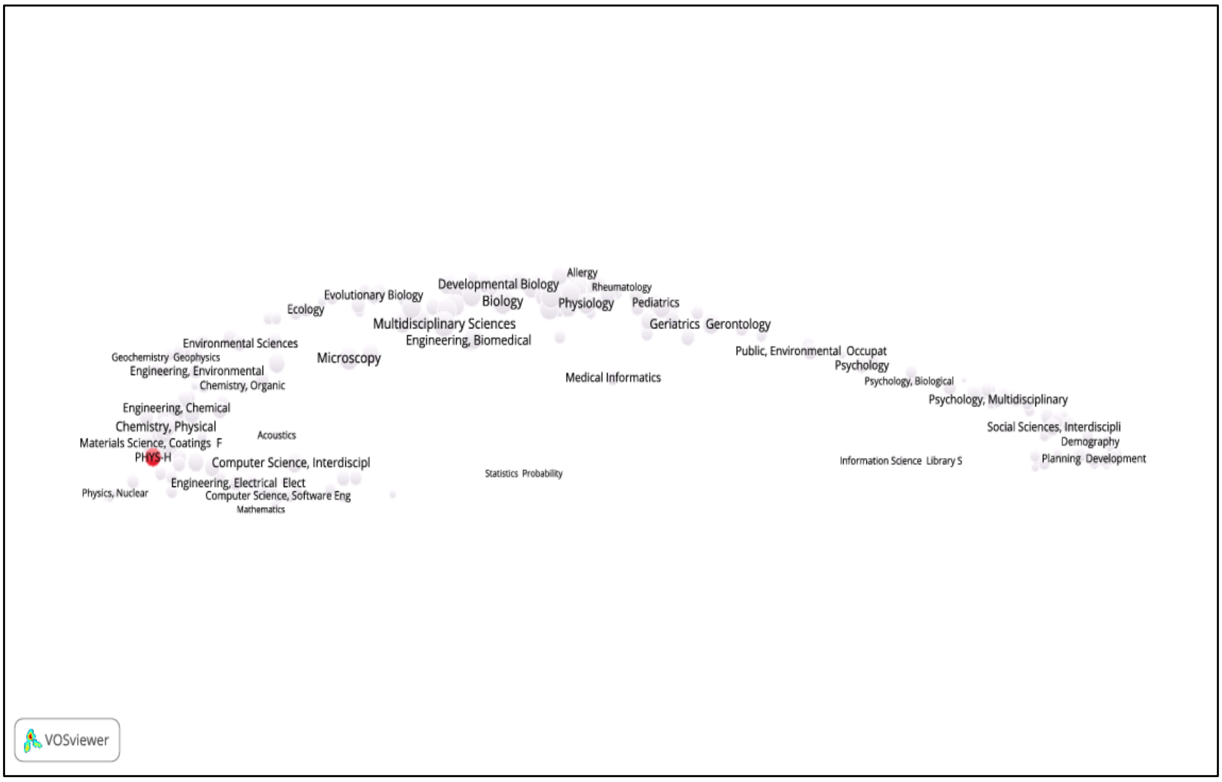


Figure 69. SAPV of the PHYS-H research group's publications in WoS SCs similarity matrix

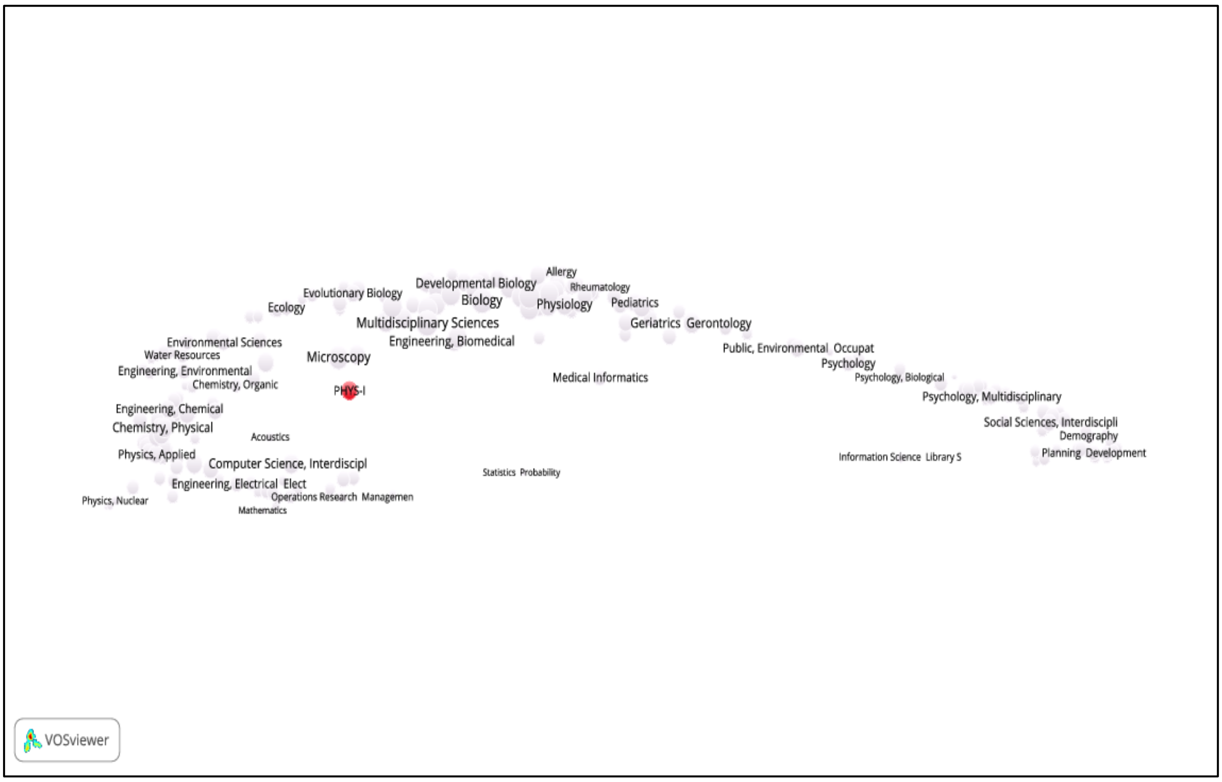


Figure 70. SAPV of the PHYS-I research group's publications in WoS SCs similarity matrix

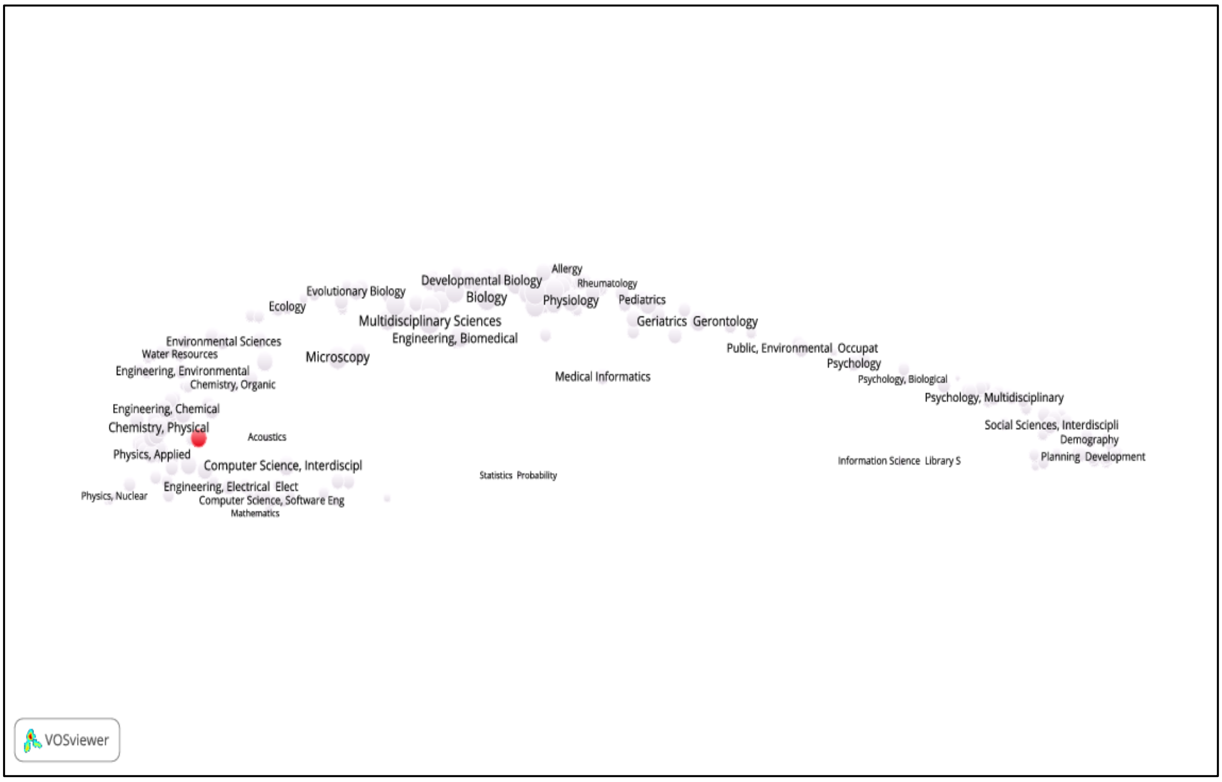


Figure 71. SAPV of the Physics research group's publications in WoS SCs similarity matrix

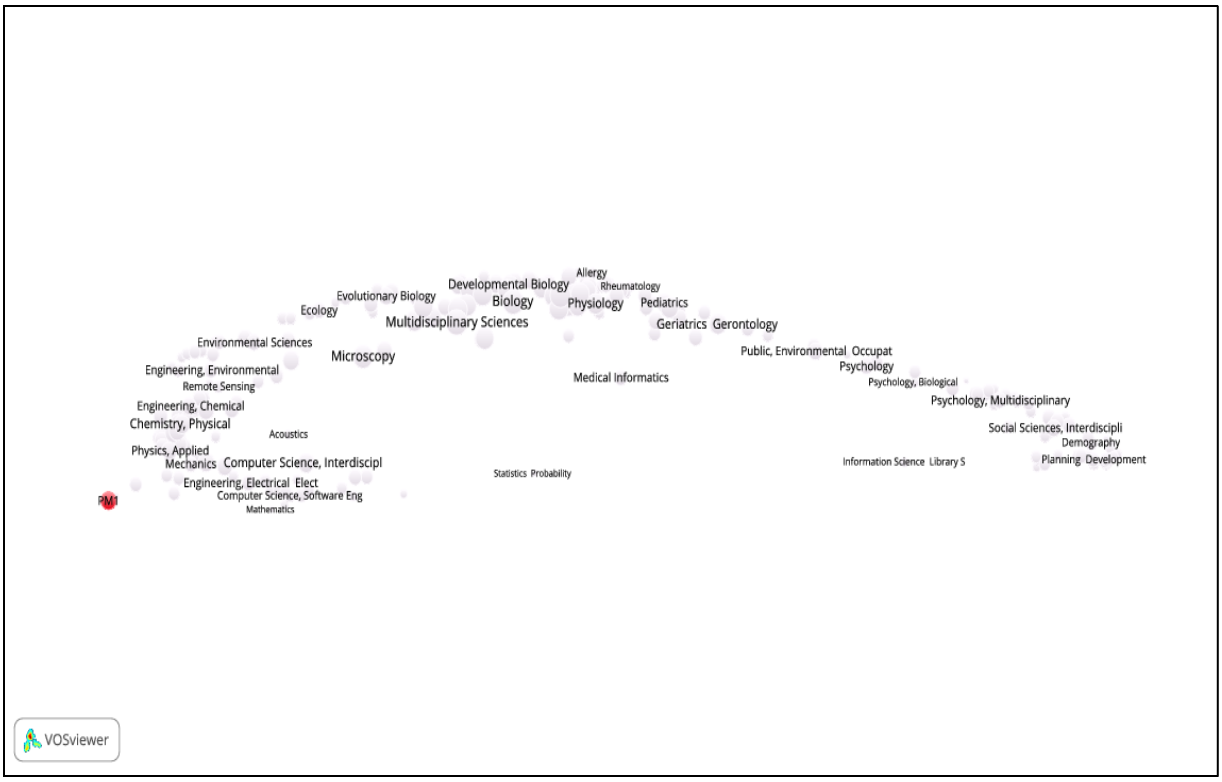


Figure 72. SAPV of the PM1's publications in WoS SCs similarity matrix

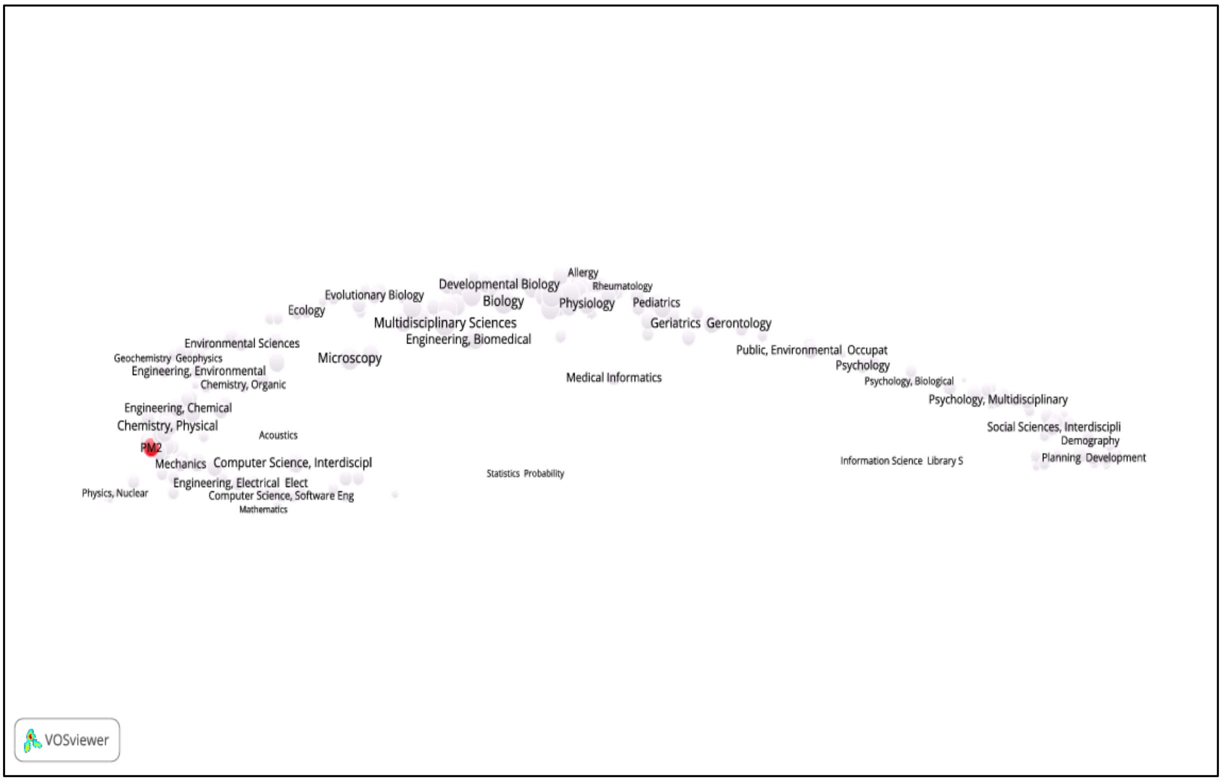


Figure 73. SAPV of the PM2's publications in WoS SCs similarity matrix

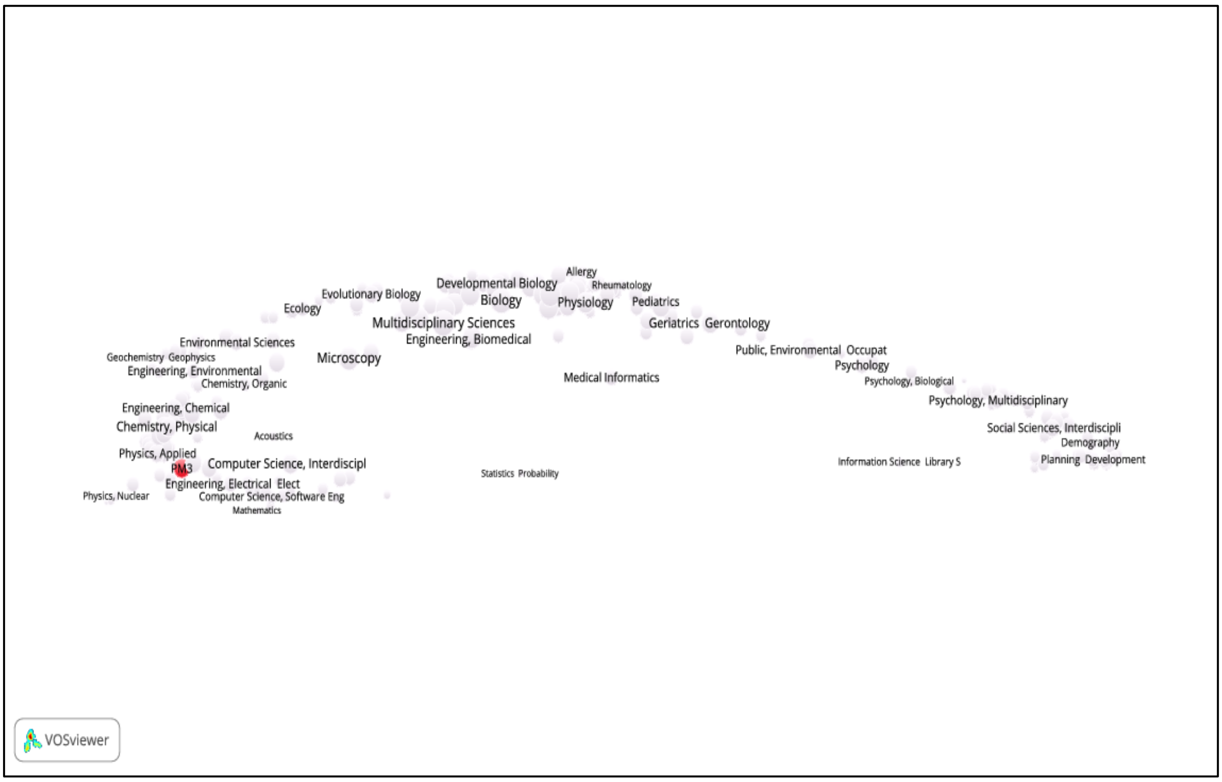


Figure 74. SAPV of the PM3's publications in WoS SCs similarity matrix

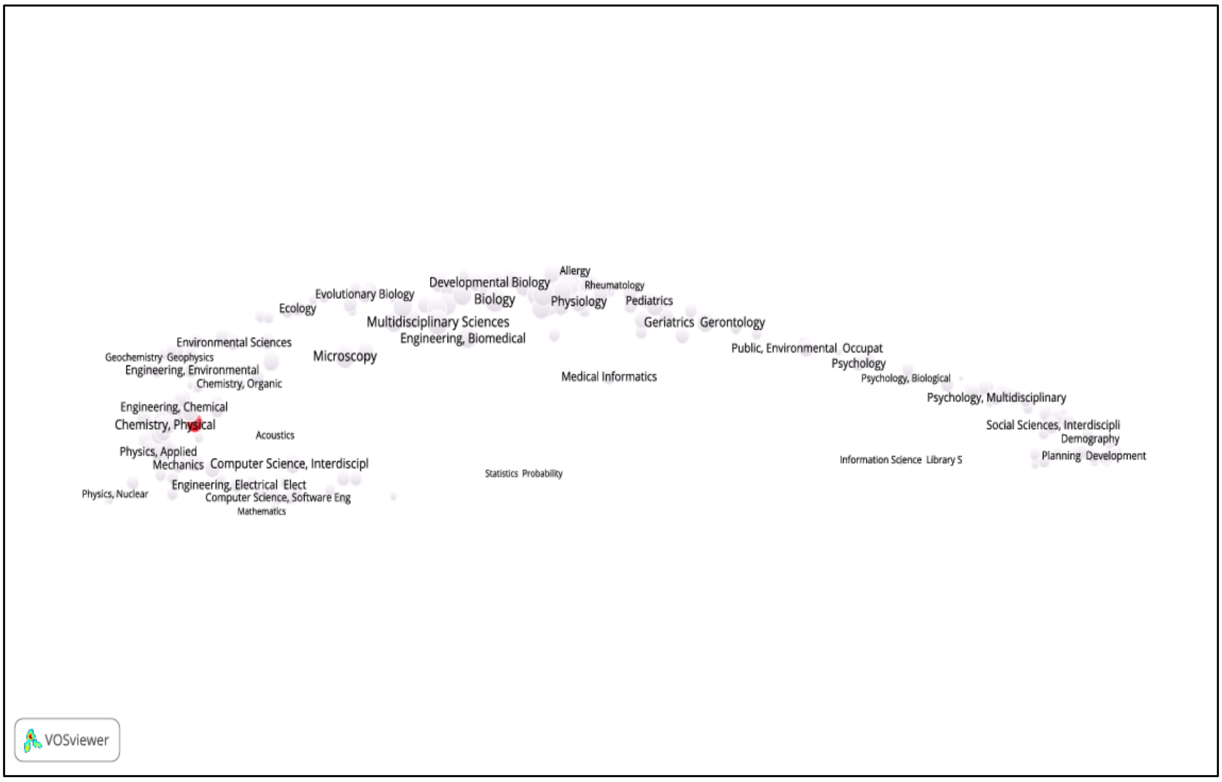


Figure 75. SAPV of the PM4's publications in WoS SCs similarity matrix

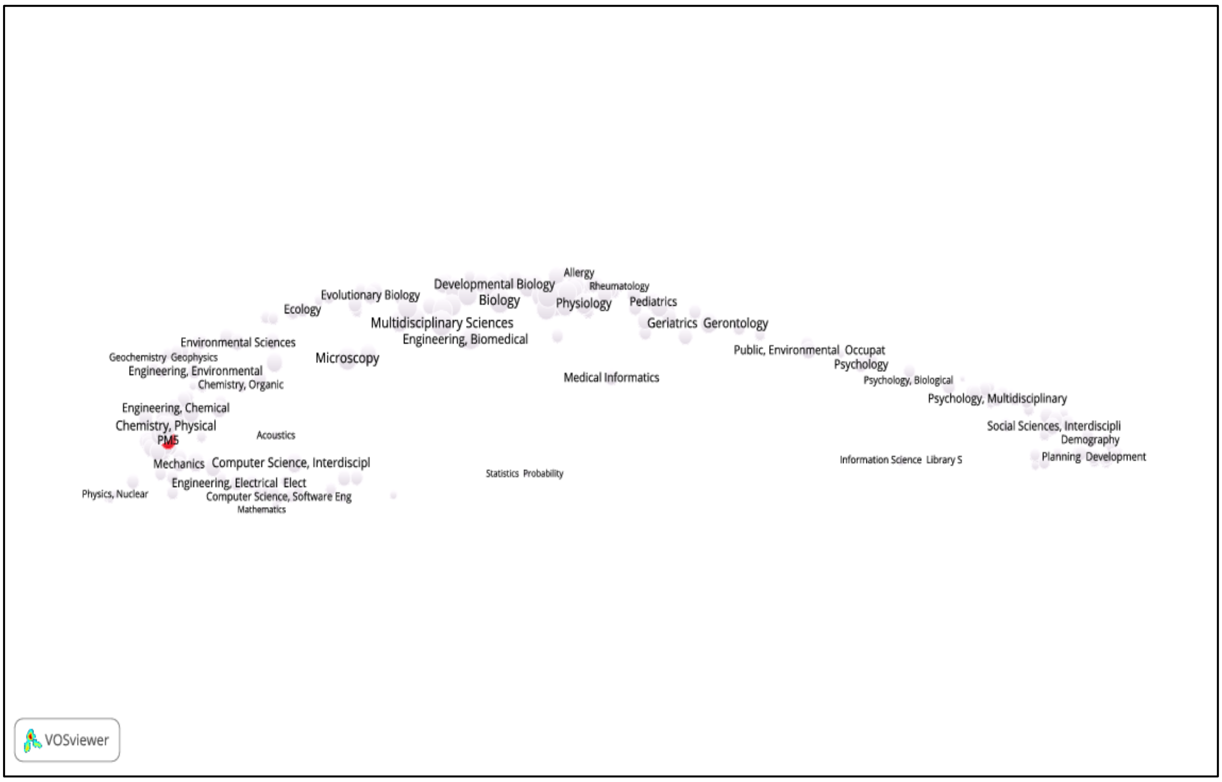


Figure 76. SAPV of the PM5's publications in WoS SCs similarity matrix

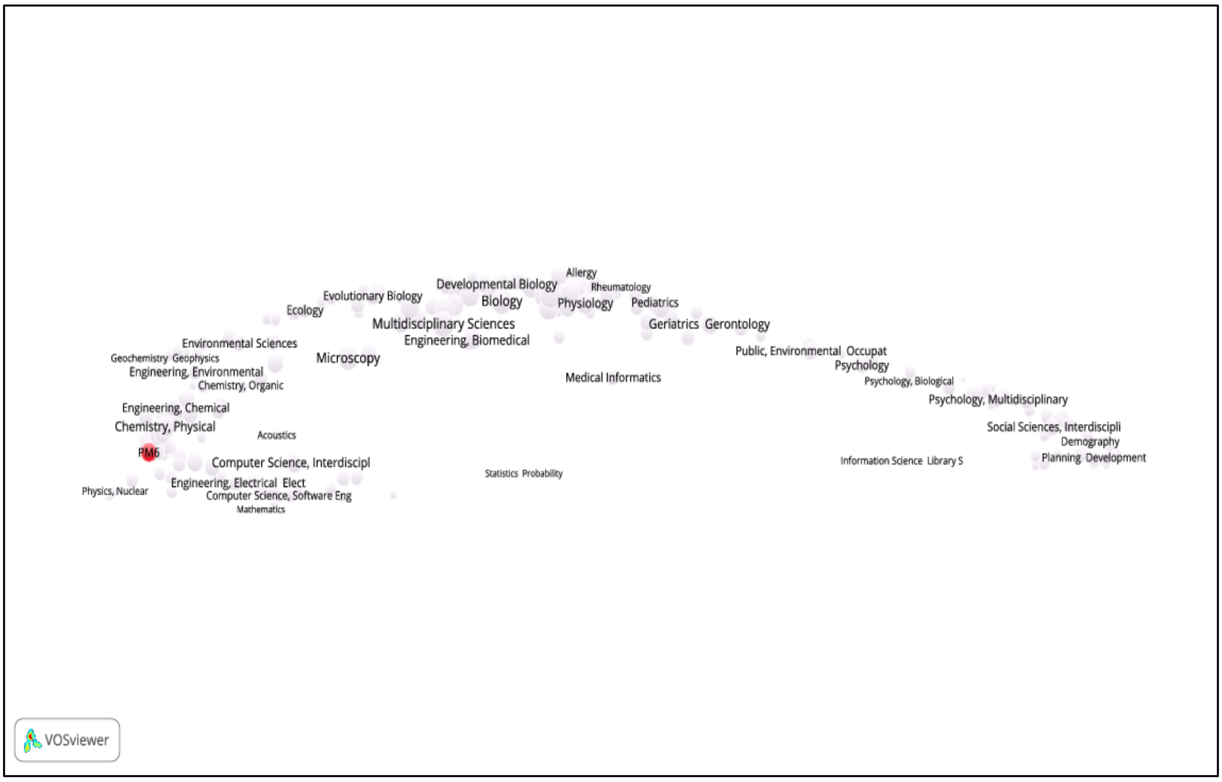


Figure 77. SAPV of the PM6's publications in WoS SCs similarity matrix

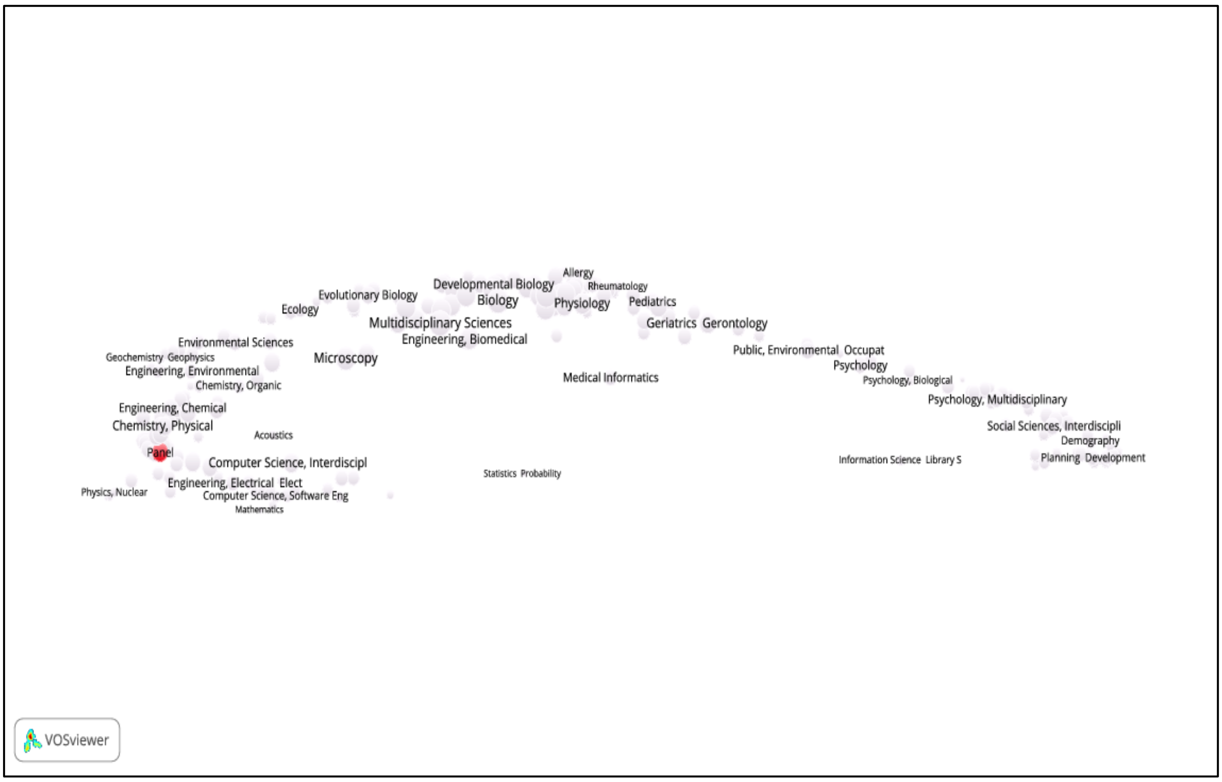


Figure 78. SAPV of the panel publications in WoS SCs similarity matrix

Appendix C

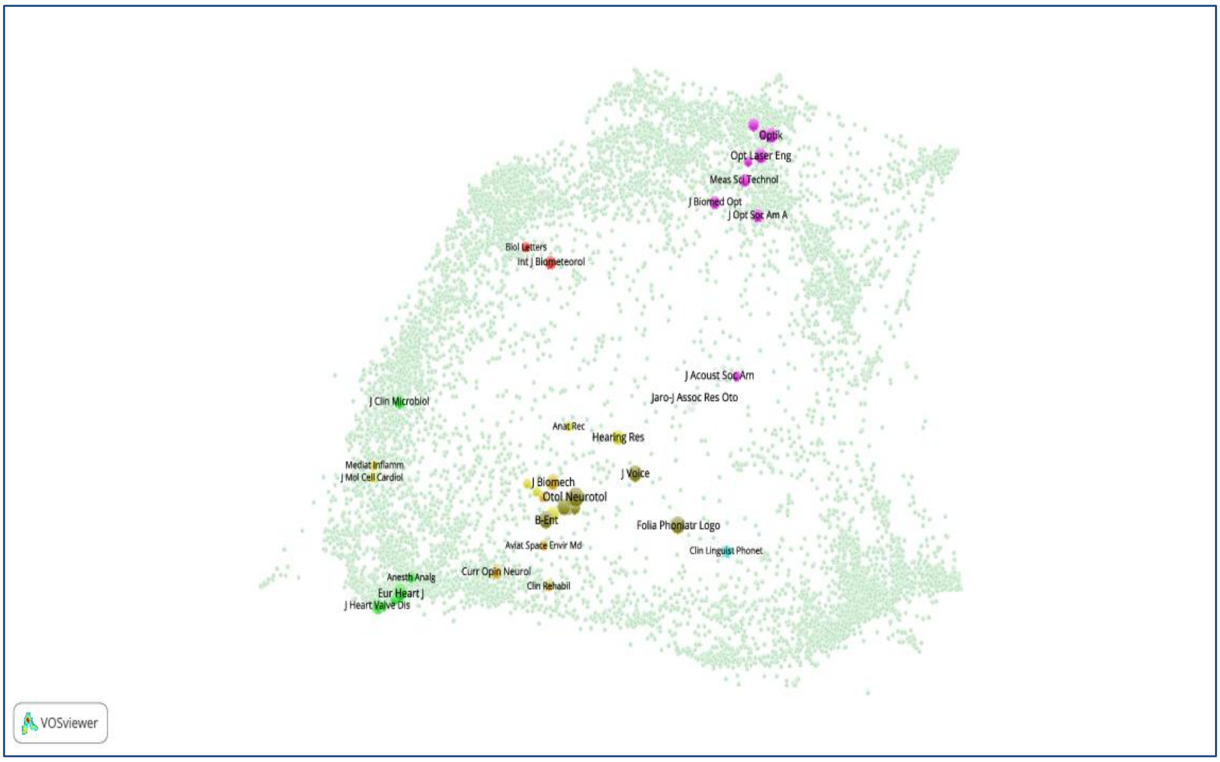


Figure 79. Journal overlay map of PHYS-A research group's publications

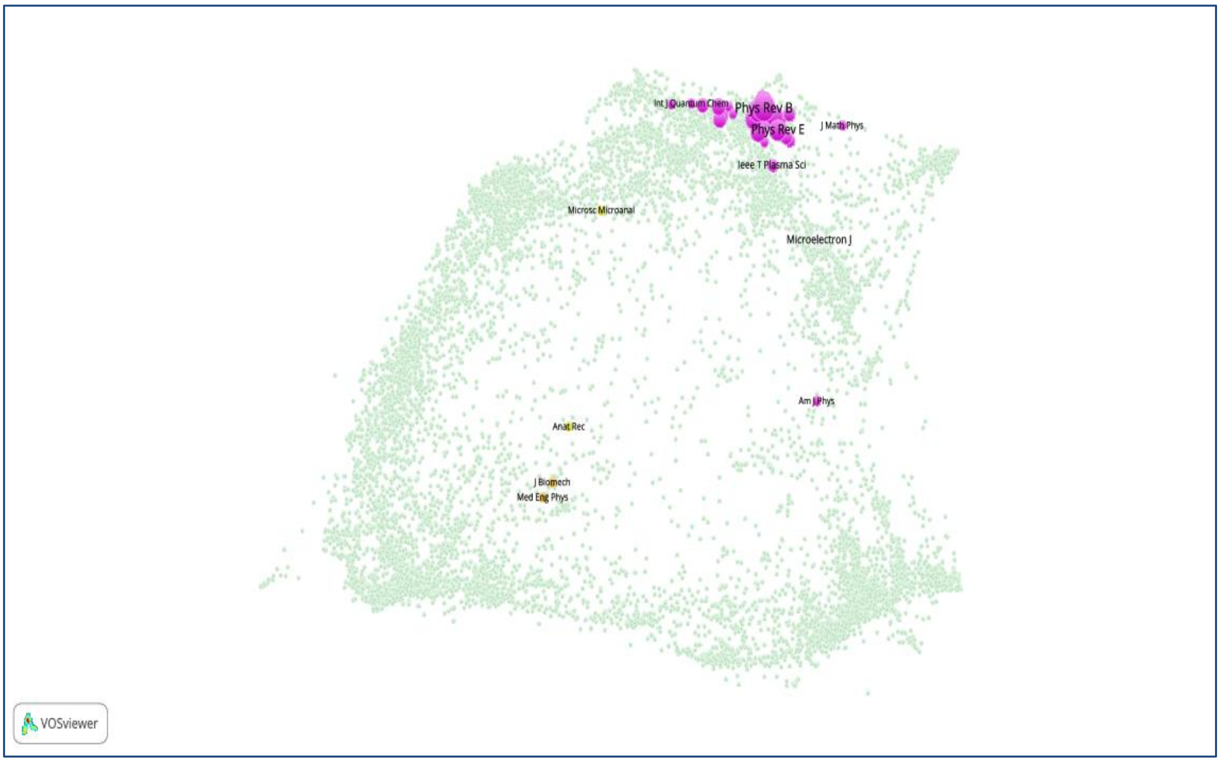


Figure 80. Journal overlay map of PHYS-B research group's publications

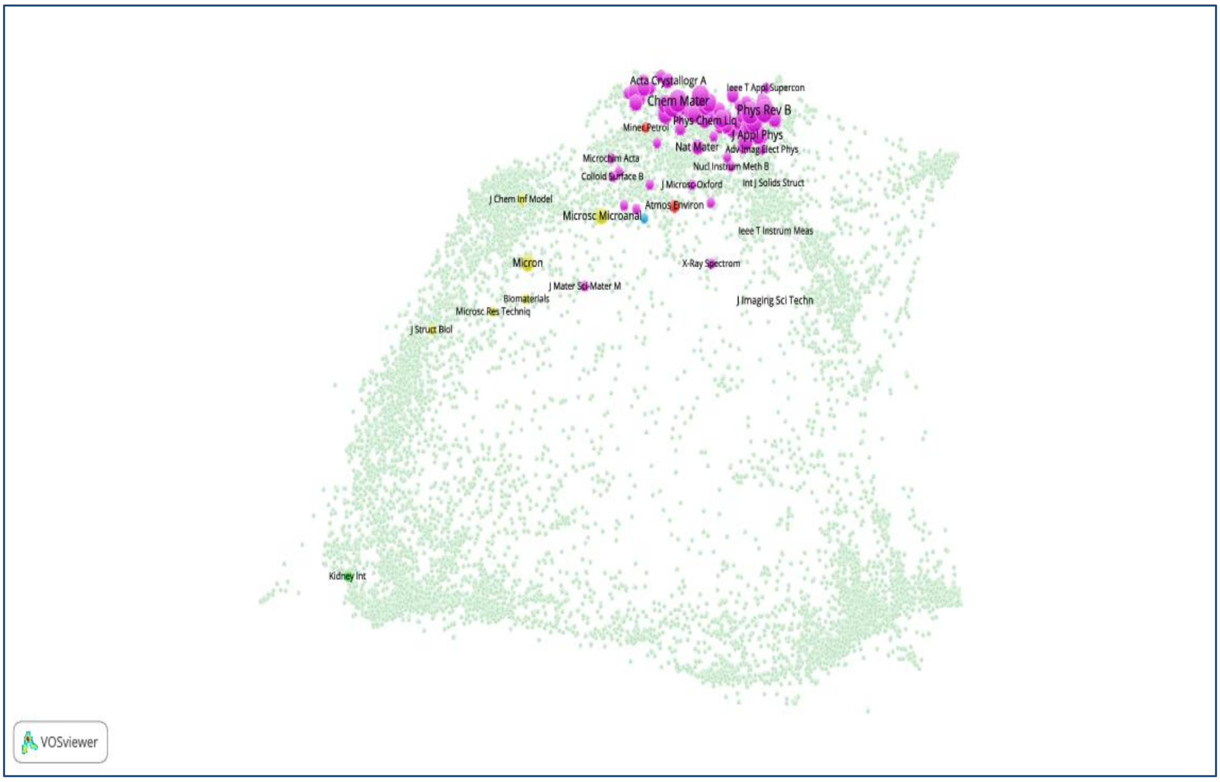


Figure 81. Journal overlay map of PHYS-C research group's publications

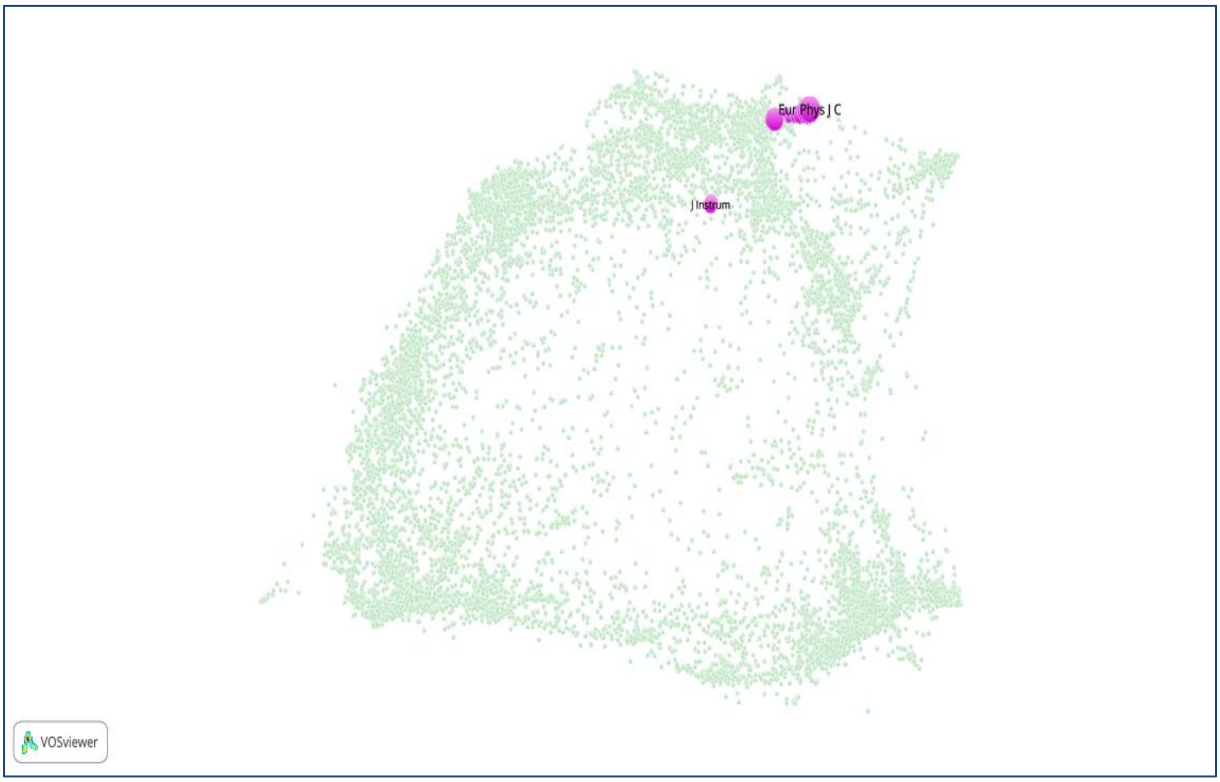


Figure 82. Journal overlay map of PHYS-D research group's publications

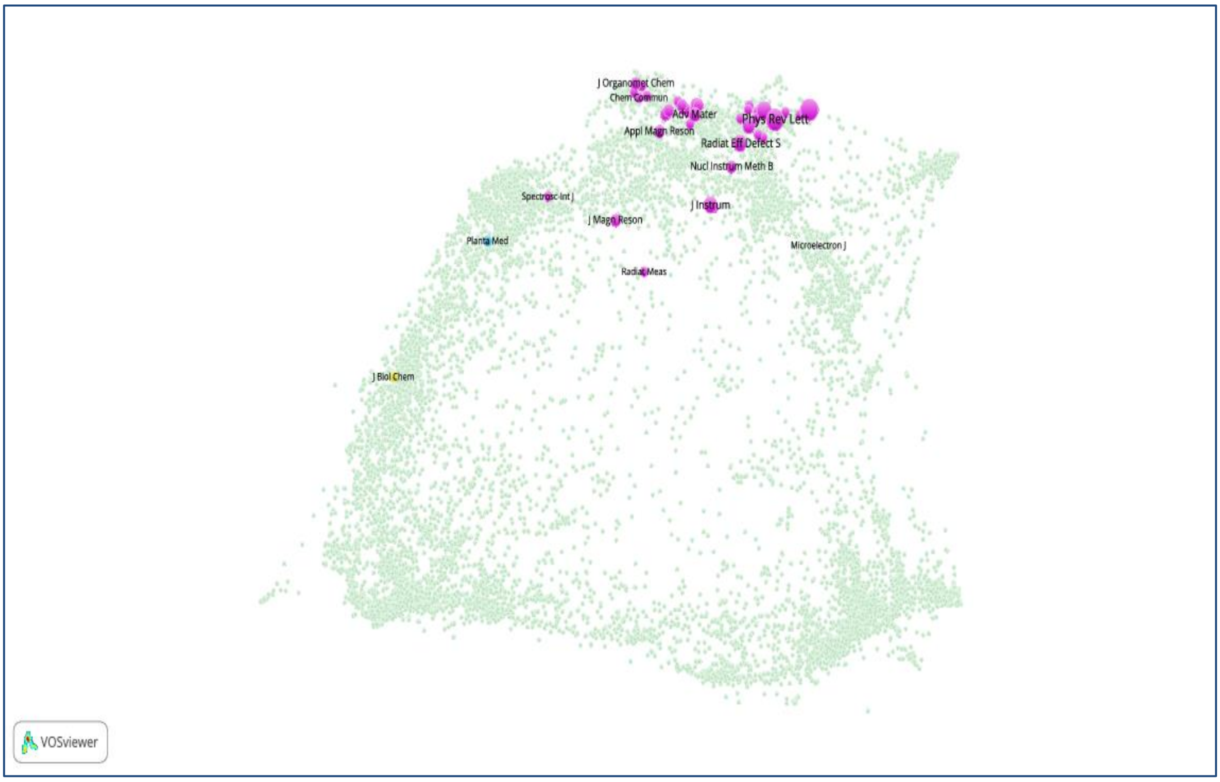


Figure 83. Journal overlay map of PHYS-E research group's publications

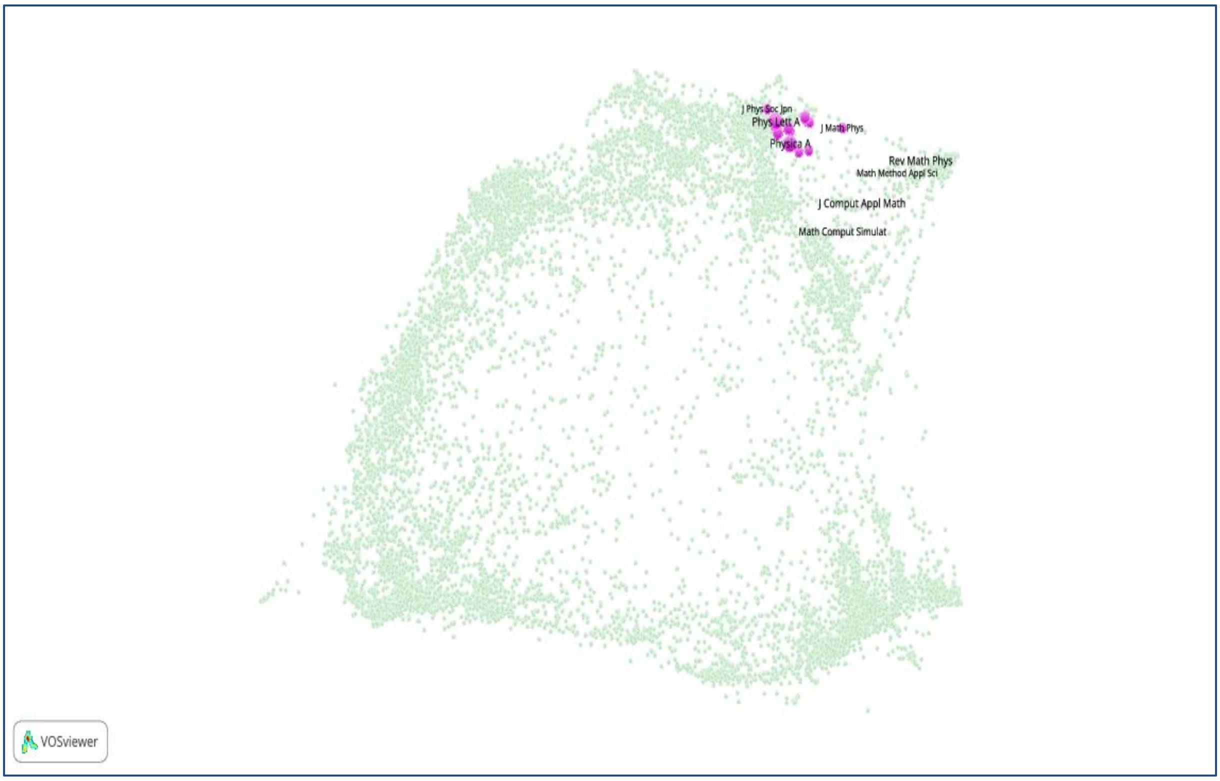


Figure 84. Journal overlay map of PHYS-F research group's publications

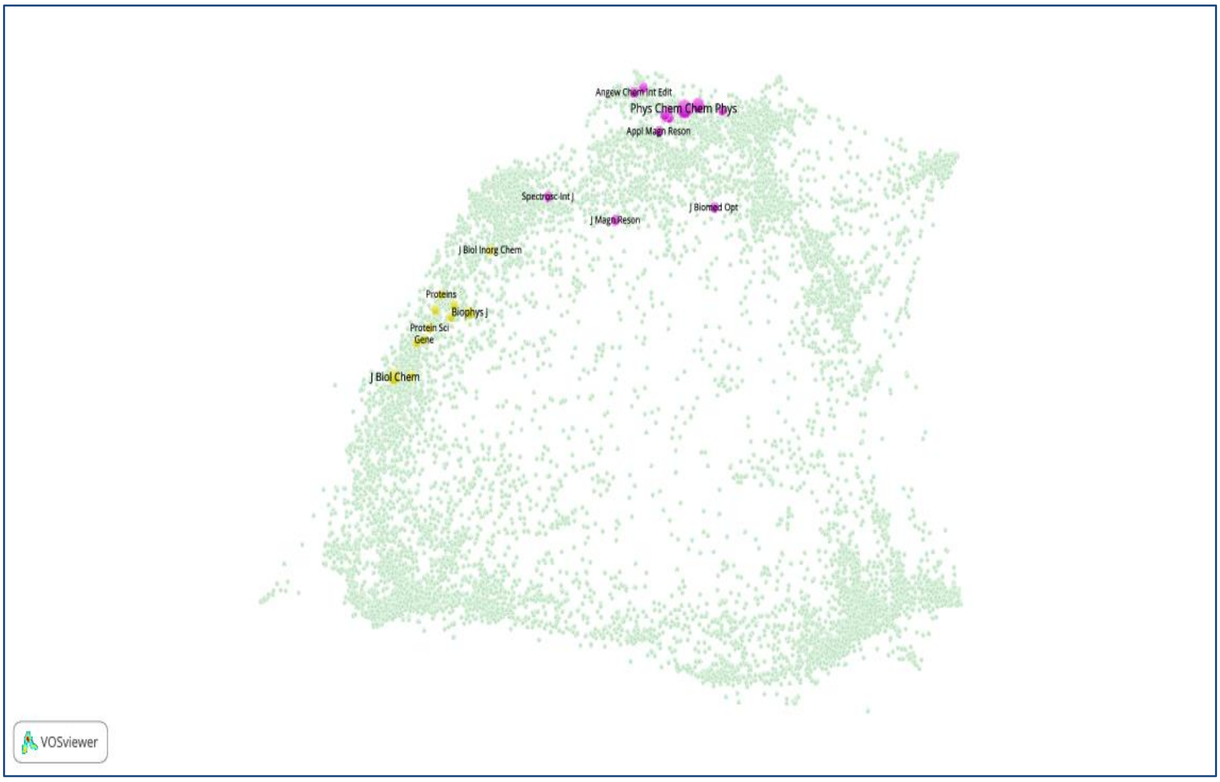


Figure 85. Journal overlay map of PHYS-G research group's publications

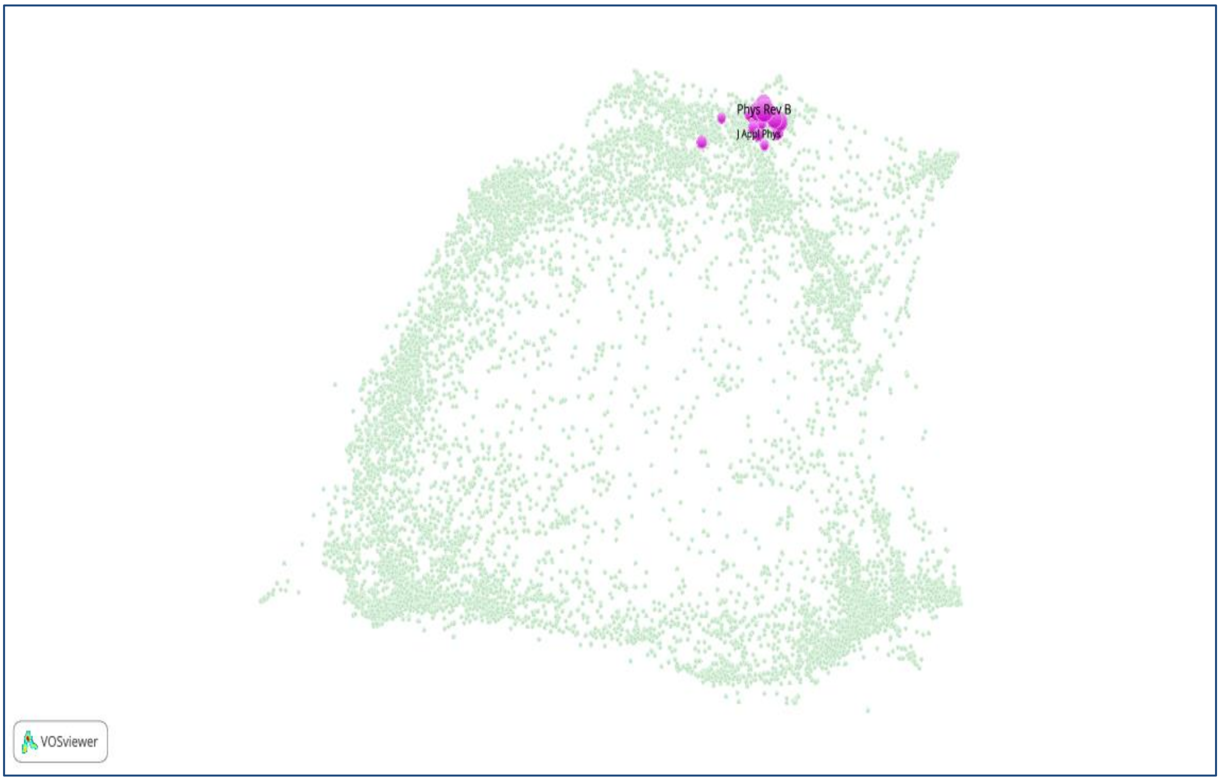


Figure 86. Journal overlay map of PHYS-H research group's publications

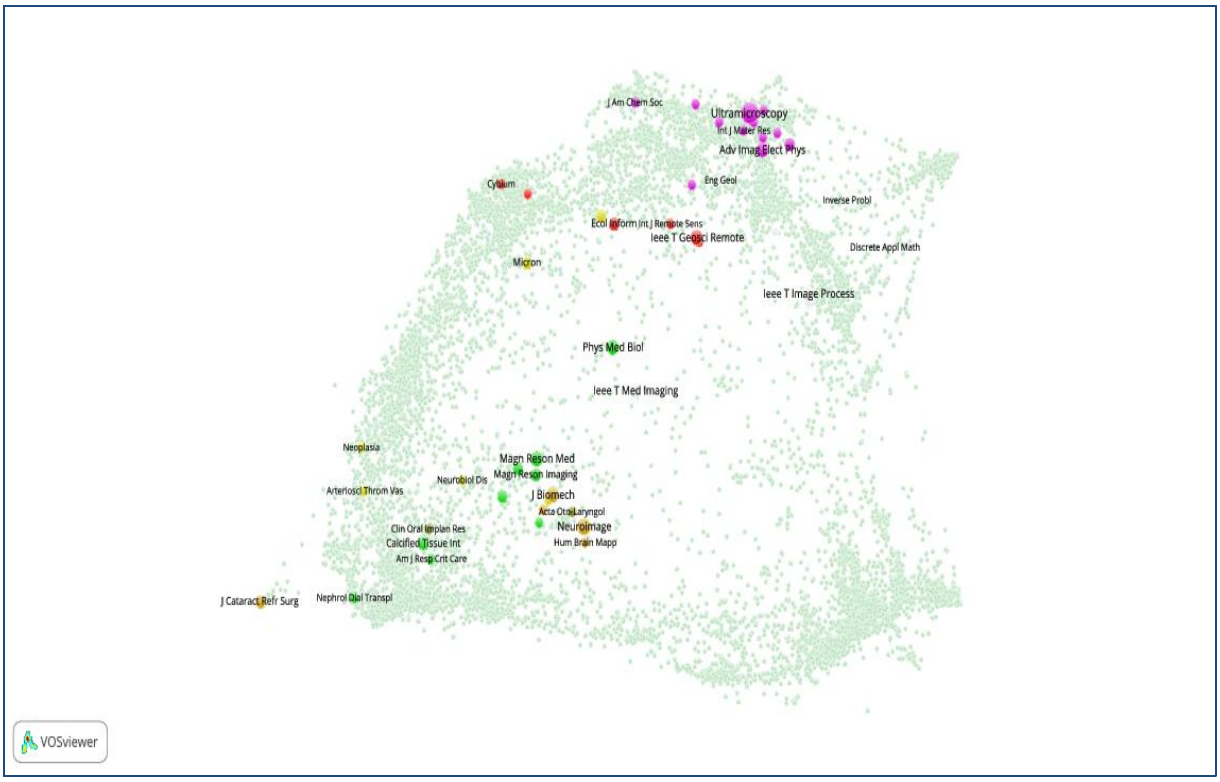


Figure 87. Journal overlay map of PHYS-I research group's publications

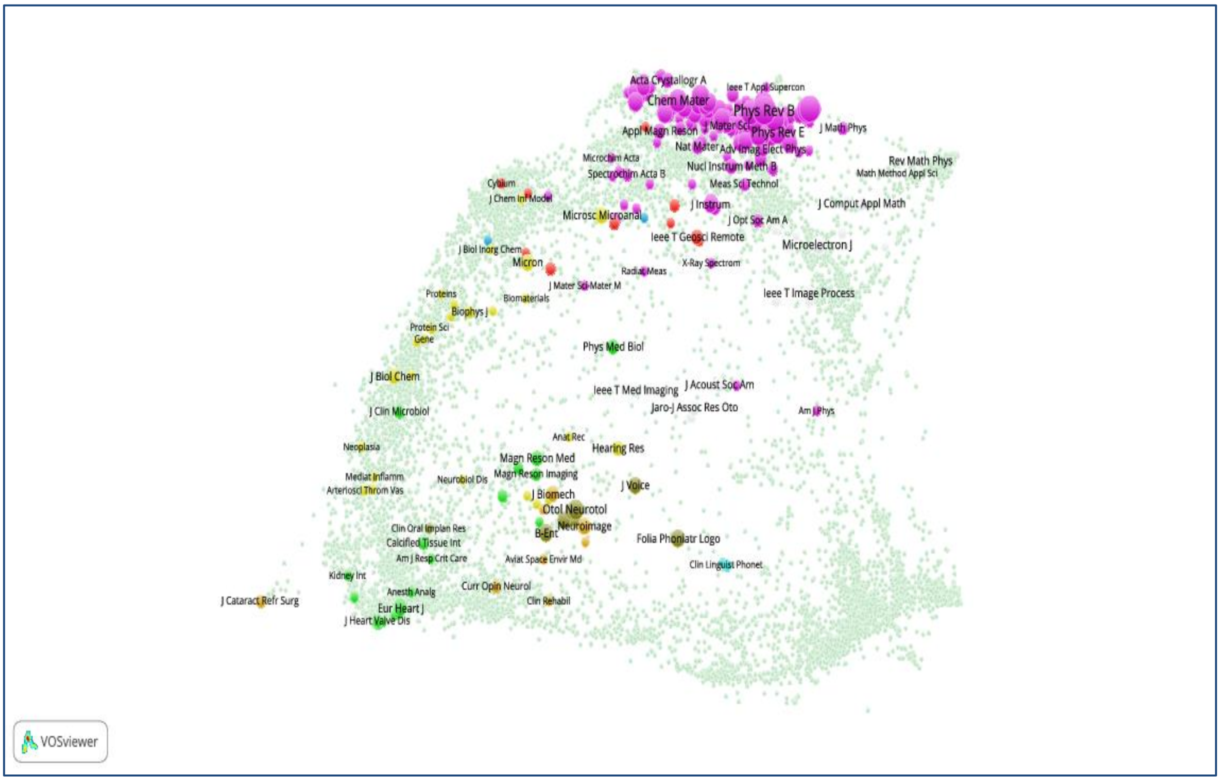


Figure 88. Figure 84. Journal overlay map of Physics research groups' publications

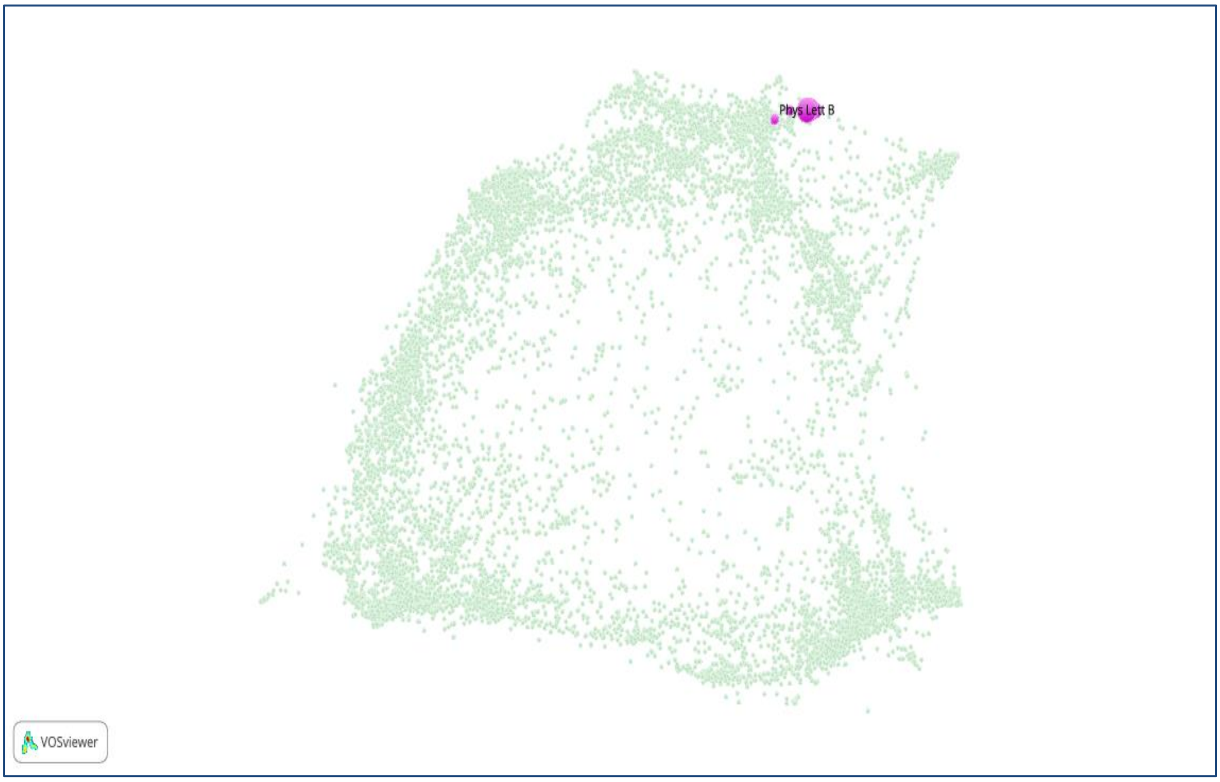


Figure 89. Journal overlay map of PM1's publications

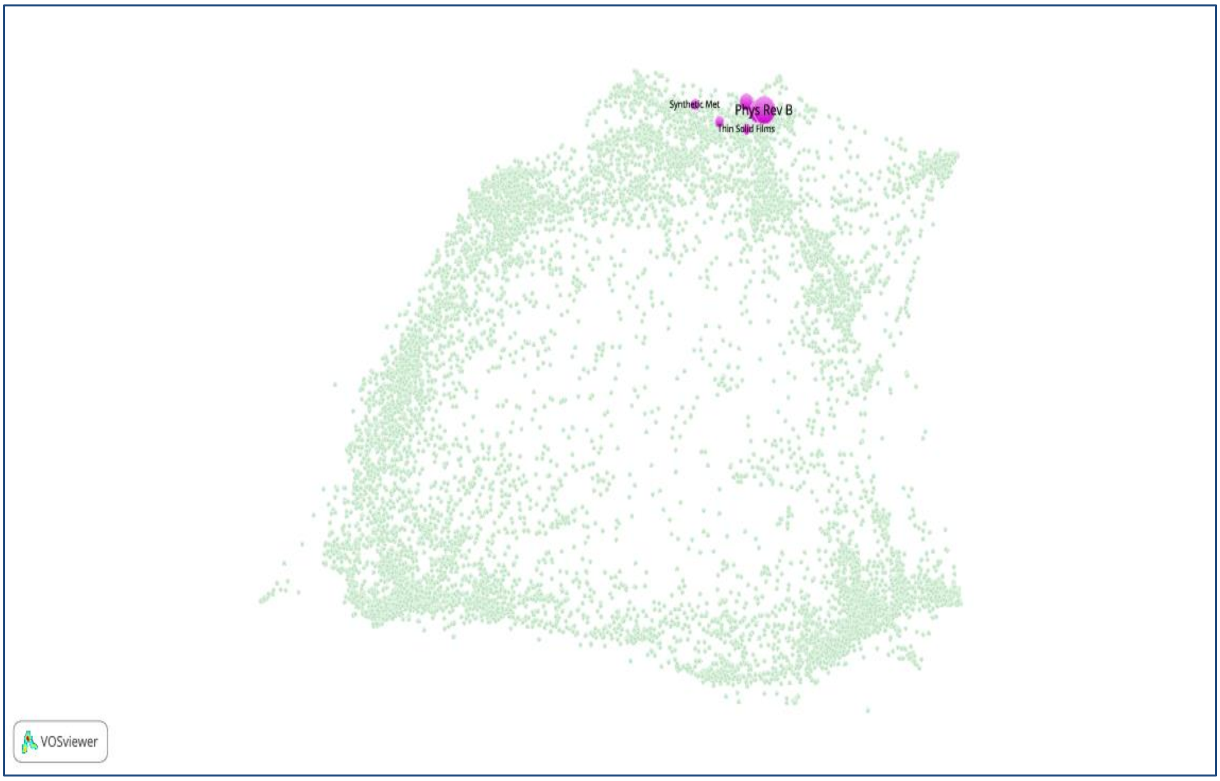


Figure 90. Journal overlay map of PM2's publications

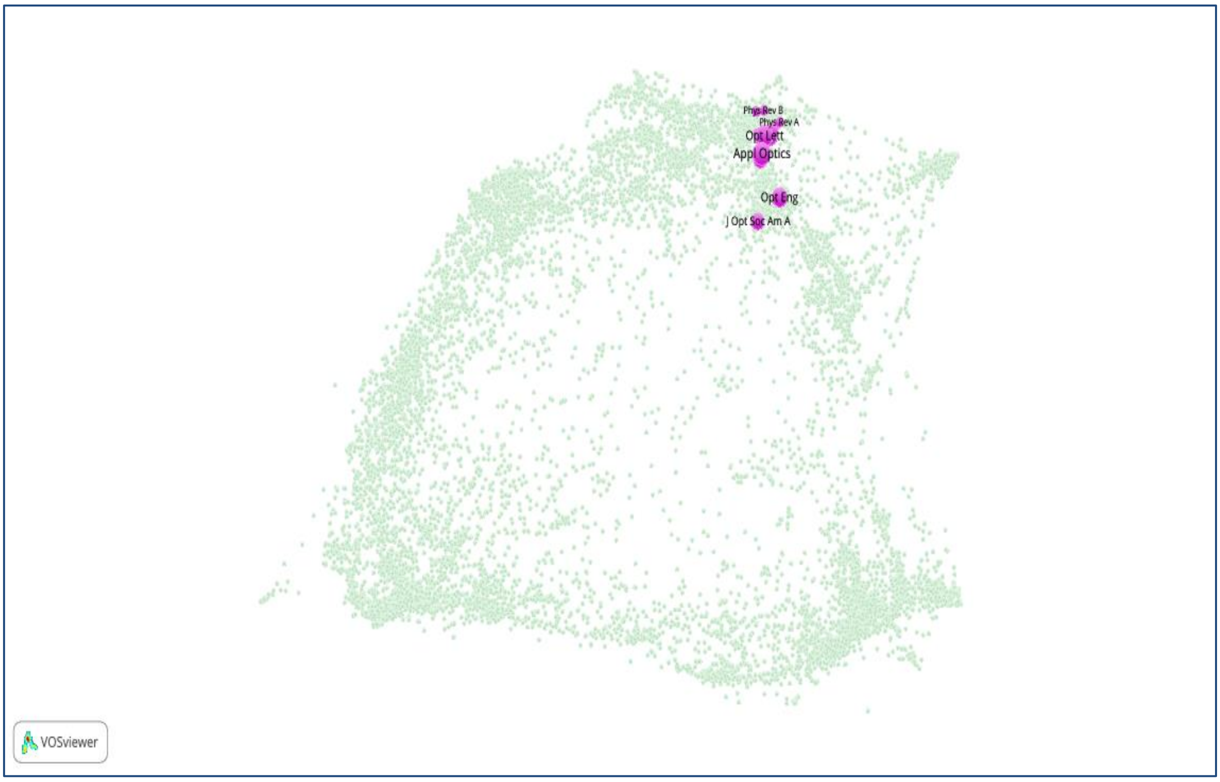


Figure 91. Journal overlay map of PM3's publications

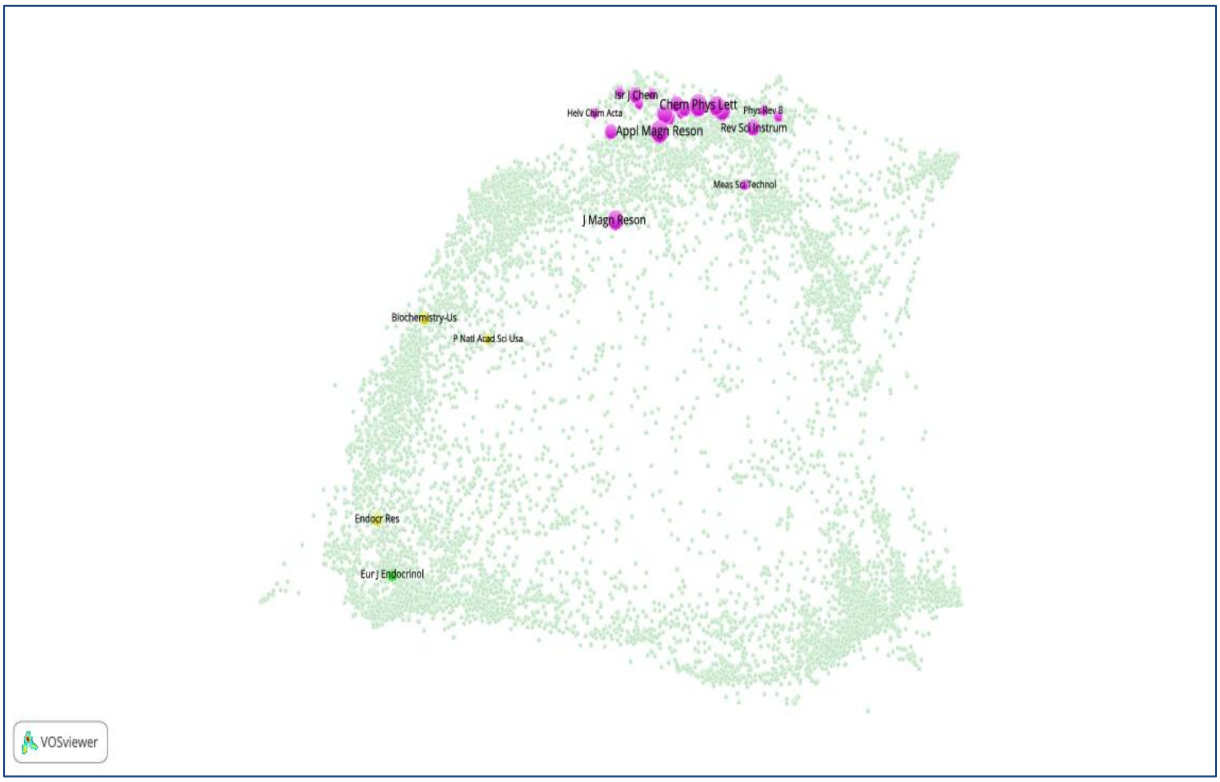


Figure 92. Journal overlay map of PM4's publications

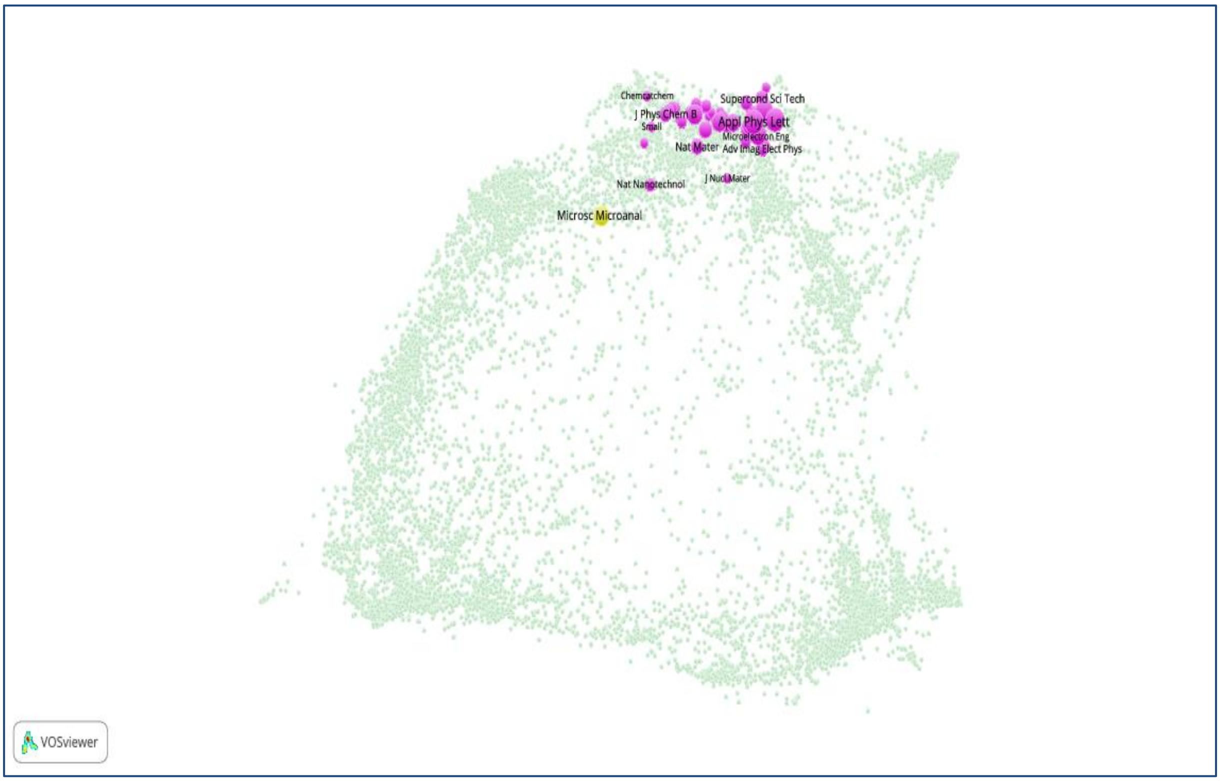


Figure 93. Journal overlay map of PM5's publications

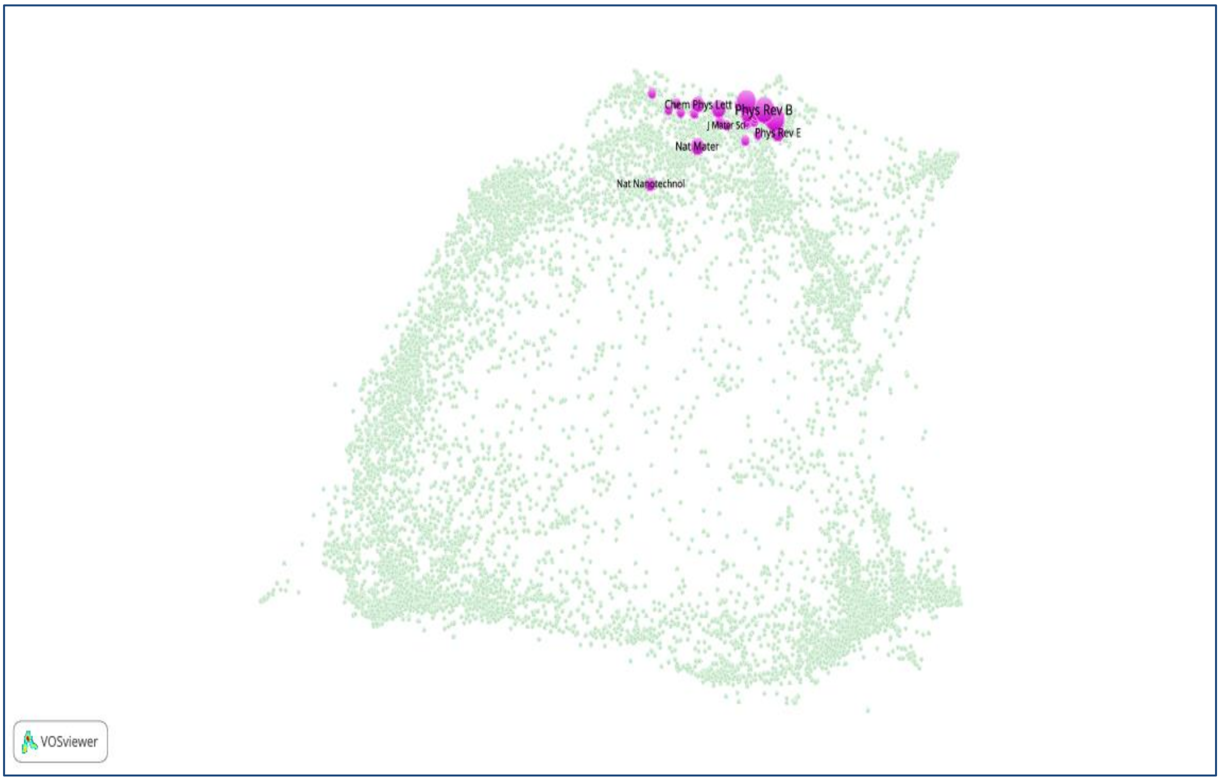


Figure 94. Journal overlay map of PM6's publications

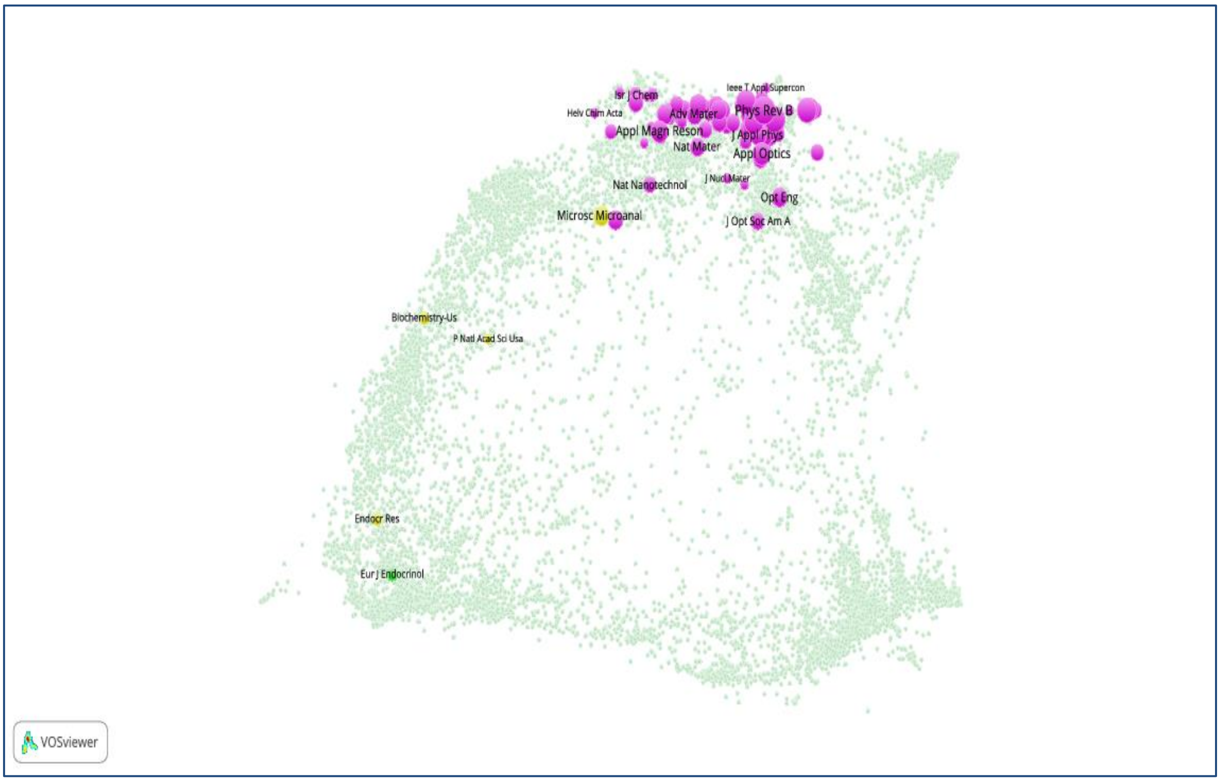


Figure 95. Journal overlay map of the panel's publications