Determining cognitive distance between publication portfolios of evaluators and evaluees in research evaluation: A case study of Pharmaceutical Sciences 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, Chemistry, Physics, 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., Pharmaceutical Sciences) 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 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 Pharmaceutical Sciences and the respective expert panels based on research evaluations carried out at the University of Antwerp.

In this technical report, we present the Pharmaceutical Sciences 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 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 Pharmaceutical Sciences 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 2009 assessment of the ten research groups of the Department of Pharmaceutical Sciences, 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 PHAR-A, PHAR-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 'PHAR-A_WoS SCs.txt' and keep both files.

Table 1: Publication statistics of Pharmaceutical Sciences research groups (2001-2008)

Group code	Number of Publications	Number of Journals	Number of WoS SCs
PHAR-A	40	22	19
PHAR-B	62	32	21
PHAR-C	61	35	25
PHAR-D	32	17	13
PHAR-E	64	42	31
PHAR-F	34	21	8
PHAR-G	67	31	14
PHAR-H	39	27	21
PHAR-I	29	10	6
PHAR-J	11	9	10
All groups	**376**	**180**	**67**

Table 1 lists the publication profile of the Pharmaceutical Sciences research groups during the eight years preceding their evaluation. The Pharmaceutical Sciences research groups generated 376 publications in 180 journals. Members of two research groups co-authored 59 publications and three research groups co-authored four publications. In total, their publications are distributed over 67 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

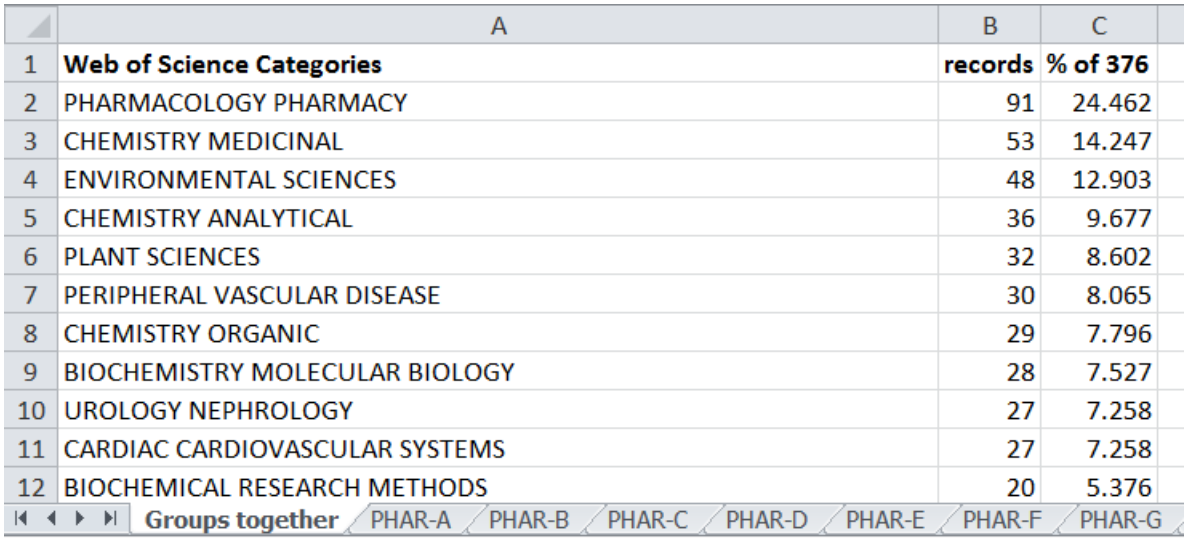
The Pharmaceutical Sciences panel was composed of five panel members (including the chair). We have obtained the names and curricula vitae of the panel members from the ADOC. We replace the original name of each panel member with a code name: PM1, PM2 etc. We perform an advanced search for each panel member in WoS through checking the SCIE and SSCI. All the publications of the individual panel members up to the year of assessment (2009) 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 Pharmaceutical Sciences panel members

Panel code	Number of Publications	Number of Journals	Number of WoS SCs
PM1	122	39	17
PM2	351	93	36
PM3	259	91	33
PM4	124	67	31
PM5	180	86	33
Panel	**1036**	**300**	**68**

Table 2 lists the publication profile of the Pharmaceutical Sciences panel members. The combined publication output of the Pharmaceutical Sciences panel members consists of 1036 publications, none of which is co-authored publications between panel members. The number of publications per panel member ranges from 122 to 351. In total, these publications appeared in 300 different journals and are assigned to 68 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 376**
2	PHARMACOLOGY PHARMACY	91	24.462
3	CHEMISTRY MEDICINAL	53	14.247
4	ENVIRONMENTAL SCIENCES	48	12.903
5	CHEMISTRY ANALYTICAL	36	9.677
6	PLANT SCIENCES	32	8.602
7	PERIPHERAL VASCULAR DISEASE	30	8.065
8	CHEMISTRY ORGANIC	29	7.796
9	BIOCHEMISTRY MOLECULAR BIOLOGY	28	7.527
10	UROLOGY NEPHROLOGY	27	7.258
11	CARDIAC CARDIOVASCULAR SYSTEMS	27	7.258
12	BIOCHEMICAL RESEARCH METHODS	20	5.376

Groups together | PHAR-A | PHAR-B | PHAR-C | PHAR-D | PHAR-E | PHAR-F | PHAR-G

Figure 1. Excerpt of Pharmaceutical Sciences 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 (PHAR-A, PHAR-B, PHAR-C, etc.), the panel members' code names, (PM1, PM2, PM3, etc.), Panel together and Groups together. The Excel file is saved as 'Pharmaceutical Sciences 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 'Pharmaceutical Sciences 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	B	C
1	**Web of Science Categories**	**records**	**% of 376**
2	PHARMACOLOGY PHARMACY	91	24.462
3	CHEMISTRY MEDICINAL	53	14.247
4	ENVIRONMENTAL SCIENCES	48	12.903
5	CHEMISTRY ANALYTICAL	36	9.677
6	PLANT SCIENCES	32	8.602
7	PERIPHERAL VASCULAR DISEASE	30	8.065
8	CHEMISTRY ORGANIC	29	7.796
9	BIOCHEMISTRY MOLECULAR BIOLOGY	28	7.527
10	UROLOGY NEPHROLOGY	27	7.258
11	CARDIAC CARDIOVASCULAR SYSTEMS	27	7.258
12	BIOCHEMICAL RESEARCH METHODS	20	5.376
13	SPECTROSCOPY	19	5.108
14	HEMATOLOGY	19	5.108
15	BIOPHYSICS	13	3.495
16	TOXICOLOGY	12	3.226
17	CHEMISTRY MULTIDISCIPLINARY	12	3.226
18	PHYSIOLOGY	10	2.688
19	PSYCHIATRY	9	2.419
20	INTEGRATIVE COMPLEMENTARY MEDICINE	9	2.419

GroupsTogether | PanelTogether

	A	B	C
1	**Web of Science Categories**	**records**	**% of 1032**
2	PHARMACOLOGY PHARMACY	509	49.465
3	CHEMISTRY MULTIDISCIPLINARY	260	25.267
4	BIOCHEMISTRY MOLECULAR BIOLOGY	99	9.621
5	CHEMISTRY ANALYTICAL	98	9.524
6	CHEMISTRY MEDICINAL	95	9.232
7	SPECTROSCOPY	55	5.345
8	BIOCHEMICAL RESEARCH METHODS	42	4.082
9	CHEMISTRY ORGANIC	35	3.401
10	ONCOLOGY	34	3.304
11	MEDICINE RESEARCH EXPERIMENTAL	32	3.11
12	BIOPHYSICS	29	2.818
13	CHEMISTRY PHYSICAL	28	2.721
14	NEUROSCIENCES	25	2.43
15	TOXICOLOGY	21	2.041
16	PHYSICS ATOMIC MOLECULAR CHEMICAL	18	1.749
17	MICROBIOLOGY	18	1.749
18	MATERIALS SCIENCE BIOMATERIALS	18	1.749
19	ENGINEERING BIOMEDICAL	18	1.749
20	CRYSTALLOGRAPHY	13	1.263

GroupsTogether | PanelTogether

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

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 "Pharmaceutical Sciences panel and research groups
                       together_WoS SCs.xlsx"
```

This produces a new Excel file called 'Pharmaceutical Sciences 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 positive value ($r = 0.70$, $\rho = 0.29$). Figure 4 shows a Log-log plot of the number of publications per WoS SCs for the Pharmaceutical Sciences panel and research groups together.

	A	B	C	D
1		**Web of Science Categories**	**records_x**	**records_y**
2	**0**	PHARMACOLOGY PHARMACY	91	509
3	**1**	CHEMISTRY MEDICINAL	53	95
4	**2**	ENVIRONMENTAL SCIENCES	48	0
5	**3**	CHEMISTRY ANALYTICAL	36	98
6	**4**	PLANT SCIENCES	32	4
7	**5**	PERIPHERAL VASCULAR DISEASE	30	7
8	**6**	CHEMISTRY ORGANIC	29	35
9	**7**	BIOCHEMISTRY MOLECULAR BIOLOGY	28	99
10	**8**	UROLOGY NEPHROLOGY	27	0
11	**9**	CARDIAC CARDIOVASCULAR SYSTEMS	27	4
12	**10**	BIOCHEMICAL RESEARCH METHODS	20	42
13	**11**	SPECTROSCOPY	19	55
14	**12**	HEMATOLOGY	19	8
15	**13**	BIOPHYSICS	13	29
16	**14**	TOXICOLOGY	12	21

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

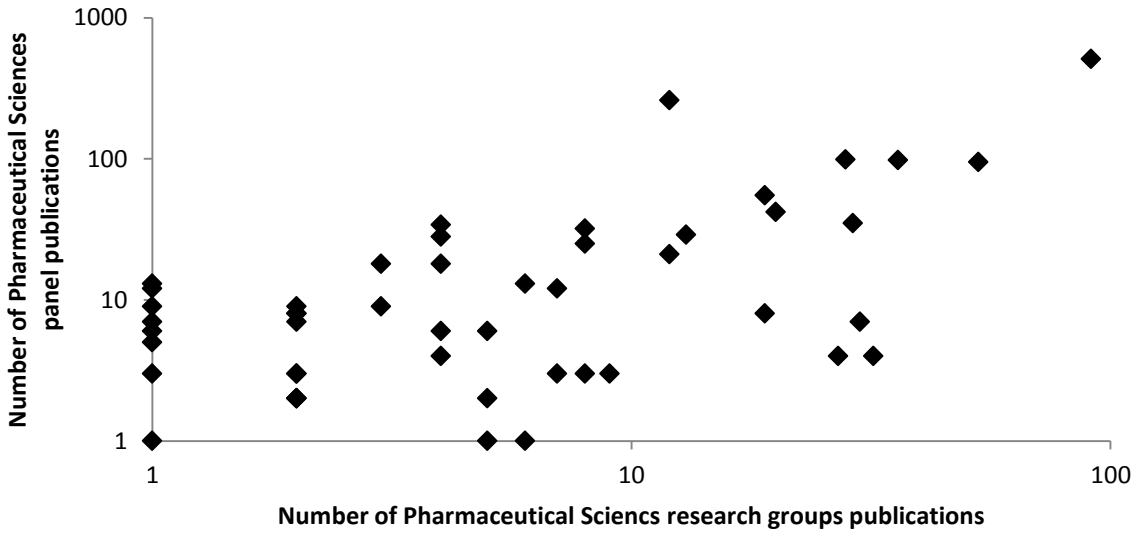


Figure 4. Log-log plot of the number of publications (log-log scale) per WoS SC for the panel (vertical axis) and research groups together (horizontal axis) of the Pharmaceutical Sciences department

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:

$$S_i = \sum_{j=i}^{n} 1/j \qquad (1)$$

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:

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

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 'Pharmaceutical Sciences 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 "Pharmaceutical Sciences panel and research groups
together_WoS SCs-joined.xlsx"
```

The outcome shows that the top-down correlation between Pharmaceutical Sciences research groups together and the panel's profile in the WoS SCs is moderate (0.53).

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 SCI 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.

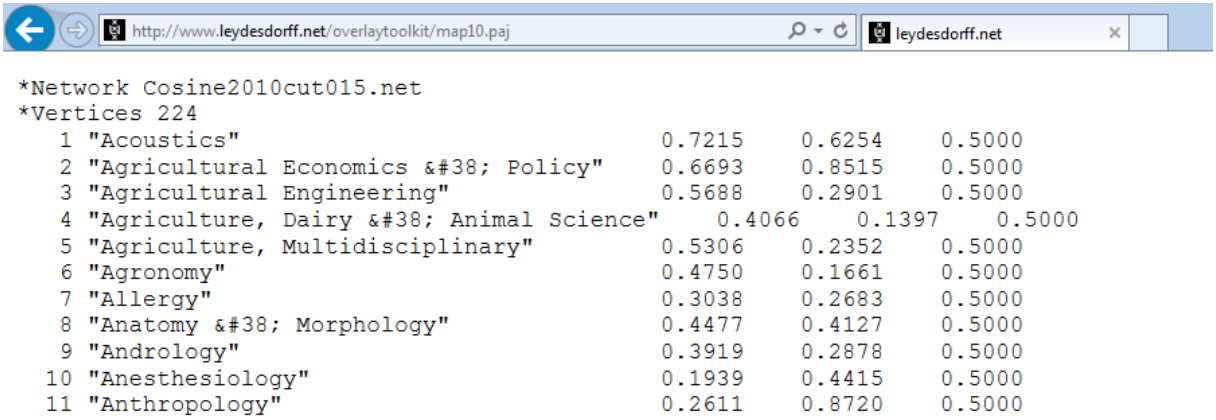


Figure 5. Excerpt of the map10.paj file

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)

ii) 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 SCIE 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).

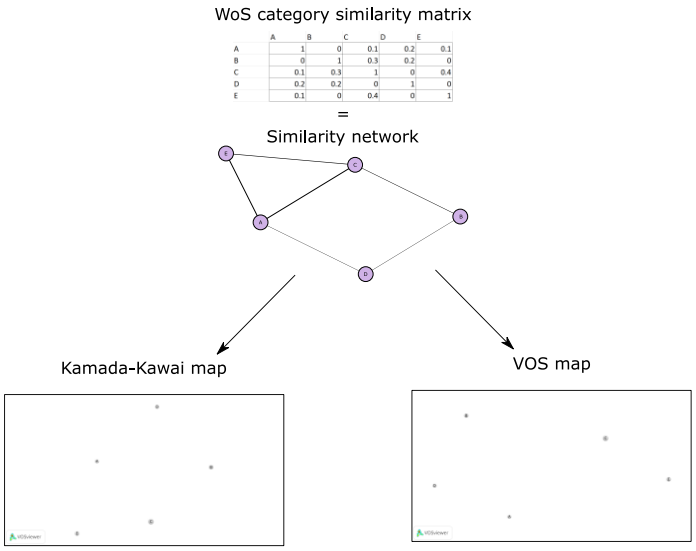


Figure 6. Transformation of WoS SCs similarity matrix to Kamada-Kawai map and VOS-map

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).

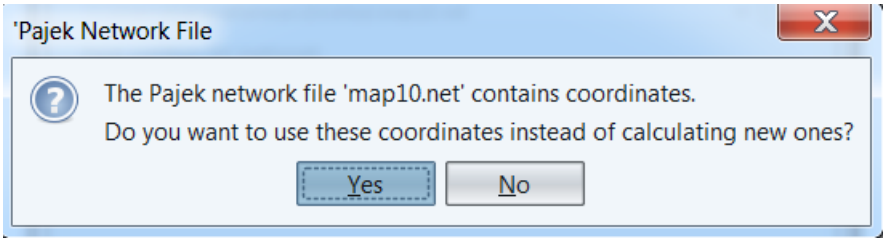


Figure 7. VOSviewer message before choosing Kamada Kawai map or VOS map data

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). 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 from map10.paj file

	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 from map10.paj file

	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 PHAR-B research group's publications overlay map in WoS SCs.

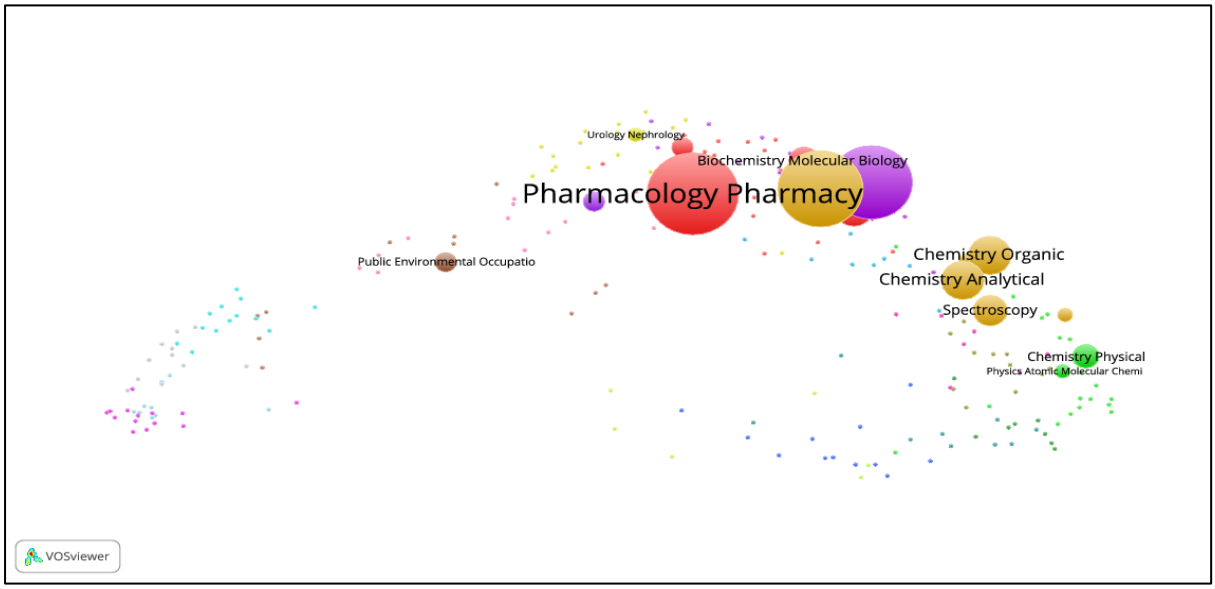


Figure 11. PHAR-B research group's publication overlay map in WoS SCs

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).

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 on 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 $C = (C_1, C_2)$, where

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

Here, Lj,1 and Lj,2 are the horizontal and vertical coordinates of WoS SC j on the map, m_j is the number of publications in WoS SC j, and $T = \sum_{j=1}^{N} m_j$ 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 ('Pharmaceutical Sciences 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" "Pharmaceutical Sciences
research groups and panel_WoS SCs.xlsx"
```

This program calculates the barycenter and generates an output file 'Pharmaceutical Sciences research groups and panel_WoS SCs-barycenter.xlsx'. Figure 12 shows the barycenter

coordinates of the Pharmaceutical Sciences individual research groups, panel members, research groups together and panel.

	A	B	C
1		x	y
2	Groups together	0.264984	0.272405
3	PHAR-A	0.593963	0.136283
4	PHAR-B	0.360218	0.271306
5	PHAR-C	-0.01657	0.420506
6	PHAR-D	0.288911	0.236003
7	PHAR-E	0.091304	0.335971
8	PHAR-F	0.449112	0.264706
9	PHAR-G	0.464561	0.079935
10	PHAR-H	0.296136	0.332864
11	PHAR-I	-0.08628	0.477194
12	PHAR-J	0.02639	0.40271
13	PM1	0.593641	0.143534
14	PM2	0.165965	0.277448
15	PM3	0.438648	0.18326
16	PM4	0.414994	0.219059
17	PM5	0.604778	0.148765
18	Panel together	0.408232	0.202479

Figure 12. Barycenter coordinates of the Pharmaceutical Sciences 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:

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

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
1		Groups together	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H
2	Groups together	0	0.356028753	0.095240164	0.318128864	0.043561582	0.184947514	0.184288907	0.277265209	0.068012064
3	PHAR-A	0.356028753	0	0.269940907	0.673448012	0.320937835	0.540871597	0.193582671	0.141138205	0.356853771
4	PHAR-B	0.095240164	0.269940907	0	0.405251773	0.07956795	0.276580095	0.089138878	0.217969335	0.088857889
5	PHAR-C	0.318128864	0.673448012	0.405251773	0	0.356874166	0.137049638	0.491052593	0.589470196	0.324755207
6	PHAR-D	0.043561582	0.320937835	0.07956795	0.356874166	0	0.221454793	0.162752504	0.234969052	0.097129594
7	PHAR-E	0.184947514	0.540871597	0.276580095	0.137049638	0.221454793	0	0.364836651	0.452632339	0.204856533
8	PHAR-F	0.184288907	0.193582671	0.089138878	0.491052593	0.162752504	0.364836651	0	0.185416011	0.167472444
9	PHAR-G	0.277265209	0.141138205	0.217969335	0.589470196	0.234969052	0.452632339	0.185416011	0	0.303874794
10	PHAR-H	0.068012064	0.356853771	0.088857889	0.324755207	0.097129594	0.204856533	0.167472444	0.303874794	0
11	PHAR-I	0.40659868	0.760885243	0.491677843	0.089848186	0.446025317	0.226888967	0.576014084	0.679144573	0.408743167
12	PHAR-J	0.271856831	0.626994309	0.358758451	0.046499794	0.310979008	0.093100619	0.444678122	0.544221738	0.278642
13	PM1	0.353019683	0.007258165	0.26610503	0.670126315	0.318450716	0.537935538	0.188603038	0.143897064	0.35263915
14	PM2	0.099147529	0.450677152	0.194350111	0.231914513	0.129743217	0.094864875	0.283433775	0.358009906	0.141476316
15	PM3	0.195208157	0.162263649	0.117913359	0.513330984	0.158755276	0.379432825	0.082115871	0.106524646	0.20661795
16	PM4	0.15921306	0.197184678	0.075698184	0.476264194	0.127216747	0.344156972	0.056988758	0.147690314	0.164555819
17	PM5	0.36158944	0.016515095	0.273543597	0.67817075	0.327693045	0.546536739	0.194098602	0.156199515	0.35937729
18	Panel together	0.159403999	0.197175036	0.083920044	0.47748471	0.123941182	0.343895047	0.074454252	0.134870454	0.171946399

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.159	0.197	0.084	0.477	0.124	0.344	0.074	0.135	0.172	0.566	0.431
PM1	0.353	**0.007**	0.266	0.670	**0.318**	0.538	**0.189**	**0.144**	0.353	0.757	0.624
PM2	0.099	0.451	**0.194**	**0.232**	**0.130**	**0.095**	0.283	0.358	**0.141**	**0.322**	**0.188**
PM3	0.195	**0.162**	**0.118**	0.513	**0.159**	0.379	**0.082**	**0.107**	**0.207**	0.602	0.467
PM4	0.159	0.197	**0.076**	0.476	**0.127**	0.344	**0.057**	**0.148**	**0.165**	0.564	0.430
PM5	0.362	**0.017**	0.274	0.678	**0.328**	0.547	**0.194**	**0.156**	0.359	0.765	0.632

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.135 (SD 0.092). 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 Pharmaceutical Sciences 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.

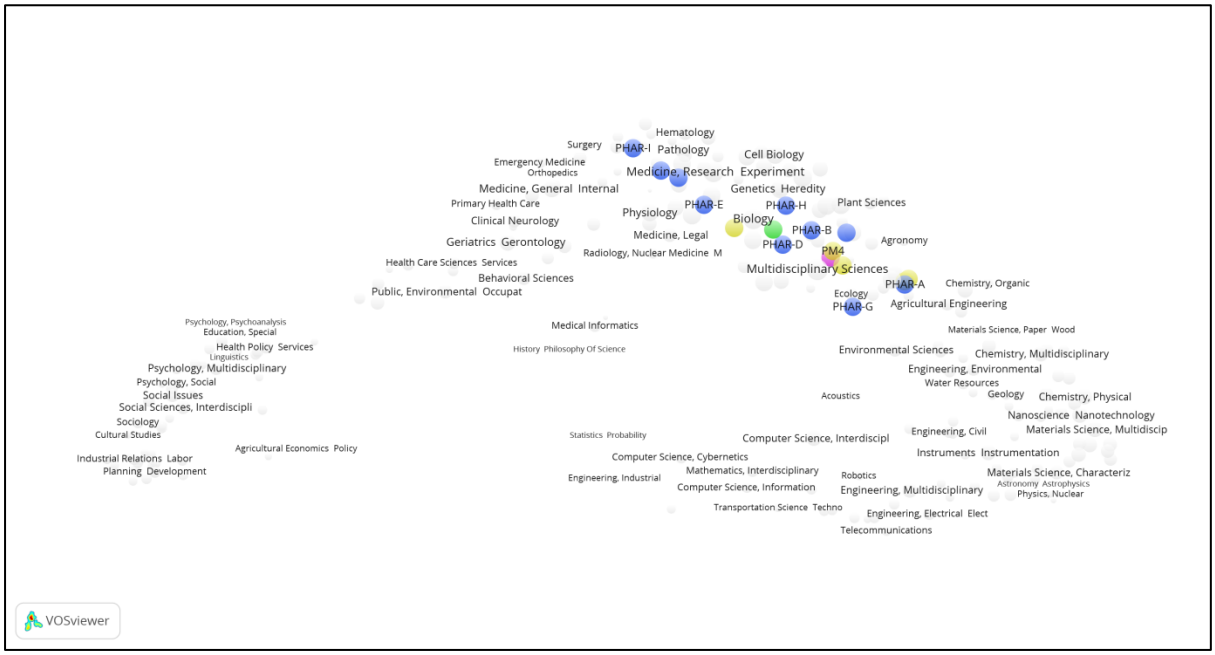


Figure 14. Barycenter overlay map of Pharmaceutical Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs

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 Pharmaceutical Sciences 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 Pharmaceutical Sciences 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

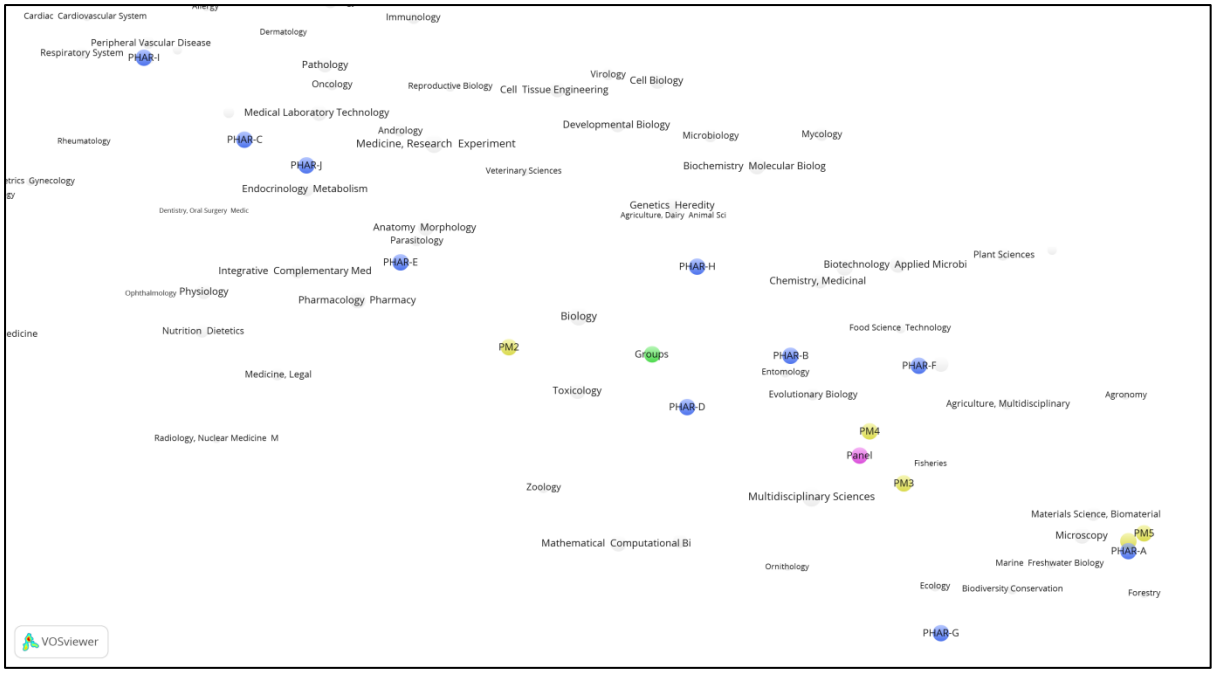


Figure 15. Barycenter map of Pharmaceutical Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed)

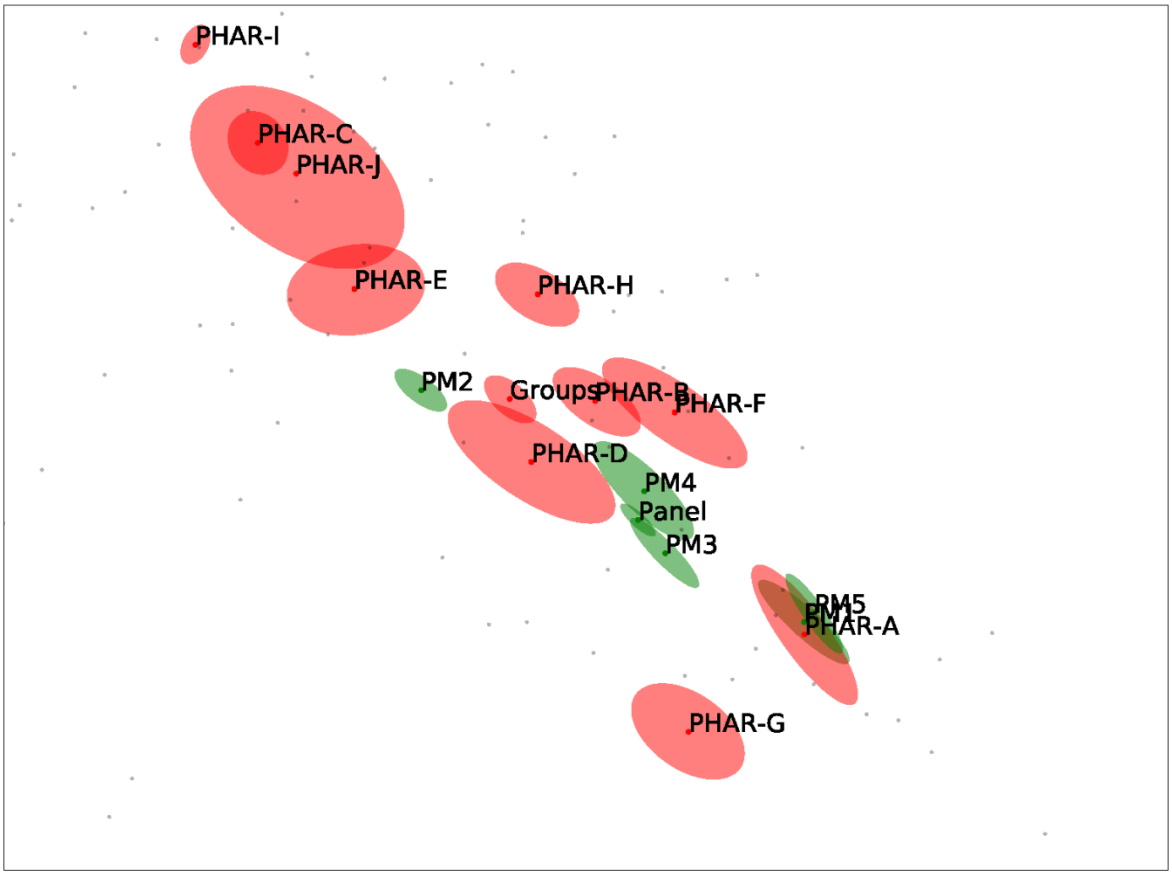


Figure 16. Barycenter overlay map of Pharmaceutical Sciences panel, panel members (PM), research groups and research groups together (groups) with their confidence regions

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 $\mathrm{C} = (\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_\mathrm{N})$, where:

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

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 $\mathrm{M} = (\mathrm{m_j})_\mathrm{j}$. The denominator is the L1-norm of the unnormalized vector.

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 Pharmaceutical Sciences 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" "Pharmaceutical
Sciences research groups and Panel_WoS SCs.xlsx"
```

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Acoustics	al Economi	tural Engir	Dairy	Anire, Multidi	Agronomy	Allergy	my	Morpl	Andrology	esthesiolon	throplog	Area 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

	A	B	C	D	E	F	G	H	I	J	K
1		Acoustics	al Economi	tural Engir	Dairy	Anire, Multidi	Agronomy	Allergy	my	Morpl	Andrology	esthesiolon
2	**Groups together**	1.52E-05	0	0.008242	0.001748	0.009736	0.005371	0.003889	0.014173	0.010126	0.004774	
3	**PHAR-A**	0	0	0.010176	0.001293	0.01189	0.00658	0.001009	0.011096	0.006522	0.001463	
4	**PHAR-B**	0	0	0.008046	0.000904	0.013324	0.00969	0.00338	0.014056	0.010261	0.003487	
5	**PHAR-C**	0	0	0.003249	0.001728	0.004288	0.001463	0.004363	0.019367	0.010831	0.010968	
6	**PHAR-D**	0	0	0.007471	0.000849	0.009606	0.003691	0.004677	0.012453	0.009221	0.004847	
7	**PHAR-E**	7.97E-05	0	0.004817	0.0022	0.006723	0.002853	0.005394	0.017509	0.011799	0.00566	
8	**PHAR-F**	0	0	0.00657	0.001845	0.011337	0.004166	0.003019	0.014296	0.010186	0.002451	
9	**PHAR-G**	0	0	0.020555	0.002772	0.015669	0.00971	0.001827	0.005521	0.004757	0.000919	
10	**PHAR-H**	0	0	0.007948	0.002144	0.012729	0.009766	0.005882	0.015823	0.011041	0.003032	
11	**PHAR-I**	0	0	0	0.000499	0.000457	0	0.005743	0.019261	0.041649	0.005851	
12	**PHAR-J**	0	0	0.003944	0.00275	0.00575	0.002742	0.002326	0.02069	0.010907	0.009802	
13	**PM1**	0	0	0.008071	0.000708	0.006508	0.002473	0.002118	0.009152	0.006563	0.00238	
14	**PM2**	3.44E-05	0	0.00611	0.000844	0.00738	0.000949	0.005901	0.01495	0.010561	0.007563	
15	**PM3**	0.000311	0	0.006876	0.001327	0.00631	0.001442	0.004175	0.012577	0.008682	0.004151	
16	**PM4**	0.000503	0	0.00646	0.001252	0.006093	0.002061	0.003659	0.01217	0.010314	0.002457	
17	**PM5**	3.2E-05	0	0.009154	0.003048	0.009958	0.003709	0.00237	0.012815	0.007589	0.000903	
18	**Panel together**	0.000162	0	0.007123	0.001372	0.007221	0.001839	0.004123	0.012883	0.008995	0.004306	

Figure 18. Excerpt of SAPV of the Pharmaceutical Sciences 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.

	A	B	C	D	E	F	G	H	I	J
1		Groups together	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H
2	Groups together	0	0.047792346	0.024743126	0.050284101	0.023303786	0.024460758	0.037939791	0.076916629	0.030015214
3	PHAR-A	0.047792346	0	0.046332165	0.092808908	0.046228047	0.067528164	0.046701358	0.072416297	0.060901783
4	PHAR-B	0.024743126	0.046332165	0	0.065794312	0.021604503	0.036549021	0.026211624	0.086721734	0.019068685
5	PHAR-C	0.050284101	0.092808908	0.065794312	0	0.062139914	0.035138944	0.074575994	0.114865323	0.061703748
6	PHAR-D	0.023303786	0.046228047	0.021604503	0.062139914	0	0.034574891	0.028142954	0.087085337	0.033469575
7	PHAR-E	0.024460758	0.067528164	0.036549021	0.035138944	0.034574891	0	0.046114151	0.097530749	0.033717905
8	PHAR-F	0.037939791	0.046701358	0.026211624	0.074575994	0.028142954	0.046114151	0	0.096553768	0.037039979
9	PHAR-G	0.076916629	0.072416297	0.086721734	0.114865323	0.087085337	0.097530749	0.096553768	0	0.092173244
10	PHAR-H	0.030015214	0.060901783	0.019068685	0.061703748	0.033469575	0.033717905	0.037039979	0.092173244	0
11	PHAR-I	0.125368779	0.156610679	0.139447835	0.102300323	0.134734074	0.118013764	0.145383398	0.167760426	0.137039576
12	PHAR-J	0.049051072	0.087661809	0.063927648	0.018654143	0.063088286	0.037196707	0.072403199	0.112243531	0.060521024
13	PM1	0.057358912	0.032837928	0.055321452	0.096617573	0.045547405	0.072457028	0.046032931	0.090656708	0.070365019
14	PM2	0.022902609	0.061898341	0.031339984	0.044596988	0.022620995	0.019918992	0.040587163	0.094071998	0.034094655
15	PM3	0.031827369	0.036849764	0.036116944	0.068810919	0.022660578	0.044116484	0.036130696	0.086318197	0.048119038
16	PM4	0.03291696	0.03575912	0.035004297	0.071642523	0.023170273	0.045539767	0.031063869	0.086739089	0.047480615
17	PM5	0.049729358	0.017981071	0.048229918	0.090823337	0.047441633	0.065240492	0.04672131	0.083769544	0.060840055
18	Panel together	0.026621799	0.03524437	0.030103004	0.066164897	0.017753635	0.039783603	0.030879155	0.085042474	0.042755384

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

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.027	0.035	0.030	0.066	0.018	0.04	0.031	0.085	0.043	0.137	0.065
PM1	0.057	**0.033**	0.055	0.097	**0.046**	0.072	**0.046**	**0.091**	0.070	**0.157**	0.094
PM2	0.023	0.062	**0.031**	**0.045**	**0.023**	**0.020**	**0.041**	**0.094**	**0.034**	**0.124**	**0.049**
PM3	0.032	0.037	**0.036**	0.069	**0.023**	0.044	**0.036**	**0.086**	**0.048**	**0.138**	**0.067**
PM4	0.033	**0.036**	**0.035**	0.072	**0.023**	0.046	**0.031**	**0.087**	**0.047**	**0.138**	**0.070**
PM5	0.050	**0.018**	0.048	0.091	0.047	0.065	0.047	**0.084**	0.061	**0.156**	**0.085**

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.046 (SD 0.034). 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).

The Euclidean distance between vectors a and b in $\mathbf{R}^N$ is:

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

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 'Pharmaceutical Sciences research groups and panel_WoS SCs-sa-vectors.xlsx' (Figure 19). From the calculated matrix of

pairwise Euclidean distances between SAPVs of Pharmaceutical Sciences groups, panel members, groups together, and panel together (Figure 18) we extract Table 4 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.

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.

We determine SAPVs of all entities (Figure 18). We take the WoS SCs similarity matrix Excel file (Figure 17). We copy PHAR-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 'PHAR-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 "PHAR-B_similarity matrix.xlsx" Sheet1
```

A	B	C	D	E	F
id	**label**	**x**	**y**	**weight**	**cluster**
1	Chemistry, Analytical	-0.7063	0.0097	13.8836	1
2	PHAR-B	-0.2368	0.1395	20	3
3	Engineering, Mechanical	-0.9306	-0.1852	13.2042	1
4	History Philosophy Of Science	0.4428	0.1276	1.393	1
5	Neuroimaging	0.264	0.1107	12.4462	1
6	Andrology	-0.0228	0.2258	16.9563	1
7	Urology Nephrology	0.1384	0.2274	7.1291	1
8	Parasitology	-0.1119	0.2297	17.5416	1
9	Orthopedics	0.2745	0.1843	6.171	1
10	Medicine, General Internal	0.2987	0.1728	24.867	1
11	Physics, Fluids Plasmas	-0.9787	-0.2242	11.1486	1
12	Business	1.3026	-0.1792	8.317	1
13	Primary Health Care	0.3913	0.1589	12.3628	1
14	Biochemistry Molecular Biology	-0.1931	0.2001	29.6091	1
15	Biophysics	-0.2537	0.1661	29.4281	1
16	Public, Environmental Occupati	0.6911	0.0716	16.4362	1
17	Education, Scientific Disciplines	0.5138	0.1147	2.5975	1
18	Cell Tissue Engineering	-0.0404	0.2178	26.2851	1
19	Medical Informatics	0.1809	0.0063	13.9865	1

Figure 20. Excerpt of PHAR-B.csv file

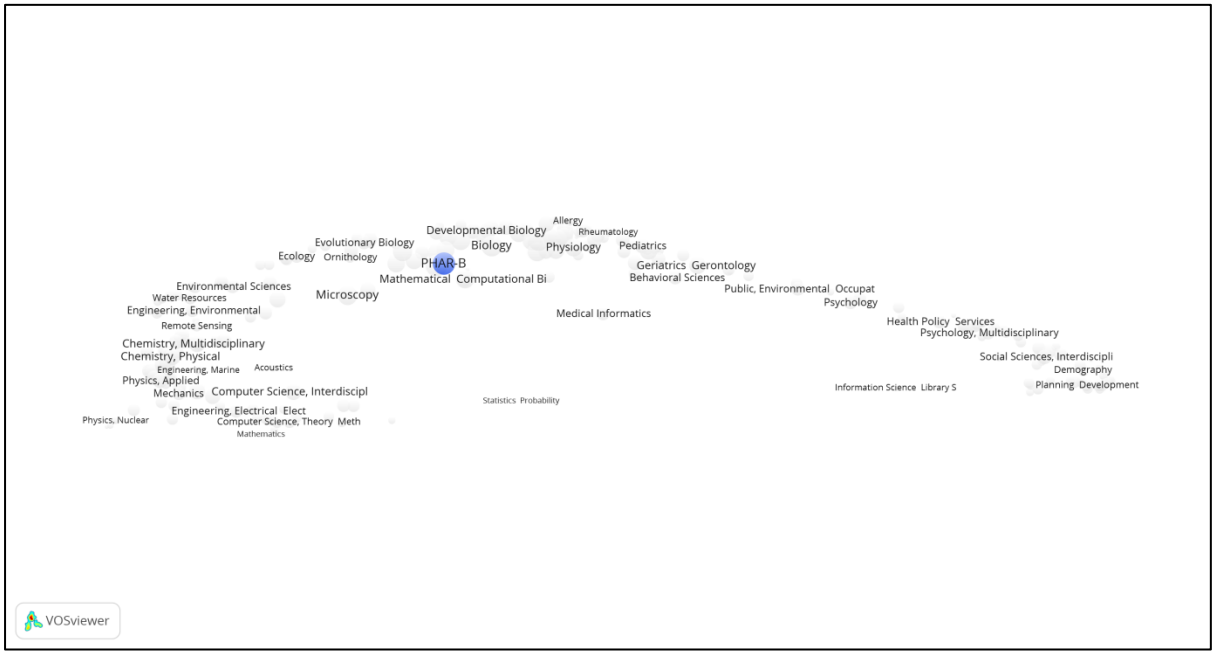


Figure 21. Location of the SAPV of PHAR-B in the WoS SCs similarity matrix

This program yields an output file named ‘PHAR-B_similarity matrix.net’. We create a map based on the network file using VOSviewer. It is not possible to easily locate PHAR-B in the map due to many different cluster colors. Therefore, we save the map data as ‘PHAR-B.txt’ file. In the text file, we can identify PHAR-B, but cannot easily change cluster number of all

the WoS SCs in the file that is necessary to highlight the PHAR-B's location in the overlay map. Therefore, we import the data from the 'PHAR-B.txt' file to 'PHAR-B.xlsx' file.

In the 'PHAR-B.xlsx' file we first identify the PHAR-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 PHAR-B and save as CSV file (Figure 20). We open the file with VOSviewer to visualize the SAPV 'location' of PHAR-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:

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

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

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 ('Pharmaceutical Sciences 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" "Pharmaceutical
Sciences research groups and panel_WoS SCs.xlsx"
```

This program calculates the WCS value in an output file as 'Pharmaceutical Sciences research groups and panel_WoS SCs-cosine.xlsx' (Figure 22).

	A	B	C	D	E	F	G	H	I	J
1		**Groups together**	**PHAR-A**	**PHAR-B**	**PHAR-C**	**PHAR-D**	**PHAR-E**	**PHAR-F**	**PHAR-G**	**PHAR-H**
2	**Groups together**	1	0.826755	0.913983	0.780488	0.904633	0.935919	0.844876	0.578366	0.900637
3	**PHAR-A**	0.826755226	1	0.816015	0.421797	0.754472	0.67119	0.777864	0.606813	0.725176
4	**PHAR-B**	0.913982847	0.816015	1	0.581479	0.951054	0.833646	0.899353	0.421791	0.970202
5	**PHAR-C**	0.780487722	0.421797	0.581479	1	0.610945	0.86827	0.528428	0.221417	0.619268
6	**PHAR-D**	0.904633259	0.754472	0.951054	0.610945	1	0.841726	0.90739	0.389753	0.920391
7	**PHAR-E**	0.935918988	0.67119	0.833646	0.86827	0.841726	1	0.799967	0.344549	0.858922
8	**PHAR-F**	0.844875553	0.777864	0.899353	0.528428	0.90739	0.799967	1	0.316758	0.861903
9	**PHAR-G**	0.578365971	0.606813	0.421791	0.221417	0.389753	0.344549	0.316758	1	0.384742
10	**PHAR-H**	0.900637033	0.725176	0.970202	0.619268	0.920391	0.858922	0.861903	0.384742	1
11	**PHAR-I**	0.319808709	0.08511	0.149126	0.383366	0.178343	0.323444	0.123203	0.055261	0.155084
12	**PHAR-J**	0.723844133	0.452826	0.516502	0.943035	0.501154	0.781697	0.463809	0.239146	0.543293
13	**PM1**	0.788319935	0.822537	0.819461	0.426273	0.881445	0.693717	0.909269	0.386875	0.740031
14	**PM2**	0.931624878	0.702427	0.902448	0.746551	0.956223	0.914117	0.839689	0.390582	0.897299
15	**PM3**	0.904014583	0.819797	0.884138	0.632341	0.94056	0.856191	0.886199	0.407056	0.84994
16	**PM4**	0.881215814	0.812429	0.876911	0.569505	0.928648	0.835525	0.922841	0.40353	0.835053
17	**PM5**	0.785739076	0.961892	0.742058	0.439642	0.685403	0.679361	0.718591	0.531437	0.674887
18	**Panel together**	0.933102529	0.863355	0.912936	0.640178	0.951393	0.874103	0.905391	0.450693	0.872321

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

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.

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.933	0.863	0.913	0.640	0.951	0.874	0.905	0.451	0.872	0.203	0.570
PM1	0.788	0.823	**0.819**	0.426	**0.881**	0.694	**0.909**	**0.387**	0.740	0.092	**0.375**
PM2	0.932	0.702	**0.902**	**0.747**	**0.956**	0.914	**0.840**	**0.391**	**0.897**	**0.274**	**0.623**
PM3	0.904	0.820	**0.884**	**0.632**	**0.941**	0.856	**0.886**	**0.407**	**0.850**	**0.203**	**0.558**
PM4	0.881	0.812	**0.877**	0.570	**0.929**	0.836	**0.923**	**0.404**	**0.835**	**0.192**	**0.488**
PM5	0.786	**0.962**	0.742	0.440	0.685	0.679	0.719	**0.531**	0.675	0.095	**0.488**

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 Pharmaceutical Sciences individual research groups, panel members, groups and panel using WoS SCs similarity matrix

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.067	0.137	0.087	0.360	0.049	0.126	0.095	0.549	0.128	0.797	0.430
PM1	0.212	0.177	0.181	0.574	0.119	0.306	0.091	0.613	0.260	0.908	0.625
PM2	0.068	0.298	**0.098**	**0.253**	**0.044**	**0.086**	0.160	0.609	**0.103**	**0.726**	**0.377**
PM3	0.096	0.180	0.116	0.368	0.059	0.144	0.114	0.593	0.150	0.797	0.442
PM4	0.119	0.188	0.123	0.430	0.071	0.164	**0.077**	0.596	0.165	0.808	0.512
PM5	0.214	**0.038**	0.258	0.560	0.315	0.321	0.281	**0.469**	0.325	0.905	0.512

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 Pharmaceutical Sciences individual research groups and panel members

Pearson Spearman	Barycenter	SAPV	WCS
Barycenter	1.00	0.75	0.69
SAPV	0.73	1.00	0.87
WCS	0.66	0.95	1.00

In Table 7, the upper triangle refers to Pearson's correlations while the lower triangle refers to Spearman's correlations. Table 7 and Figure 23 show that there is moderate correlation between barycenter and WCS methods, while the correlation is strong between SAPV and WCS methods.

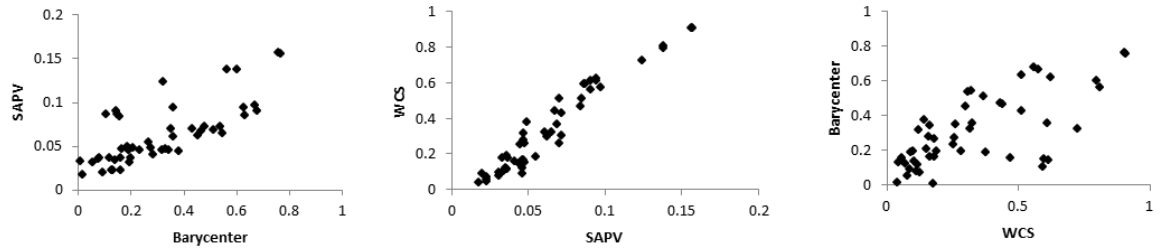


Figure 23. Scatter plot of the correlation between barycenter, SAPV and WCS

We calculate the Pearson's correlation coefficient (r) and the Spearman's rank-order correlation coefficient (ρ) between the three methods: barycenter, SAPV and WCS. These calculations are based on all Euclidean distances between barycenter and SAPV of individual research groups and panel members, and WCS value of individual research groups and panel members only. Although there are co-publications between groups, the barycenter distances between panel and combined group and separate groups, and combined groups and individual panel member can be (or at least are) considered independent, and have been included in the correlation calculation. The results are shown in Table 7.

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 'PHAR-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.

	A	B	C
1	**Source Titles**	**records**	**% of 376**
2	KIDNEY INTERNATIONAL	13	3.495
3	PLANTA MEDICA	11	2.957
4	ENVIRONMENTAL SCIENCE TECHNOLOGY	8	2.151
5	JOURNAL OF MASS SPECTROMETRY	7	1.882
6	CHEMOSPHERE	7	1.882
7	CARDIOVASCULAR RESEARCH	7	1.882
8	BRITISH JOURNAL OF PHARMACOLOGY	7	1.882
9	RAPID COMMUNICATIONS IN MASS SPECTROMETRY	6	1.613
10	JOURNAL OF MEDICINAL CHEMISTRY	6	1.613
11	JOURNAL OF ETHNOPHARMACOLOGY	6	1.613
12	JOURNAL OF ENVIRONMENTAL MONITORING	6	1.613
13	CIRCULATION	6	1.613
14	THROMBOSIS AND HAEMOSTASIS	5	1.344
15	PHARMAZIE	5	1.344

Groups Together | PHAR-A | PHAR-B | PHAR-C | PHAR-D | PHAR-E | PHAR-F | PHAR-G | PHAR-H | PH

Figure 24. Excerpt of Pharmaceutical Sciences research groups and panel_journal title.xlsx file

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 (PHAR-A, PHAR-B, PHAR-C, etc.), the panel members' code names, (PM1, PM2, PM3, etc.), Panel members together and Groups together. The Excel file is saved as 'Pharmaceutical Sciences research groups and panel_ journals title.xlsx' (Figure 24).

Publication statistics for each research groups and panel members have shown in the Table 1 and Table 2 respectively.

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 'Pharmaceutical Sciences 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 25).

	A	B	C
1	**Source Titles**	**records**	**% of 376**
2	KIDNEY INTERNATIONAL	13	3.495
3	PLANTA MEDICA	11	2.957
4	ENVIRONMENTAL SCIENCE TECHNOLOGY	8	2.151
5	JOURNAL OF MASS SPECTROMETRY	7	1.882
6	CHEMOSPHERE	7	1.882
7	CARDIOVASCULAR RESEARCH	7	1.882
8	BRITISH JOURNAL OF PHARMACOLOGY	7	1.882
9	RAPID COMMUNICATIONS IN MASS SPECTROM	6	1.613
10	JOURNAL OF MEDICINAL CHEMISTRY	6	1.613
11	JOURNAL OF ETHNOPHARMACOLOGY	6	1.613
12	JOURNAL OF ENVIRONMENTAL MONITORING	6	1.613
13	CIRCULATION	6	1.613
14	THROMBOSIS AND HAEMOSTASIS	5	1.344
15	PHARMAZIE	5	1.344
16	JOURNAL OF THE AMERICAN SOCIETY OF NEPH	5	1.344
17	EUROPEAN JOURNAL OF PHARMACEUTICS ANI	5	1.344
18	BIOORGANIC MEDICINAL CHEMISTRY LETTERS	5	1.344
19	ARTERIOSCLEROSIS THROMBOSIS AND VASCUL	5	1.344
20	ACTA CARDIOLOGICA	5	1.344

GroupsTogether | PanelTogether

	A	B	C
1	**Source Titles**	**records**	**% of 1032**
2	PHARMACEUTICAL RESEARCH	52	5.053
3	BRITISH JOURNAL OF CLINICAL PHARMACOLOG	35	3.401
4	ARCHIV DER PHARMAZIE	35	3.401
5	CLINICAL PHARMACOLOGY THERAPEUTICS	27	2.624
6	MONATSHEFTE FUR CHEMIE	23	2.235
7	INTERNATIONAL JOURNAL OF PHARMACEUTIC	23	2.235
8	JOURNAL OF PHARMACOLOGY AND EXPERIMEN	22	2.138
9	JOURNAL OF CONTROLLED RELEASE	21	2.041
10	JOURNAL OF PHARMACEUTICAL SCIENCES	20	1.944
11	BIOCHEMICAL PHARMACOLOGY	19	1.846
12	ANALYTICAL CHEMISTRY	18	1.749
13	EUROPEAN JOURNAL OF CLINICAL PHARMACO	17	1.652
14	CLINICAL PHARMACOKINETICS	15	1.458
15	JOURNAL OF CHROMATOGRAPHY	14	1.361
16	JOURNAL OF PHARMACY AND PHARMACOLOGY	13	1.263
17	PHARMACEUTISCH WEEKBLAD SCIENTIFIC EDIT	12	1.166
18	BIOMATERIALS	12	1.166
19	ARZNEIMITTEL FORSCHUNG DRUG RESEARCH	12	1.166
20	XENOBIOTICA	11	1.069

GroupsTogether | PanelTogether

Figure 25. Excerpt of the Pharmaceutical Sciences panel and research groups together_journals title.xlsx file

	A	B	C	D
1		**Source Titles**	**records_x**	**records_y**
2	**0**	KIDNEY INTERNATIONAL	13	0
3	**1**	PLANTA MEDICA	11	1
4	**2**	ENVIRONMENTAL SCIENCE TECHNOLOGY	8	0
5	**3**	JOURNAL OF MASS SPECTROMETRY	7	9
6	**4**	CHEMOSPHERE	7	0
7	**5**	CARDIOVASCULAR RESEARCH	7	0
8	**6**	BRITISH JOURNAL OF PHARMACOLOGY	7	3
9	**7**	RAPID COMMUNICATIONS IN MASS SPECTROMETRY	6	7
10	**8**	JOURNAL OF MEDICINAL CHEMISTRY	6	7
11	**9**	JOURNAL OF ETHNOPHARMACOLOGY	6	0
12	**10**	JOURNAL OF ENVIRONMENTAL MONITORING	6	0
13	**11**	CIRCULATION	6	0
14	**12**	THROMBOSIS AND HAEMOSTASIS	5	5
15	**13**	PHARMAZIE	5	7
16	**14**	JOURNAL OF THE AMERICAN SOCIETY OF NEPHROLOGY	5	0
17	**15**	EUROPEAN JOURNAL OF PHARMACEUTICS AND BIOPHARMACEUTICS	5	9
18	**16**	BIOORGANIC MEDICINAL CHEMISTRY LETTERS	5	1

Figure 26. Excerpt of the Pharmaceutical Sciences panel and research groups together_journals title-joined.xlsx file

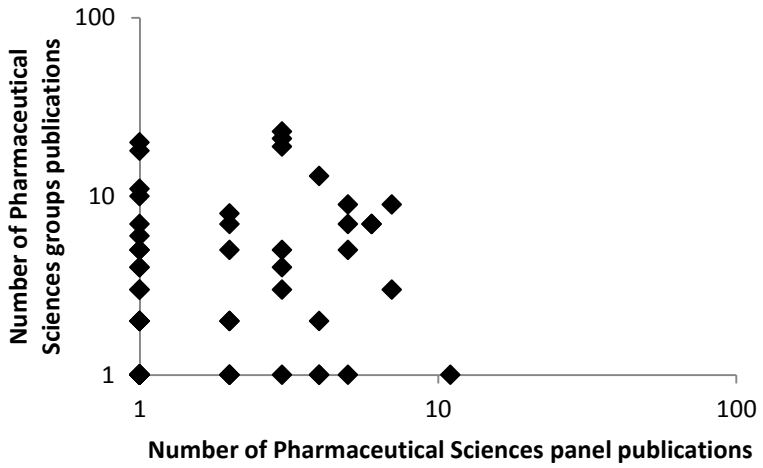


Figure 27. Log-log plot of the number of publications (log-log scale) per journals for the panel (horizontal axis) and research groups together (vertical axis) of the Pharmaceutical Sciences department

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 "Pharmaceutical Sciences panel and research groups
together_journals title.xlsx"
```

It produces a new Excel file named 'Pharmaceutical Sciences panel and research groups together_journals title-joined.xlsx' (Figure 26). 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 using SPSS and obtain negative correlation coefficient value ($r = -0.5$, $\rho = -0.47$). A log-log plot of the number of publications per journal for the Pharmaceutical Sciences panel and research groups together is shown in Figure 27.

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 'Pharmaceutical Sciences research groups and panel_journal title- joined.xlsx' (Figure 26) 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 "Pharmaceutical Sciences research groups and panel_
journals title-joined.xlsx"
```

The outcome shows that the Top-down correlation between Pharmaceutical Sciences research groups together and the panel based on the journals in which they publish is low (0.09).

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 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

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

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.

	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 UNIVERS	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 28. Excerpt of the journals VOS map data

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 28).

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.

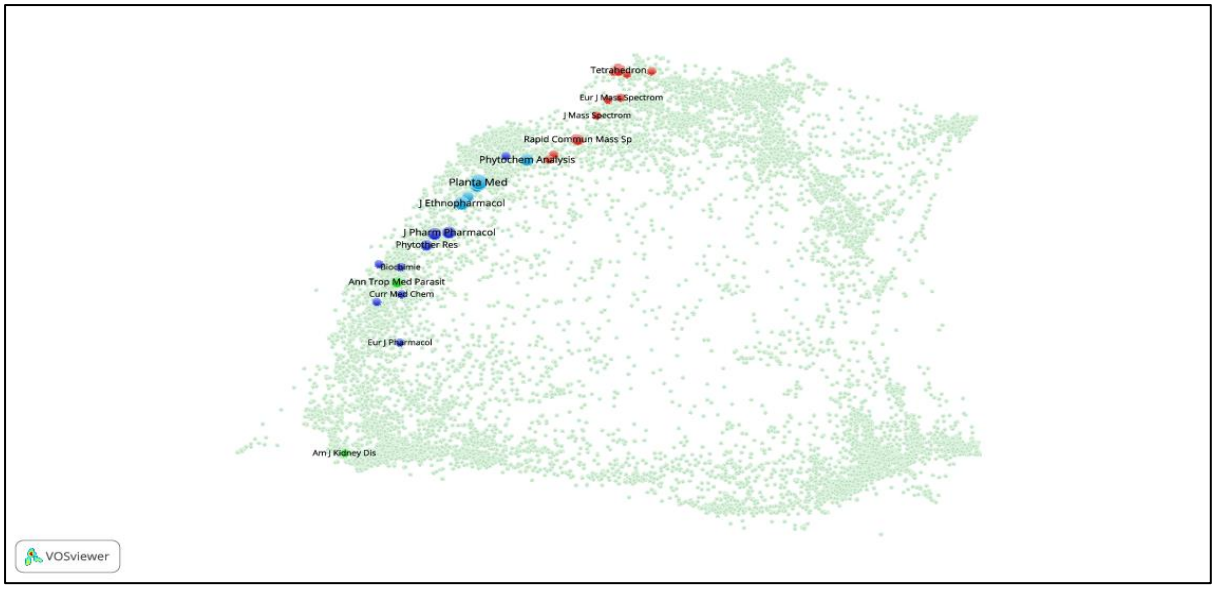


Figure 29. Journal overlay map of the PHAR-B research group

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 29 shows the journal overlay map of the PHAR-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, $\mathrm{m_j}$ is the number of publications in journal j, and $\mathrm{T} = \sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{m_j}$ is the total number of publications of the entity.

Based on formula 3, Python script ‘journals-barycenter.py’ is used. This script takes ‘Journals_VOS_map.xlsx’ (Figure 28) and ‘Pharmaceutical Sciences research groups and panel_journals title.xlsx’ (Figure 24) as input. We run the program as follows

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

At this point, we notice that our program indicates that the journal titles in our Pharmaceutical Sciences 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 30). 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 30. 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.

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 31) to translate 'old' titles to 'correct' titles.

We keep the 'Pharmaceutical Sciences research groups and panel_journals title.xlsx' (Figure 24), 'Journals_VOS_map.xlsx' (Figure 28), 'translate.xlsx' (Figure 30), and 'Journal name change.xlsx' (Figure 31) 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" "Pharmaceutical Sciences
research groups and panel journals title.xlsx"
```

This program calculates the barycenter coordinates of Pharmaceutical Sciences individual research groups, panel members, groups and panel in the journals VOS map in an output file 'Pharmaceutical Sciences research groups and panel_ journals title -barycenters.xlsx' (Figure 32).

	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 31. Excerpt of journal name change.xlsx file

	A	B	C
1		**x**	**y**
2	**Groups Together**	-0.54660515	0.117209756
3	**PHAR-A**	-0.25527838	0.503043243
4	**PHAR-B**	-0.46875	0.425341667
5	**PHAR-C**	-0.84614667	-0.23502
6	**PHAR-D**	-0.53415667	0.28592
7	**PHAR-E**	-0.66281452	-0.101977419
8	**PHAR-F**	-0.47425938	0.4249875
9	**PHAR-G**	-0.17799375	0.35755
10	**PHAR-H**	-0.60679189	0.274737838
11	**PHAR-I**	-0.95846296	-0.441433333
12	**PHAR-J**	-0.6813	-0.278011111
13	**PM1**	-0.26936408	0.602659223
14	**PM2**	-0.67866231	-0.116471733
15	**PM3**	-0.49688474	0.264682731
16	**PM4**	-0.48673137	0.265651961
17	**PM5**	-0.4150439	0.407007317
18	**Panel together**	-0.52095432	0.19065865

Figure 32. Barycenter coordinates of the Pharmaceutical Sciences individual research groups, panel members, research groups together, and panel using journal VOS map

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.

	A	B	C	D	E	F	G	H	I
1		Groups Tog	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G
2	Groups Together	0	0.483465	0.317816	0.462375	0.169169	0.248088	0.316166	0.440043
3	PHAR-A	0.4834654	0	0.227173	0.945443	0.353434	0.729476	0.232477	0.164746
4	PHAR-B	0.3178155	0.227173	0	0.760596	0.154001	0.561895	0.005521	0.298555
5	PHAR-C	0.4623753	0.945443	0.760596	0	0.60722	0.226519	0.757569	0.893066
6	PHAR-D	0.1691689	0.353434	0.154001	0.60722	0	0.408677	0.151418	0.363294
7	PHAR-E	0.248088	0.729476	0.561895	0.226519	0.408677	0	0.559683	0.667994
8	PHAR-F	0.3161662	0.232477	0.005521	0.757569	0.151418	0.559683	0	0.303844
9	PHAR-G	0.4400429	0.164746	0.298555	0.893066	0.363294	0.667994	0.303844	0
10	PHAR-H	0.1686343	0.419148	0.204297	0.563155	0.073491	0.380858	0.200349	0.436722
11	PHAR-I	0.6940526	1.177499	0.995549	0.234992	0.842068	0.450154	0.992541	1.116918
12	PHAR-J	0.4175431	0.889686	0.734767	0.17036	0.582812	0.177002	0.732853	0.810713
13	PM1	0.5590383	0.100607	0.266826	1.017047	0.412842	0.807042	0.2712	0.261586
14	PM2	0.2684141	0.750368	0.581055	0.205194	0.427552	0.021476	0.578756	0.689468
15	PM3	0.155629	0.339396	0.163104	0.609661	0.042898	0.402458	0.161894	0.332138
16	PM4	0.1600624	0.33155	0.160699	0.616321	0.051575	0.407623	0.159823	0.322124
17	PM5	0.3182625	0.186408	0.056749	0.773336	0.169853	0.566088	0.061885	0.242154
18	Panel together	0.0777991	0.410083	0.240419	0.535679	0.096172	0.325208	0.238936	0.381411

Figure 33. Excerpt of Euclidean distances matrix of the barycenter of the Pharmaceutical Sciences groups, panel members, research groups together and panel using the journal VOS map

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.078	0.410	0.240	0.536	0.096	0.325	0.239	0.381	0.120	0.769	0.495
PM1	0.559	**0.101**	0.267	1.017	0.413	0.807	**0.271**	**0.262**	0.471	1.251	0.972
PM2	0.268	0.750	0.581	**0.205**	0.428	**0.021**	0.579	0.689	0.398	**0.429**	**0.162**
PM3	0.156	**0.339**	**0.163**	0.610	**0.043**	0.402	**0.162**	**0.332**	**0.110**	0.844	**0.573**
PM4	0.160	**0.332**	**0.161**	0.616	**0.052**	0.408	**0.160**	**0.322**	**0.120**	0.850	**0.577**
PM5	0.318	**0.186**	**0.057**	0.773	**0.170**	0.566	**0.062**	**0.242**	**0.233**	1.008	0.735

For each research group we determined the panel member at the shortest distance. Average shortest distances is 0.124 (SD 0.013). 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).

From the matrix of Euclidean distances between all entity pairs (Figure 33), 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.

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 28) file and manually add the Pharmaceutical Sciences individual groups, panel members, research groups together and panel's coordinates (Figure 32) 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 Pharmaceutical Sciences department in the journal level.csv'. After that, we open the file with VOSviewer to visualize the barycenters (Figure 34). Figure 35 shows a zoomed in version of Figure 34.

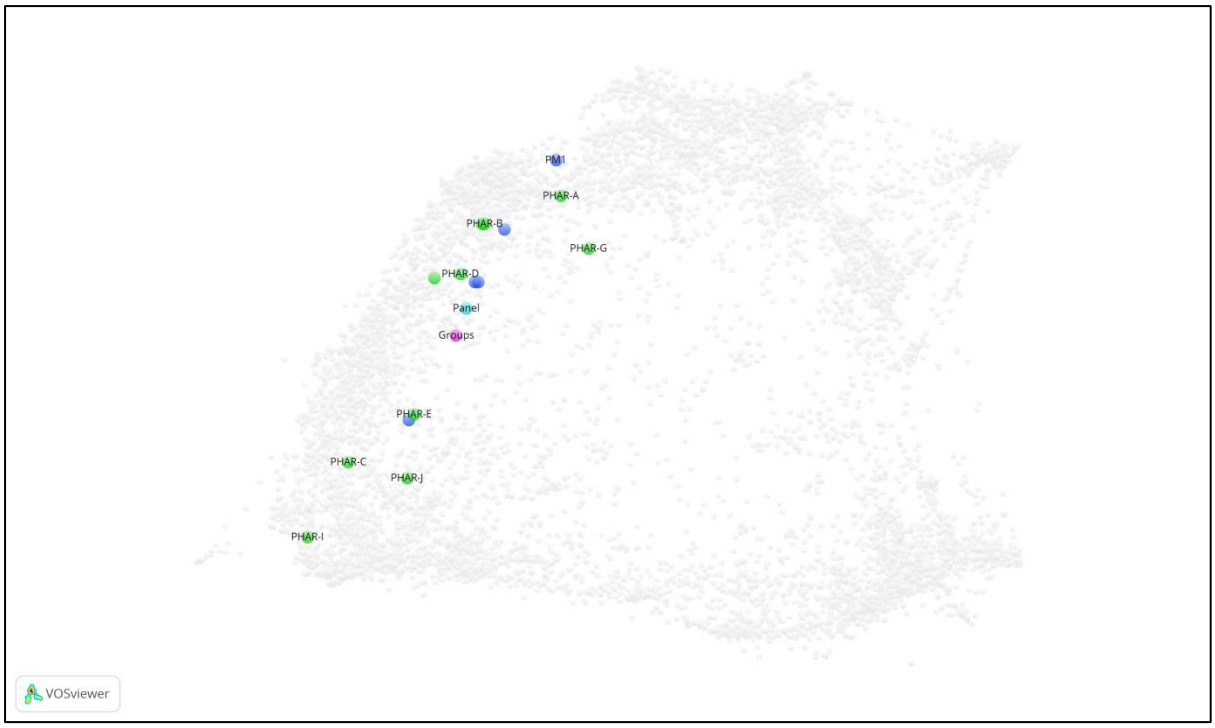


Figure 34. Barycenter overlay map of Pharmaceutical Sciences panel, panel members (PM), research groups and research groups together

We also create the barycenter overlap map of Pharmaceutical Sciences 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 36). 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 36 (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.

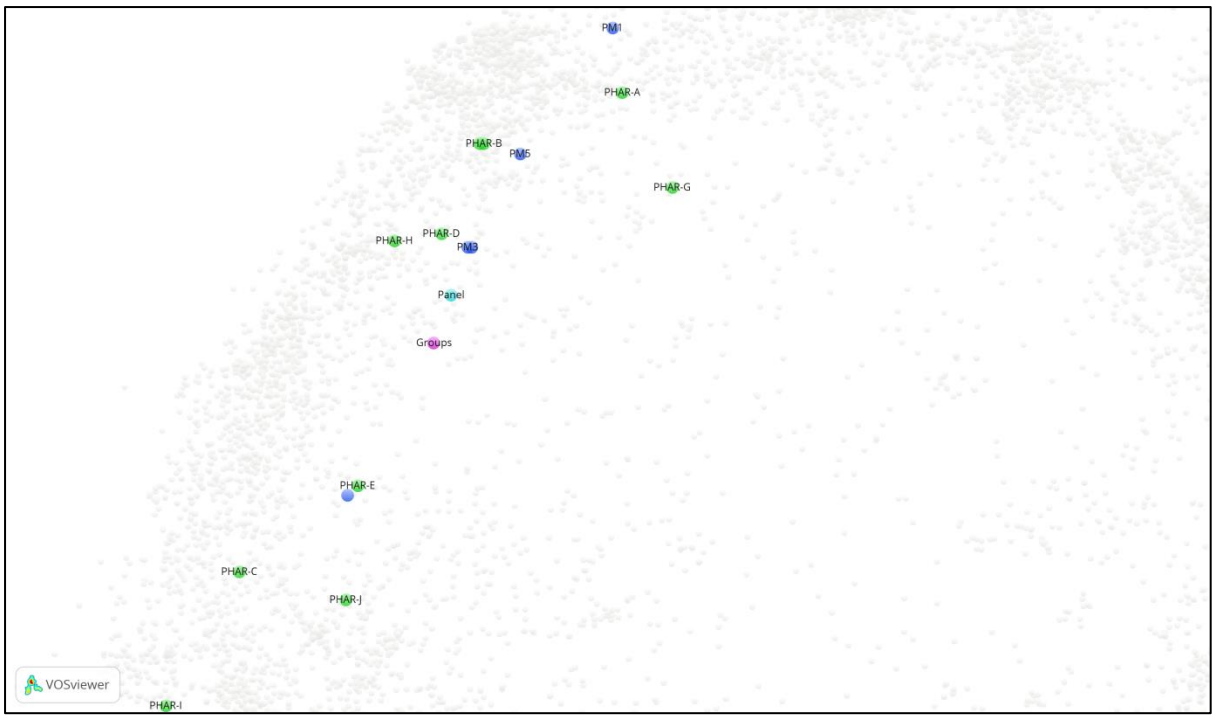


Figure 35. Barycenter overlay map of Pharmaceutical Sciences panel, panel members (PM), research groups and research groups together (zoomed)

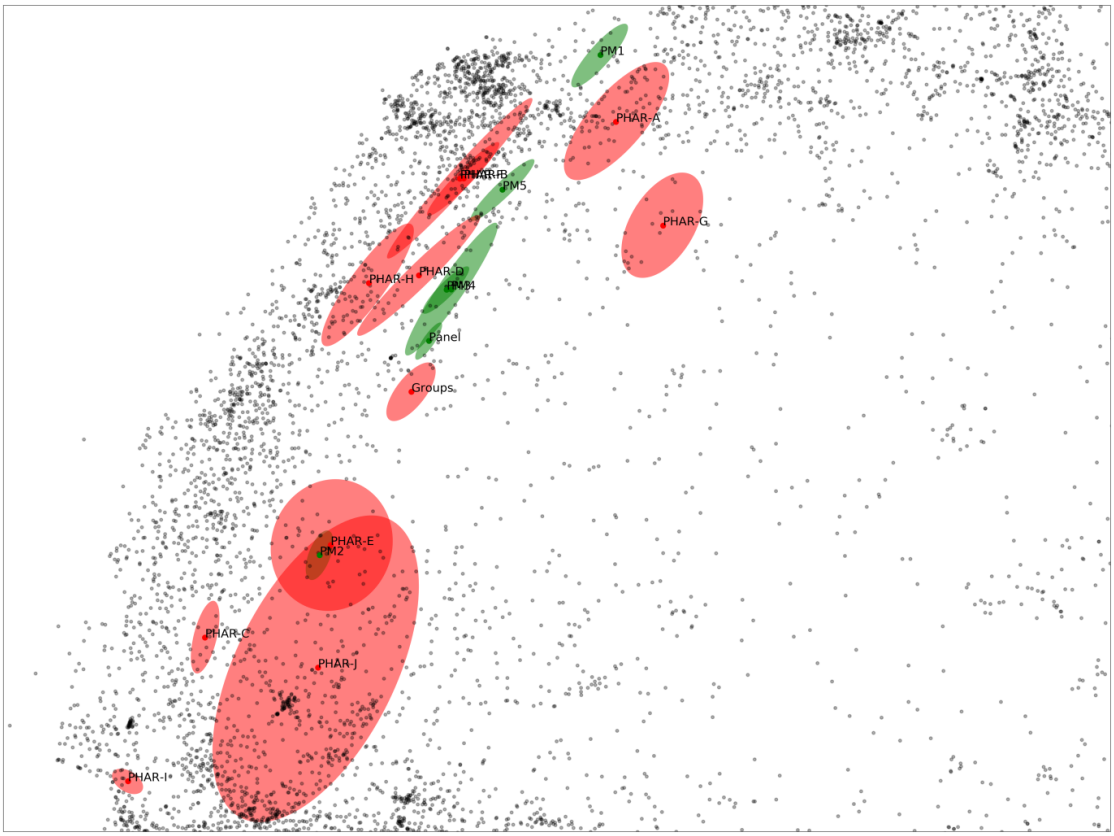


Figure 36. Barycenter overlay map of Pharmaceutical Sciences panel, panel members (PM), research groups and research groups together with their confidence regions

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 $C = (C_1, C_2, \ldots, C_N)$, where:

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

where 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 $M = (m_j)_j$. The denominator is the L1-norm of the unnormalized vector.

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

	A	B	C	D	E	F	G	H	I	J
1		4OR A QU	OHN JOUR	PG BULLE	PS JOURN	PHARMSC	TCC REVI	EMINAL IM	TISCHEN S	ND APPLIE
2	**Groups Together**	4.39E-07	7.14E-05	1.8E-05	0.000362	0.000166	8.25E-05	3.79E-05	2.07E-06	1.06E-05
3	**PHAR-A**	2.45E-07	4.42E-05	4.42E-05	0.000302	8.17E-05	0.000127	1.48E-05	8.46E-07	1.04E-05
4	**PHAR-B**	1.27E-08	3.43E-05	1.5E-05	0.000431	0.00019	9.03E-05	2.26E-05	2.64E-06	9.49E-06
5	**PHAR-C**	0	7.93E-05	1.39E-05	0.000279	3.02E-05	5.76E-05	5.03E-05	2.58E-06	1.17E-05
6	**PHAR-D**	2.34E-06	3.65E-05	1.22E-05	0.001246	0.001855	0.000163	3.26E-05	2.28E-06	1.01E-05
7	**PHAR-E**	9.96E-08	5.73E-05	1.54E-05	0.00032	3.91E-05	6.33E-05	4.01E-05	2.26E-06	1.25E-05
8	**PHAR-F**	0	2.57E-05	1.53E-05	0.000359	6.12E-05	8.82E-05	1.5E-05	1.98E-06	1.29E-05
9	**PHAR-G**	2.64E-06	0.000177	3.43E-05	0.000202	5.9E-05	0.000138	1.78E-05	1.45E-06	6.11E-06
10	**PHAR-H**	0	3.57E-05	1.46E-05	0.000421	0.00017	7.59E-05	2.37E-05	2.54E-06	1.14E-05
11	**PHAR-I**	1.63E-08	8.65E-05	9.25E-06	0.000313	2.23E-05	3.32E-05	8.08E-05	1.34E-06	7.91E-06
12	**PHAR-J**	4.24E-08	0.000142	1.09E-05	0.000201	2.79E-05	4.92E-05	4.26E-05	1.44E-06	9.29E-06
13	**PM1**	6.07E-07	2.39E-05	1.45E-05	0.000461	0.000353	0.000119	1.29E-05	1.16E-06	9.43E-06
14	**PM2**	8.84E-08	7.56E-05	1.18E-05	0.000675	0.000252	5.9E-05	5.62E-05	2.13E-06	1.02E-05
15	**PM3**	5.35E-08	3.48E-05	1.67E-05	0.000769	0.00084	0.000137	3.49E-05	2.07E-06	1.41E-05
16	**PM4**	1.81E-07	5.04E-05	1.32E-05	0.000413	0.000171	6.91E-05	7.57E-05	1.51E-06	9.36E-06
17	**PM5**	1.37E-07	2.39E-05	2.44E-05	0.000322	7.52E-05	0.00013	9.54E-06	1.16E-06	1.52E-05
18	**Panel together**	1.39E-07	4.91E-05	1.58E-05	0.000593	0.000374	9.79E-05	4.05E-05	1.81E-06	1.21E-05

Figure 37. Excerpt of SAPV of the Pharmaceutical Sciences research groups, research groups together, panel members and panel using journal similarity matrix

matrix and the weights (number of publications) per journals ('Pharmaceutical Sciences groups and panel_ journals title.xlsx', Figure 24). 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 30), 'Journal name change.xlsx' (Figure 31) and 'Pharmaceutical Sciences research groups and panel_ journals title.xlsx' in a folder. We run the program as:

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

This program calculates the SAPV (Figure 37) of each entity and stores the results in an output file named 'Pharmaceutical Sciences research groups and panel_ journals title-SA-Vector.xlsx' (Figure 38).

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. 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).

	A	B	C	D	E	F	G	H	I	J	K
1		**Groups To**	**PHAR-A**	**PHAR-B**	**PHAR-C**	**PHAR-D**	**PHAR-E**	**PHAR-F**	**PHAR-G**	**PHAR-H**	**PHAR-I**
2	**Groups Together**	0	0.009135	0.007535	0.005523	0.010388	0.003872	0.007934	0.015666	0.00505	0.011932
3	**PHAR-A**	0.009135	0	0.007673	0.013677	0.01299	0.011886	0.010334	0.014331	0.009029	0.01823
4	**PHAR-B**	0.007535	0.007673	0	0.011763	0.01057	0.009692	0.008317	0.017079	0.004295	0.016828
5	**PHAR-C**	0.005523	0.013677	0.011763	0	0.013239	0.004786	0.011001	0.019406	0.008979	0.011239
6	**PHAR-D**	0.010388	0.01299	0.01057	0.013239	0	0.011884	0.012763	0.019315	0.009995	0.017508
7	**PHAR-E**	0.003872	0.011886	0.009692	0.004786	0.011884	0	0.008047	0.018332	0.006421	0.012345
8	**PHAR-F**	0.007934	0.010334	0.008317	0.011001	0.012763	0.008047	0	0.019313	0.007085	0.017016
9	**PHAR-G**	0.015666	0.014331	0.017079	0.019406	0.019315	0.018332	0.019313	0	0.017229	0.02223
10	**PHAR-H**	0.00505	0.009029	0.004295	0.008979	0.009995	0.006421	0.007085	0.017229	0	0.014871
11	**PHAR-I**	0.011932	0.01823	0.016828	0.011239	0.017508	0.012345	0.017016	0.02223	0.014871	0
12	**PHAR-J**	0.009111	0.015417	0.014572	0.005887	0.015668	0.009728	0.014901	0.020336	0.012744	0.011547
13	**PM1**	0.01343	0.010873	0.010997	0.017368	0.015091	0.01486	0.008182	0.020461	0.012212	0.021431
14	**PM2**	0.004706	0.012471	0.010336	0.005301	0.010987	0.004475	0.010602	0.018189	0.007639	0.011052
15	**PM3**	0.006353	0.010437	0.009072	0.009465	0.006845	0.00701	0.008043	0.017985	0.007013	0.015373
16	**PM4**	0.006017	0.009884	0.008229	0.009484	0.011256	0.006613	0.006137	0.018066	0.006583	0.014237
17	**PM5**	0.006741	0.007132	0.008291	0.010369	0.012024	0.007719	0.00672	0.016689	0.006977	0.016651
18	**Panel together**	0.003377	0.009146	0.007643	0.00719	0.009022	0.004346	0.006694	0.017111	0.005171	0.013426

Figure 38. Excerpt of pairwise Euclidean distances matrix between the SAPV of the Pharmaceutical Sciences individual research groups, panel members, panel and research groups together using the journal similarity matrix

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.003	0.009	0.008	0.007	0.009	0.004	0.007	0.017	0.005	0.013	0.011
PM1	0.013	**0.011**	**0.011**	0.017	0.015	0.015	**0.008**	**0.020**	0.012	0.021	0.020
PM2	0.005	0.012	**0.010**	**0.005**	0.011	**0.004**	**0.011**	**0.018**	**0.008**	**0.011**	**0.008**
PM3	0.006	**0.010**	**0.009**	0.009	**0.007**	0.007	**0.008**	**0.018**	**0.007**	0.015	**0.013**
PM4	0.006	**0.010**	**0.008**	0.009	0.011	**0.007**	**0.006**	**0.018**	**0.007**	0.014	**0.012**
PM5	0.007	**0.007**	**0.008**	0.010	0.012	0.008	**0.007**	**0.017**	**0.007**	0.017	**0.014**

For each research group we determined the panel member at the shortest distance. Average shortest distance is 0.008 (SD 0.004). 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).

From the calculated matrix of pairwise Euclidean distances between SAPVs of Pharmaceutical Sciences individual research groups, panel members, research groups together and panel (Figure 38), 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.

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:

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

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

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, M^k denotes the column matrix of publications of panel member PMk 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 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 30), 'Journal name change.xlsx' (Figure 31) and 'Pharmaceutical Sciences 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 ('Pharmaceutical Sciences research groups and panel_ journals title.xlsx', Figure 24), 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  "Pharmaceutical Sciences research groups and panel_journals title.xlsx"
```

This program calculates the WCS value in an output file as 'Pharmaceutical Sciences research groups and panel _ journals title-cosine.xlsx' (Figure 39).

From the calculated WCS value matrix of Pharmaceutical Sciences individual research groups, panel members, research groups together and panel in journals (Figure 39), 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.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Groups To	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
2	Groups Together	1	0.586845	0.674276	0.769649	0.46011	0.828342	0.659293	0.463693	0.743112	0.445641	0.609733
3	PHAR-A	0.586845	1	0.624652	0.236482	0.205362	0.317021	0.351313	0.33862	0.518663	0.079184	0.23816
4	PHAR-B	0.674276	0.624652	1	0.330596	0.408225	0.445481	0.509055	0.17952	0.92458	0.119518	0.227102
5	PHAR-C	0.769649	0.236482	0.330596	1	0.233559	0.781313	0.465626	0.109137	0.444227	0.338521	0.831167
6	PHAR-D	0.46011	0.205362	0.408225	0.233559	1	0.277159	0.225169	0.087121	0.429884	0.096725	0.152277
7	PHAR-E	0.828342	0.317021	0.445481	0.781313	0.277159	1	0.729258	0.157276	0.601067	0.317461	0.529201
8	PHAR-F	0.659293	0.351313	0.509055	0.465626	0.225169	0.729258	1	0.094334	0.616731	0.159189	0.260541
9	PHAR-G	0.463693	0.33862	0.17952	0.109137	0.087121	0.157276	0.094334	1	0.159229	0.049146	0.130876
10	PHAR-H	0.743112	0.518663	0.92458	0.444227	0.429884	0.601067	0.616731	0.159229	1	0.172337	0.277847
11	PHAR-I	0.445641	0.079184	0.119518	0.338521	0.096725	0.317461	0.159189	0.049146	0.172337	1	0.277385
12	PHAR-J	0.609733	0.23816	0.227102	0.831167	0.152277	0.529201	0.260541	0.130876	0.277847	0.277385	1
13	PM1	0.504251	0.39145	0.501802	0.205825	0.297796	0.444512	0.850172	0.103374	0.521015	0.074858	0.124838
14	PM2	0.711447	0.286275	0.407696	0.644707	0.469267	0.717466	0.455987	0.160661	0.517927	0.298935	0.478292
15	PM3	0.607959	0.266433	0.365109	0.420662	0.871454	0.527409	0.442068	0.120945	0.462292	0.159628	0.263035
16	PM4	0.685345	0.373733	0.519894	0.457779	0.354474	0.710798	0.847416	0.127815	0.621341	0.207418	0.300149
17	PM5	0.678987	0.812121	0.468922	0.444415	0.24052	0.578108	0.527156	0.270586	0.470988	0.153451	0.325191
18	Panel together	0.826337	0.504097	0.543229	0.608275	0.646329	0.772198	0.692259	0.203506	0.639602	0.25325	0.423993

Figure 39. Excerpt of WCS value matrix of the Pharmaceutical Sciences individual research groups, panel members, groups and panel using the journal similarity matrix

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.826	0.504	0.543	0.608	0.646	0.772	0.692	0.204	0.640	0.253	0.424
PM1	0.504	0.391	**0.502**	0.206	0.298	0.445	**0.850**	0.103	**0.521**	0.075	0.125
PM2	0.711	0.286	**0.408**	**0.645**	0.469	**0.717**	0.456	**0.161**	**0.518**	**0.299**	**0.478**
PM3	0.608	0.266	**0.365**	0.421	**0.871**	0.527	0.442	0.121	**0.462**	0.160	0.263
PM4	0.685	0.374	**0.520**	0.458	0.354	0.711	**0.847**	0.128	**0.621**	0.207	0.300
PM5	0.679	**0.812**	**0.469**	0.444	0.241	0.578	0.527	**0.271**	0.471	0.153	0.325

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

	Groups	PHAR-A	PHAR-B	PHAR-C	PHAR-D	PHAR-E	PHAR-F	PHAR-G	PHAR-H	PHAR-I	PHAR-J
Panel	0.174	0.496	0.457	0.392	0.354	0.228	0.308	0.796	0.360	0.747	0.576
PM1	0.496	0.609	**0.498**	0.794	0.702	0.555	**0.150**	0.897	**0.479**	0.925	0.875
PM2	0.289	0.714	**0.592**	**0.355**	0.531	**0.283**	0.544	**0.839**	**0.482**	**0.701**	**0.522**
PM3	0.392	0.734	**0.635**	0.579	**0.129**	0.473	0.558	0.879	**0.538**	0.840	0.737
PM4	0.315	0.626	**0.480**	0.542	0.646	0.289	**0.153**	0.872	**0.379**	0.793	0.700
PM5	0.321	**0.188**	**0.531**	0.556	0.759	0.422	0.473	**0.729**	0.529	0.847	0.675

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

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.

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.

4 Heat map

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. For the sake of simplicity, the results are shown under the WCS method. We create an MS Excel file (Figure 40) containing:

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

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

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

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

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

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

	A	B	C	D	E	F
1	Barycenter_ PHAR_WoS SCs	Barycenter_ PHAR_Journals	SAPV_ PHAR_WoS SCs	SAPV_ PHAR_Journals	WCS_ PHAR_WoS SCs	WCS_ PHAR_Journals
2	0.007	0.101	0.033	0.011	0.177	0.609
3	0.266	0.267	0.055	0.011	0.181	0.498
4	0.670	1.017	0.097	0.017	0.574	0.794
5	0.318	0.413	0.046	0.015	0.119	0.702
6	0.538	0.807	0.072	0.015	0.306	0.555
7	0.189	0.271	0.046	0.008	0.091	0.150
8	0.144	0.262	0.091	0.020	0.613	0.897
9	0.353	0.471	0.070	0.012	0.260	0.479
10	0.757	1.251	0.157	0.021	0.908	0.925
11	0.624	0.972	0.094	0.020	0.625	0.875
12	0.451	0.750	0.062	0.012	0.298	0.714
13	0.194	0.581	0.031	0.010	0.098	0.592
14	0.232	0.205	0.045	0.005	0.253	0.355
15	0.130	0.428	0.023	0.011	0.044	0.531
16	0.095	0.021	0.020	0.004	0.086	0.283
17	0.283	0.579	0.041	0.011	0.160	0.544

Figure 40. Excerpt of the dissimilarities/distances between panel members and individual research groups according to each of the six approaches

We import data from the MS Excel file (Figure 40) to SPSS, and calculate the Spearman's rank-order correlation coefficient between the six approaches.

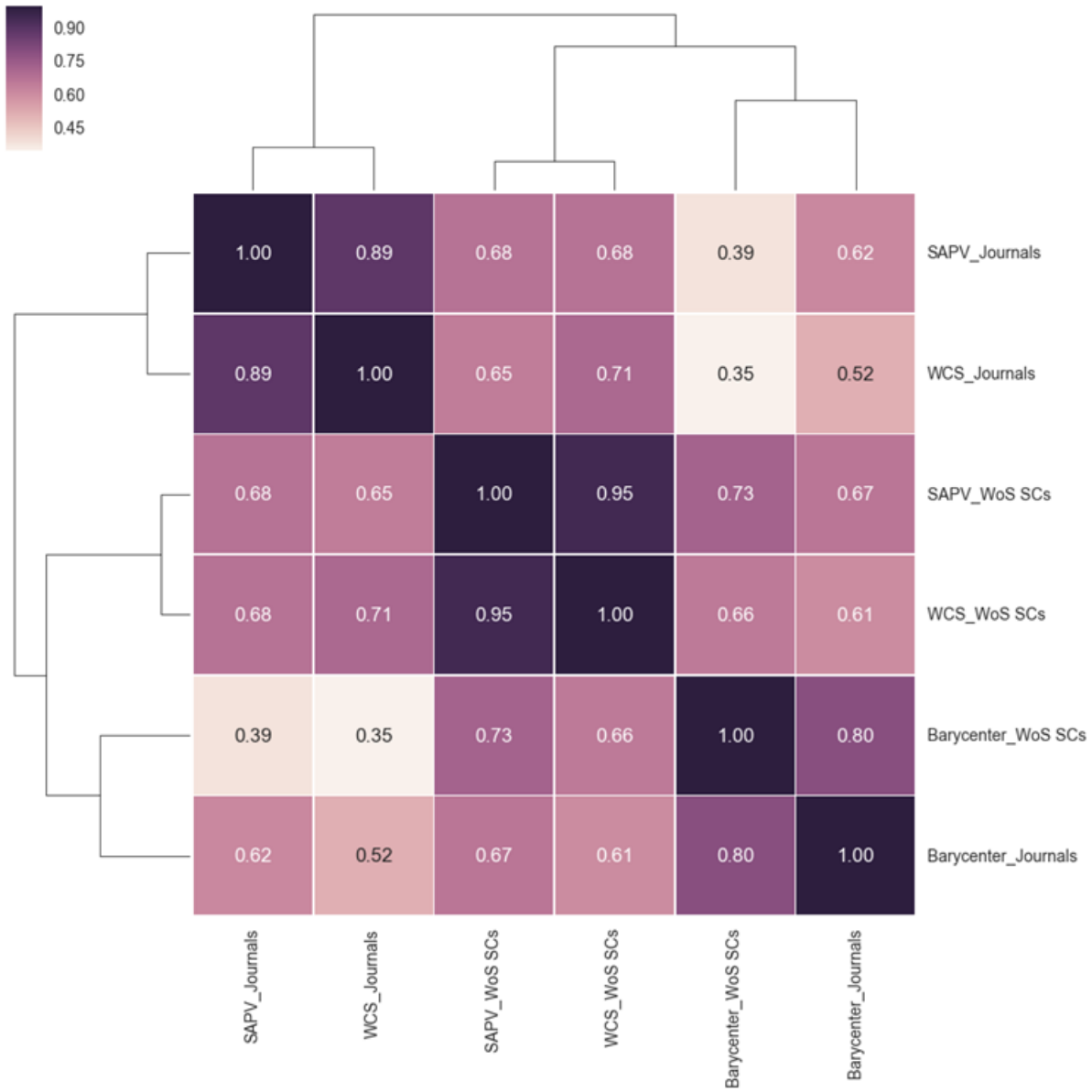


Figure 41. Heat map with hierarchical clustering based on correlation coefficient between six approaches in the Pharmaceuticals sciences department

The heat map with hierarchical clustering (Figure 41) shows that correlations between the two levels of aggregation based on barycenter ($\rho = 0.80$), SAPV ($\rho = 0.68$) and WCS ($\rho = 0.71$) are moderate to strong. The correlations between the barycenter methods on the one hand and the SAPV and WCD methods on the other are moderate to low. 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

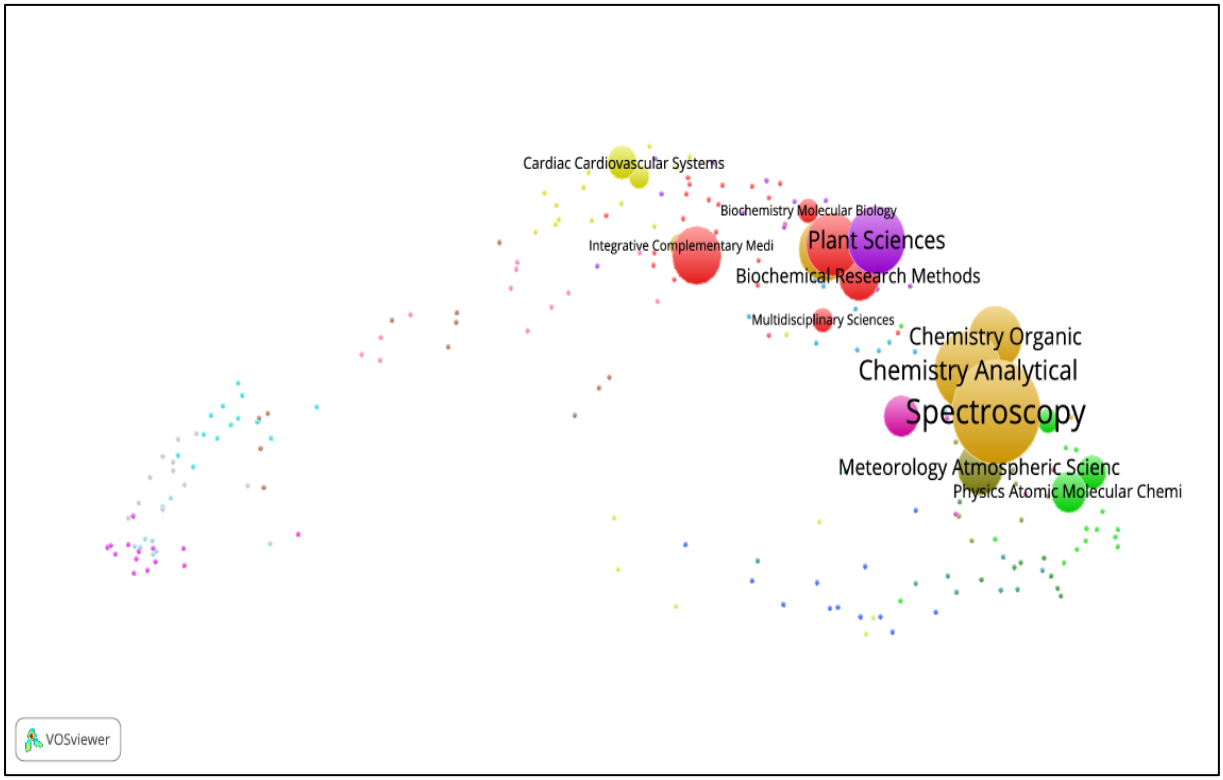


Figure 42. WoS SCs overlay map of PHAR-A research group's publications

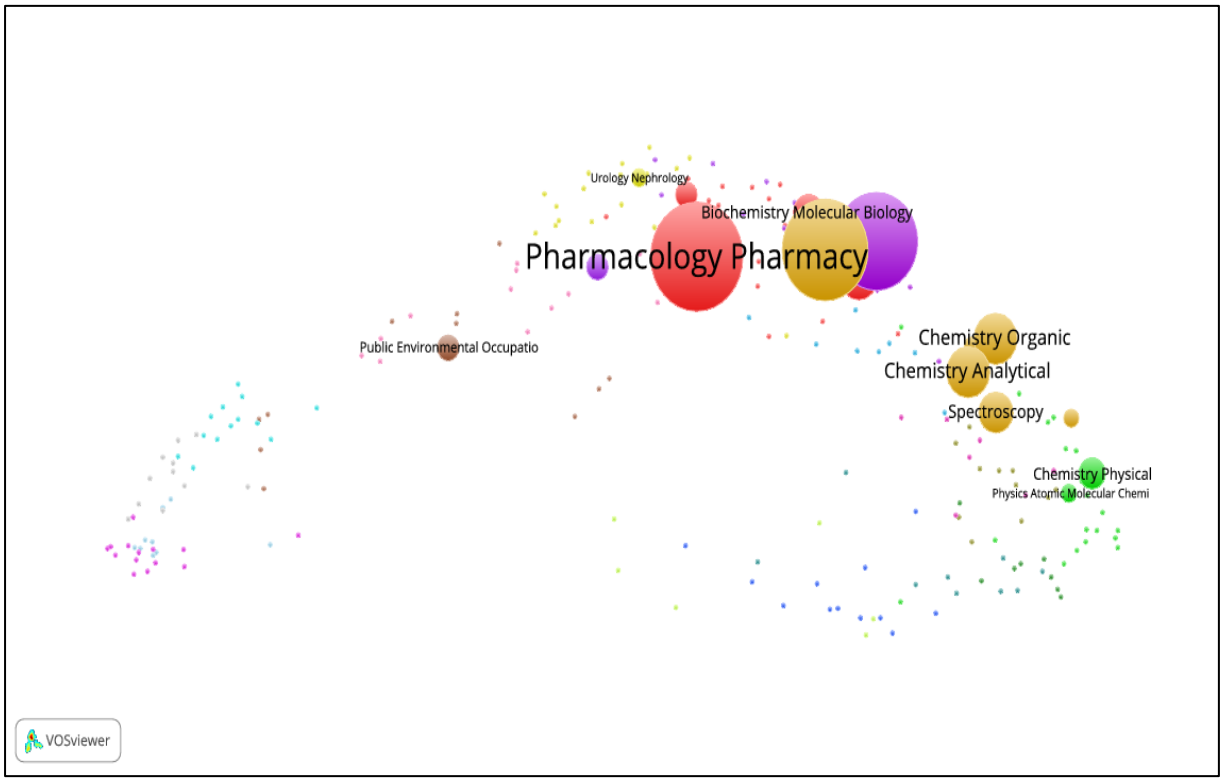


Figure 43. WoS SCs overlay map of PHAR-B research group's publications

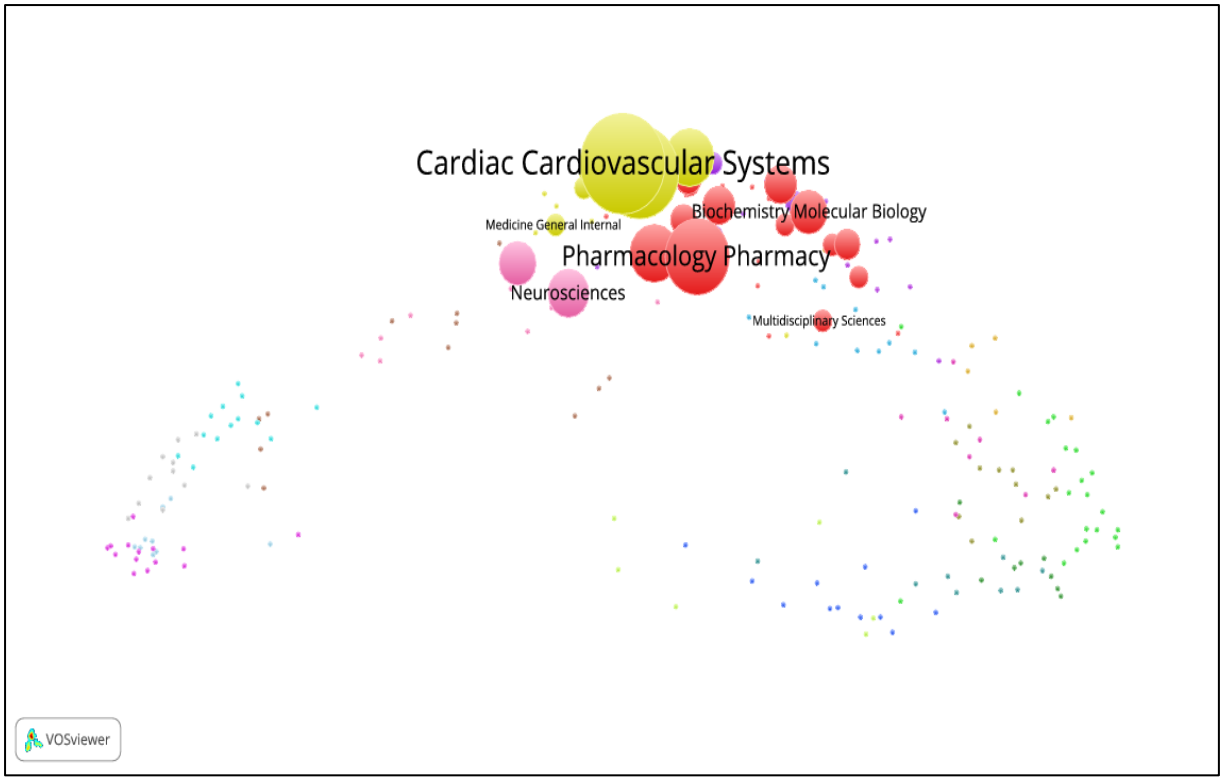


Figure 44. WoS SCs overlay map of PHAR-C research group's publications

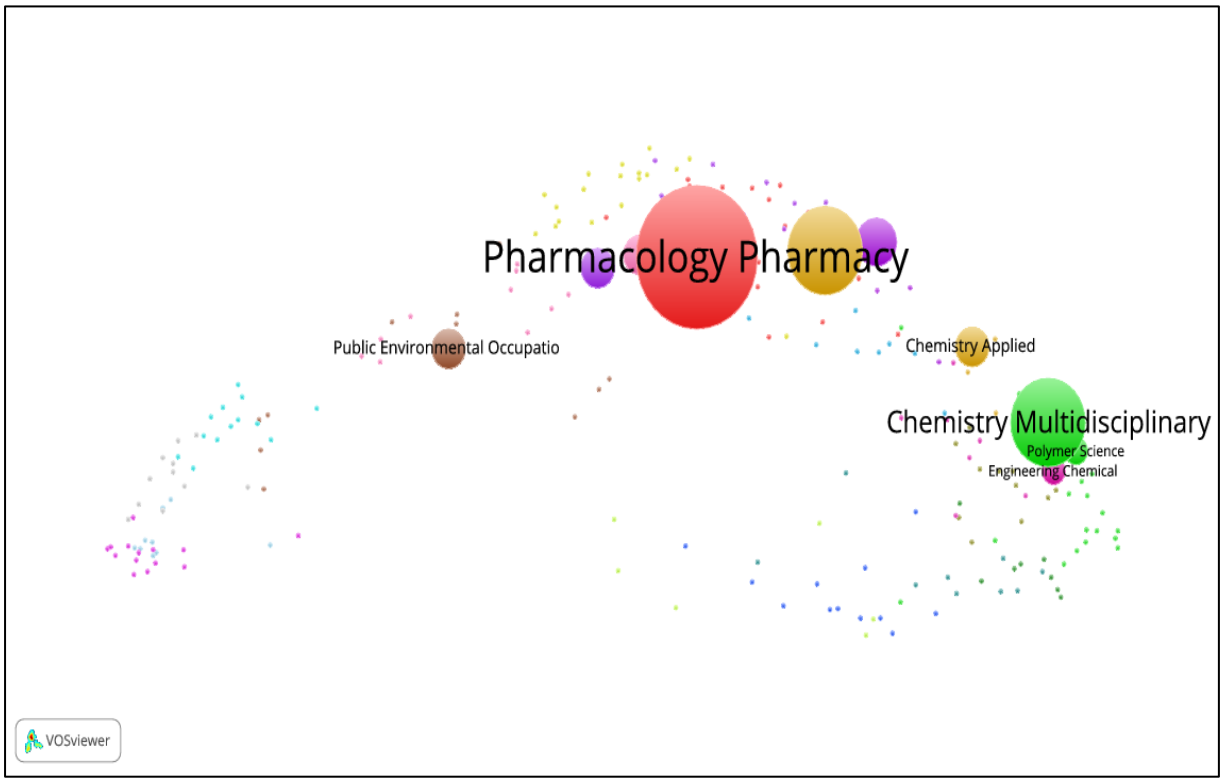


Figure 45. WoS SCs overlay map of PHAR-D research group's publications

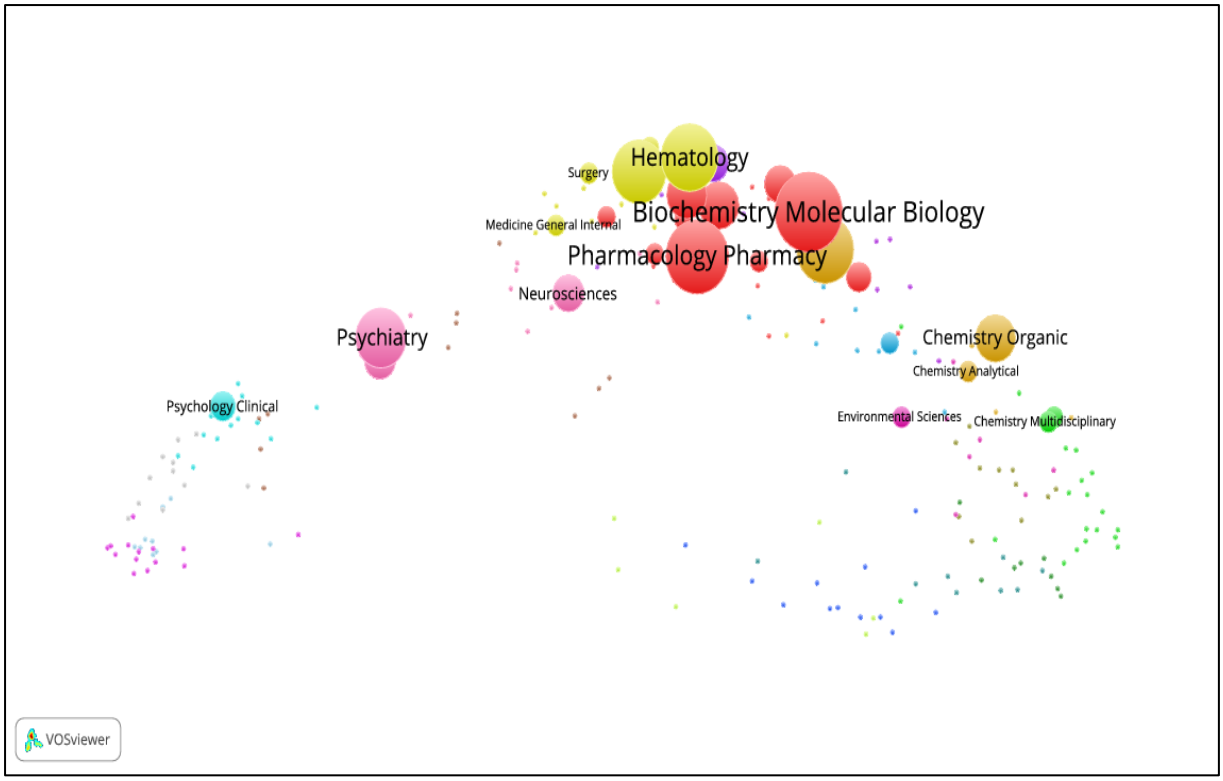


Figure 46. WoS SCs overlay map of PHAR-E research group's publications

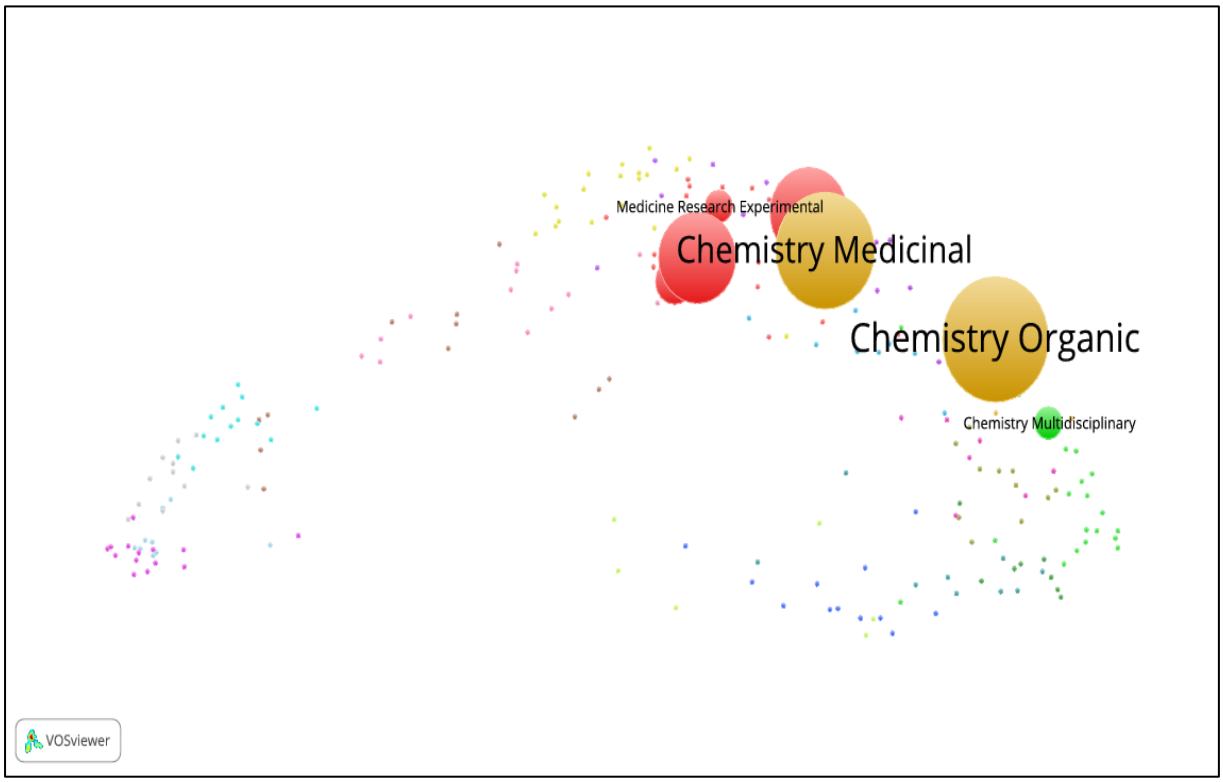


Figure 47. WoS SCs overlay map of PHAR-F research group's publications

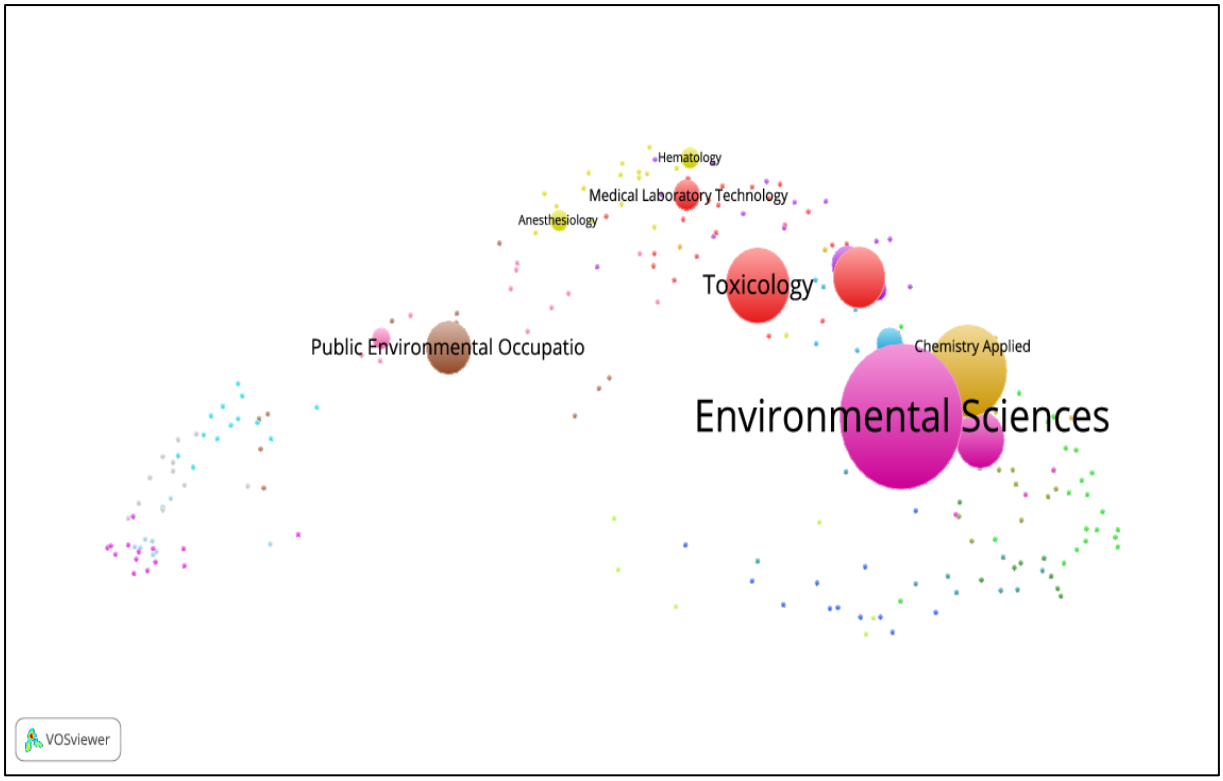


Figure 48. WoS SCs overlay map of PHAR-G research group's publications

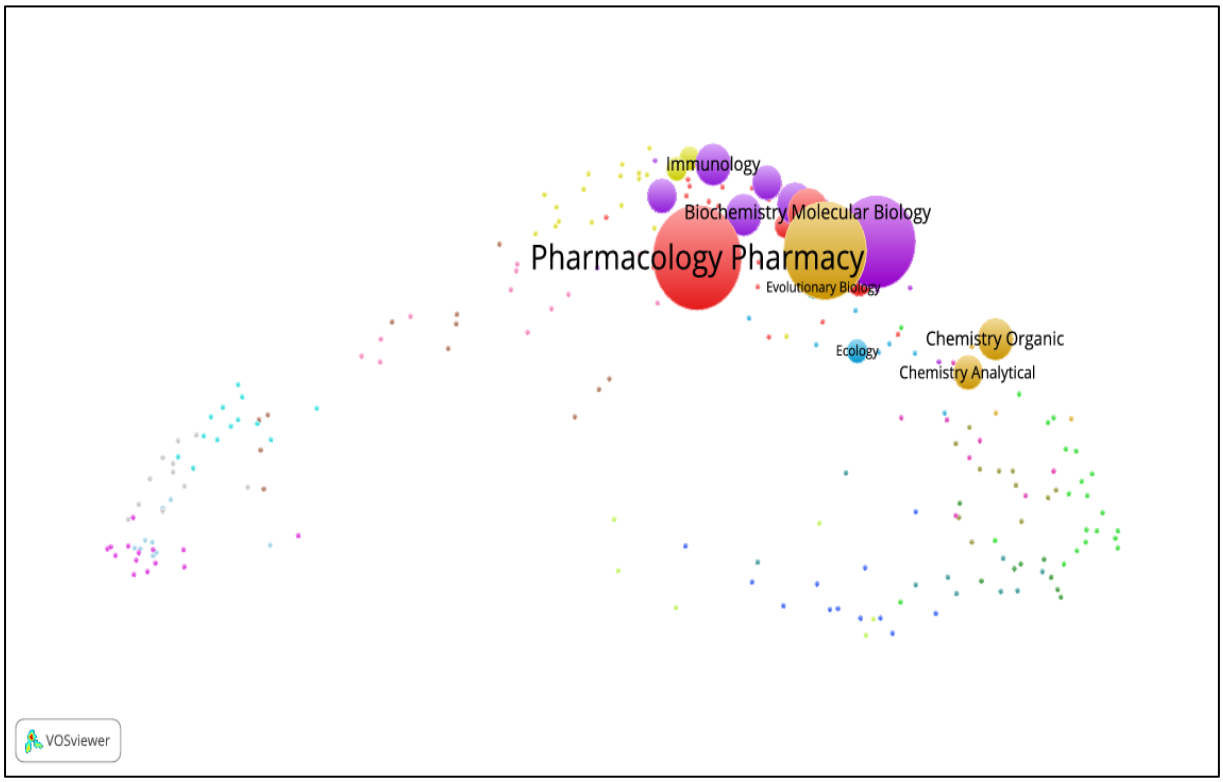


Figure 49. WoS SCs overlay map of PHAR-H research group's publications

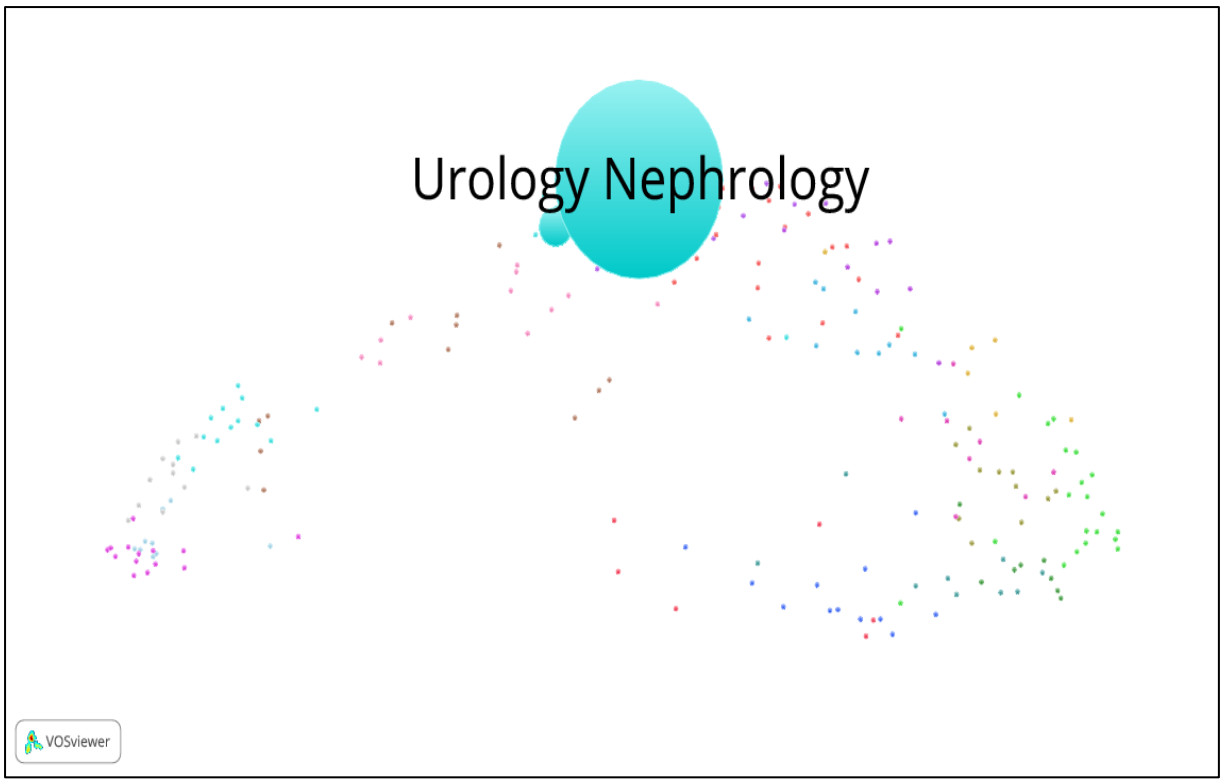


Figure 50. WoS SCs overlay map of PHAR-I research group's publications

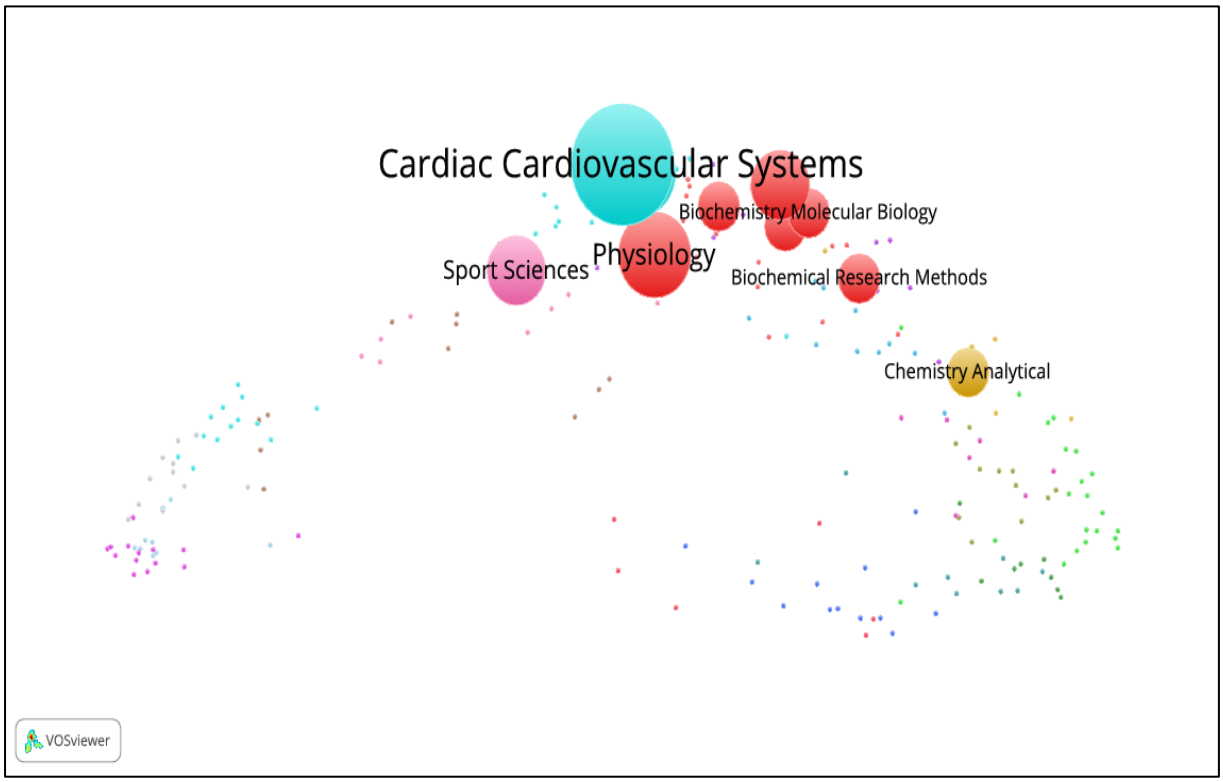


Figure 51. WoS SCs overlay map of PHAR-J research group's publications

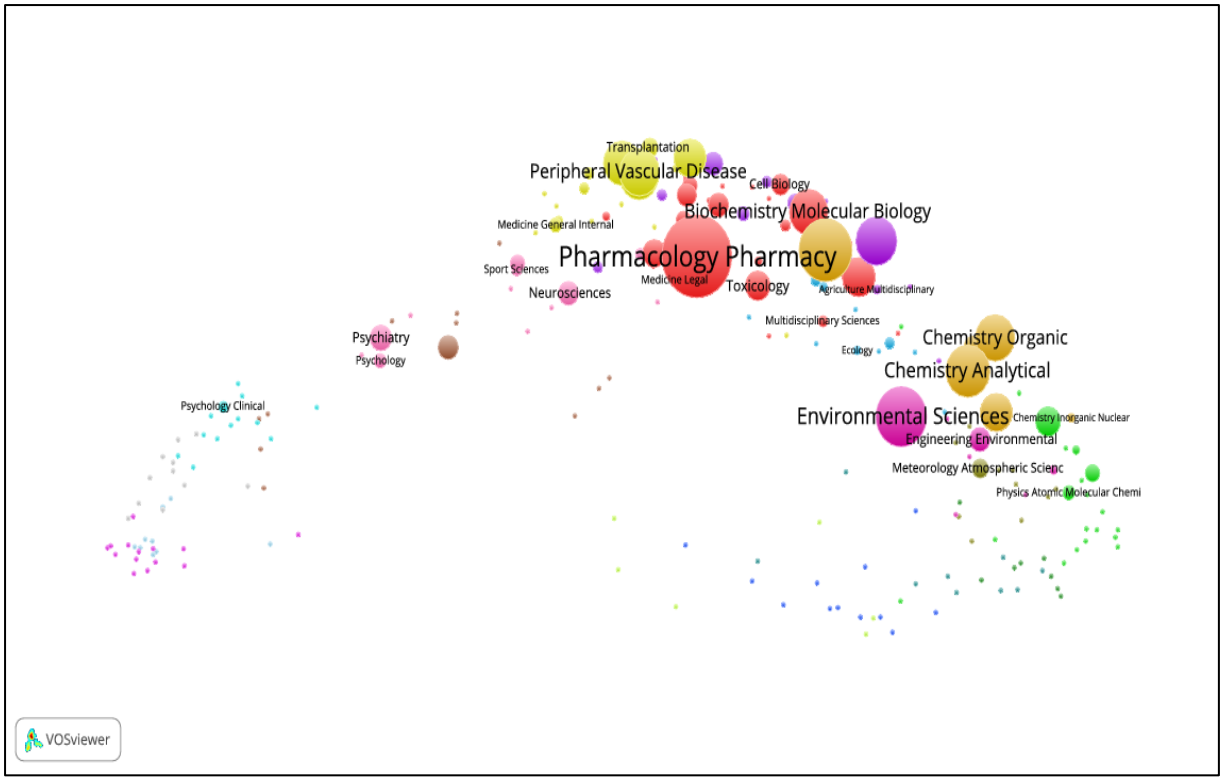


Figure 52. WoS SCs overlay map of Pharmacy research groups' publications

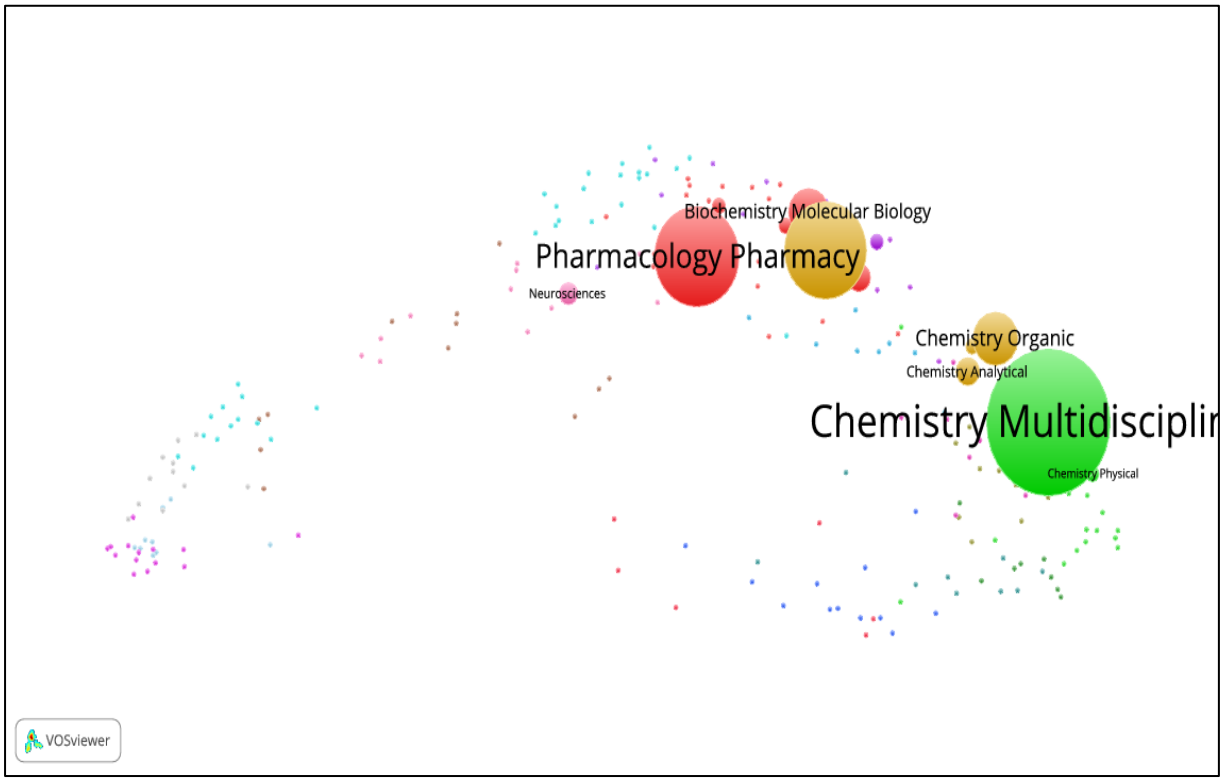


Figure 53. WoS SCs overlay map of PM1's publications

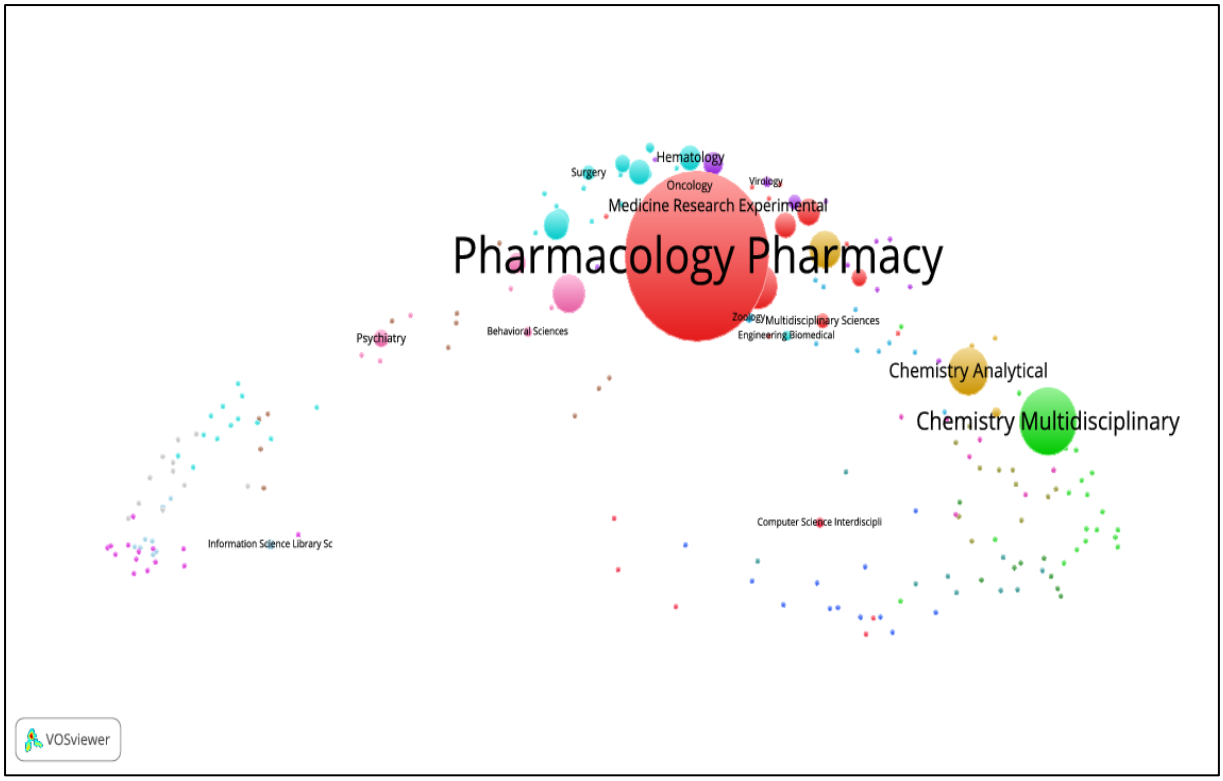


Figure 54. WoS SCs overlay map of PM2's publications

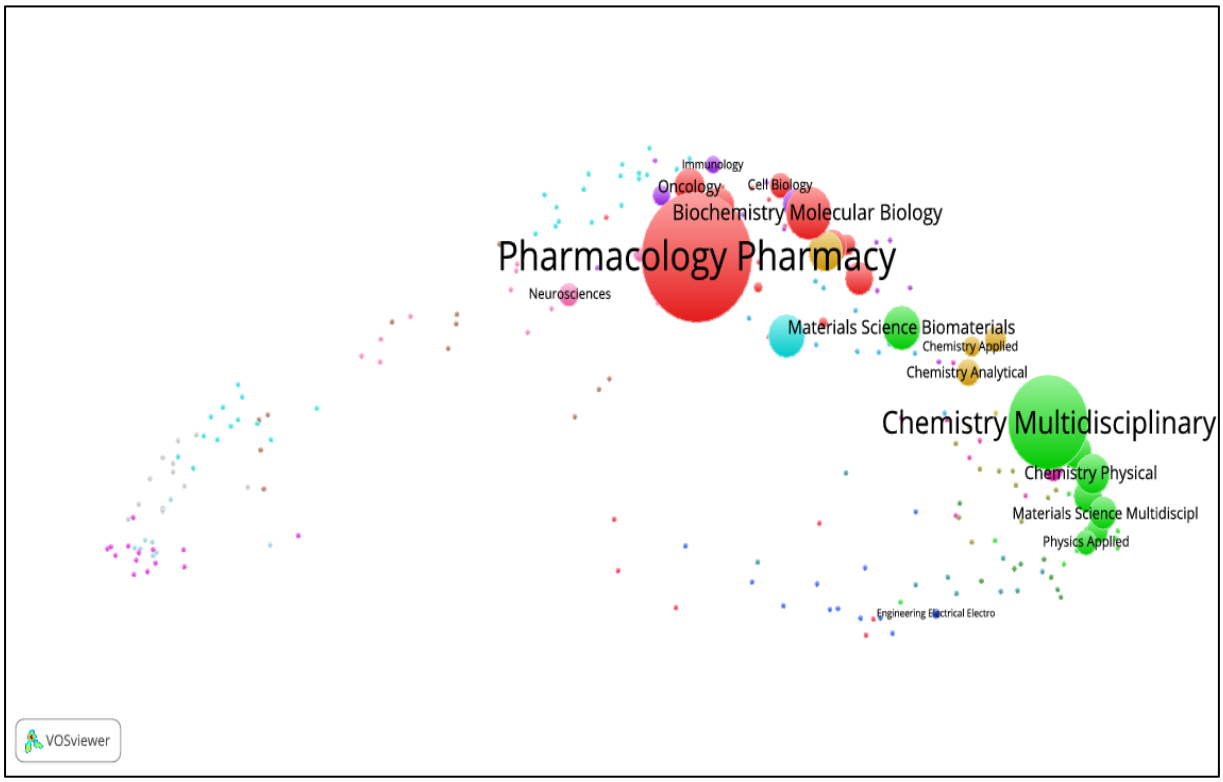


Figure 55. WoS SCs overlay map of PM3's publications

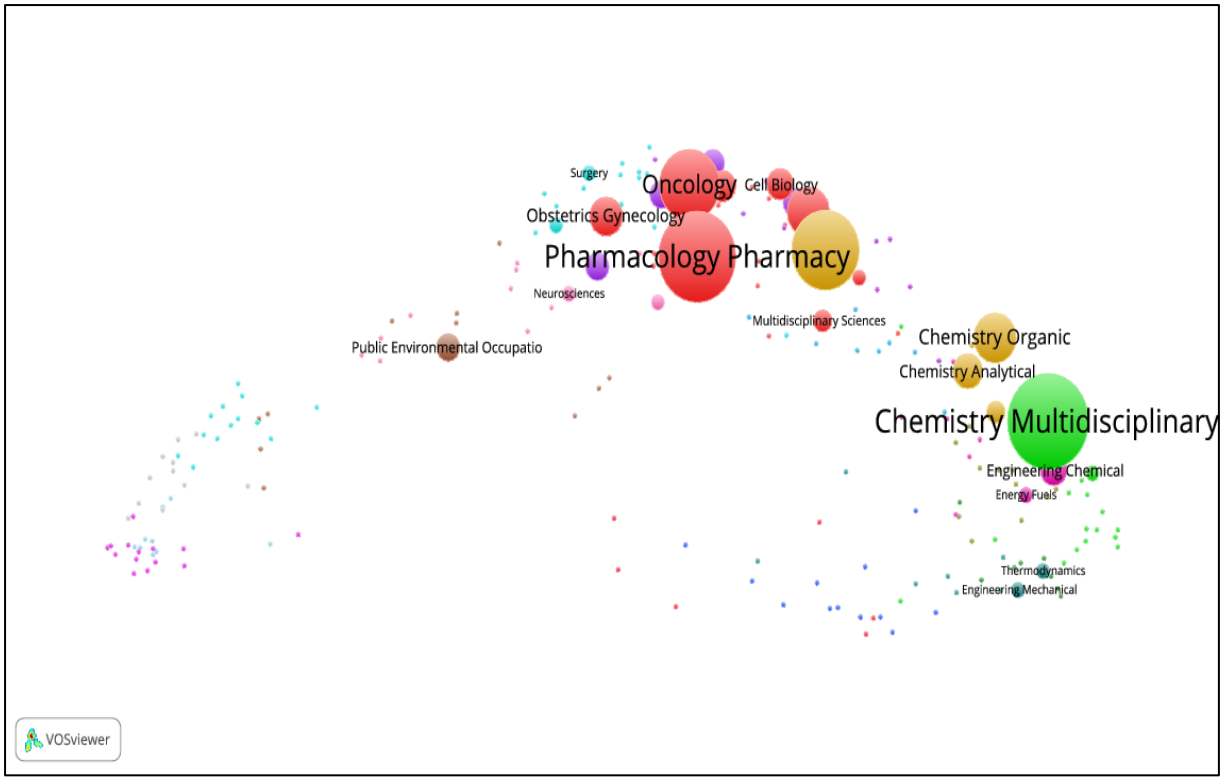


Figure 56. WoS SCs overlay map of PM4's publications

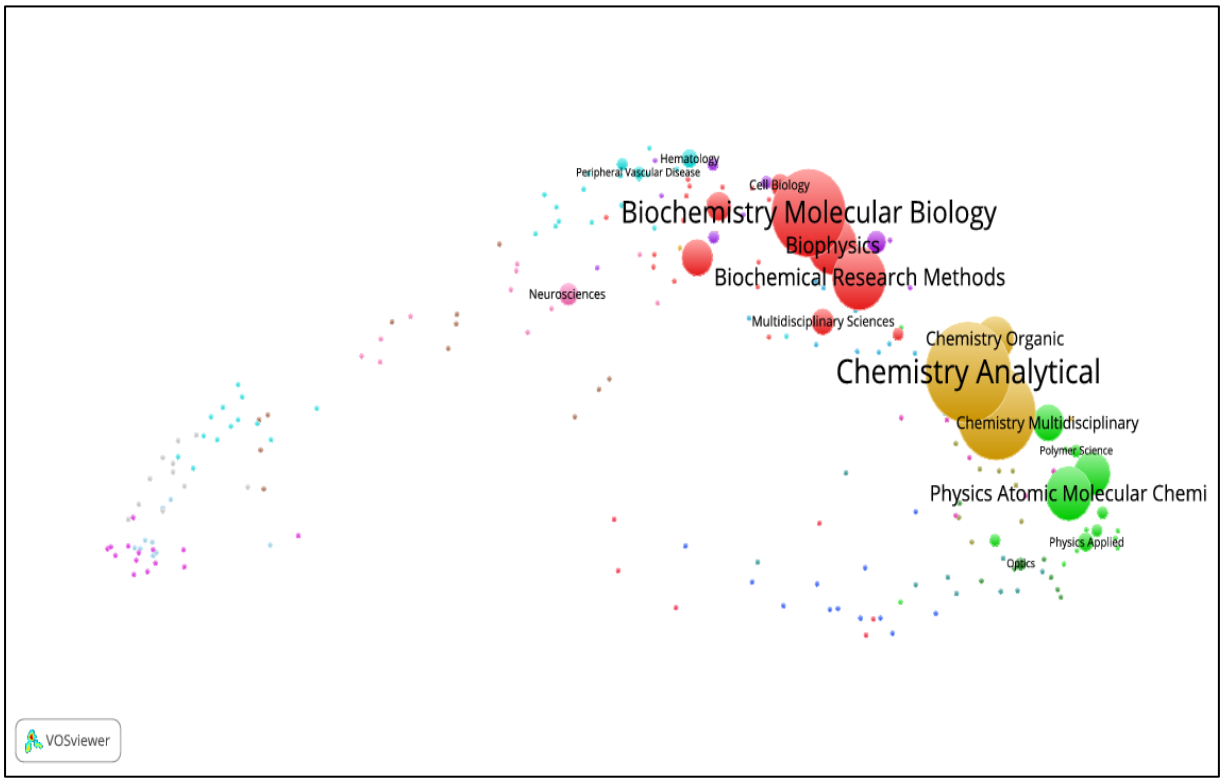


Figure 57. WoS SCs overlay map of PM5's publications

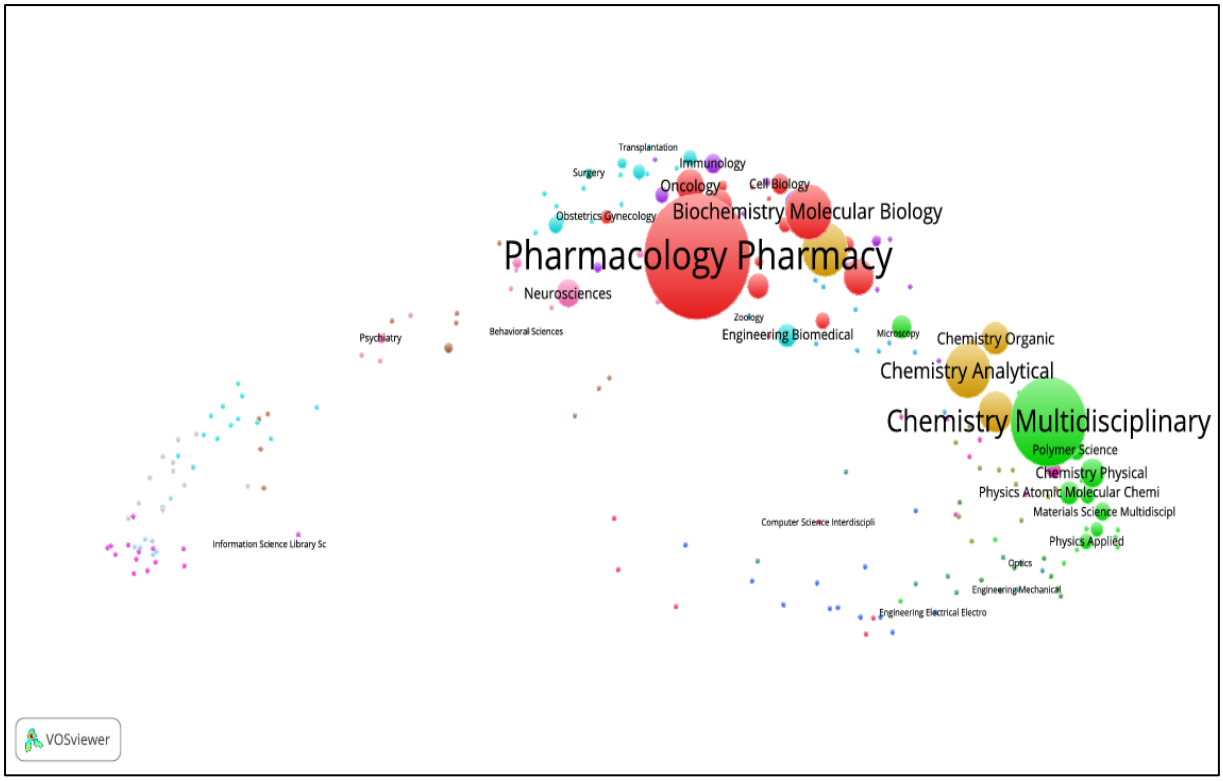


Figure 58. WoS SCs overlay map of panel's publications

Appendix B

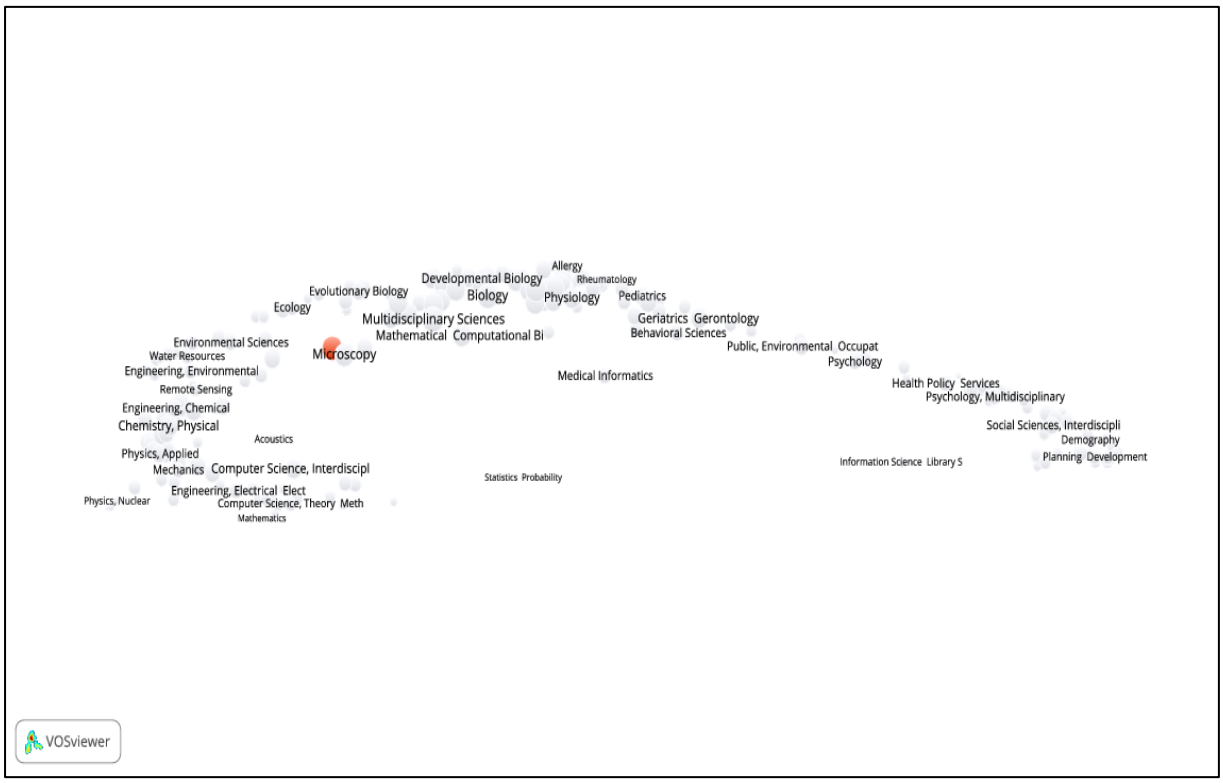


Figure 59. SAPV of the PHAR-A research group's publications in WoS SCs similarity matrix

Figure 60. SAPV of the PHAR-B research group's publications in WoS SCs similarity matrix

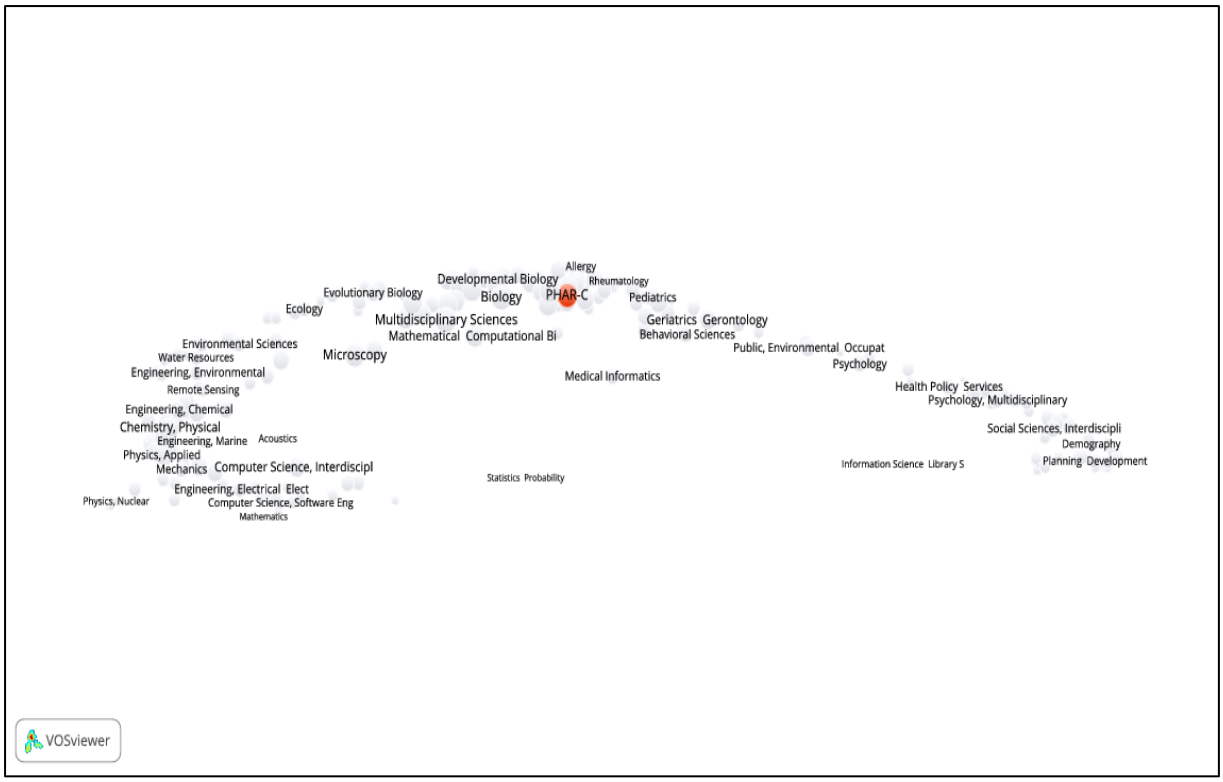


Figure 61. SAPV of the PHAR-C research group's publications in WoS SCs similarity matrix

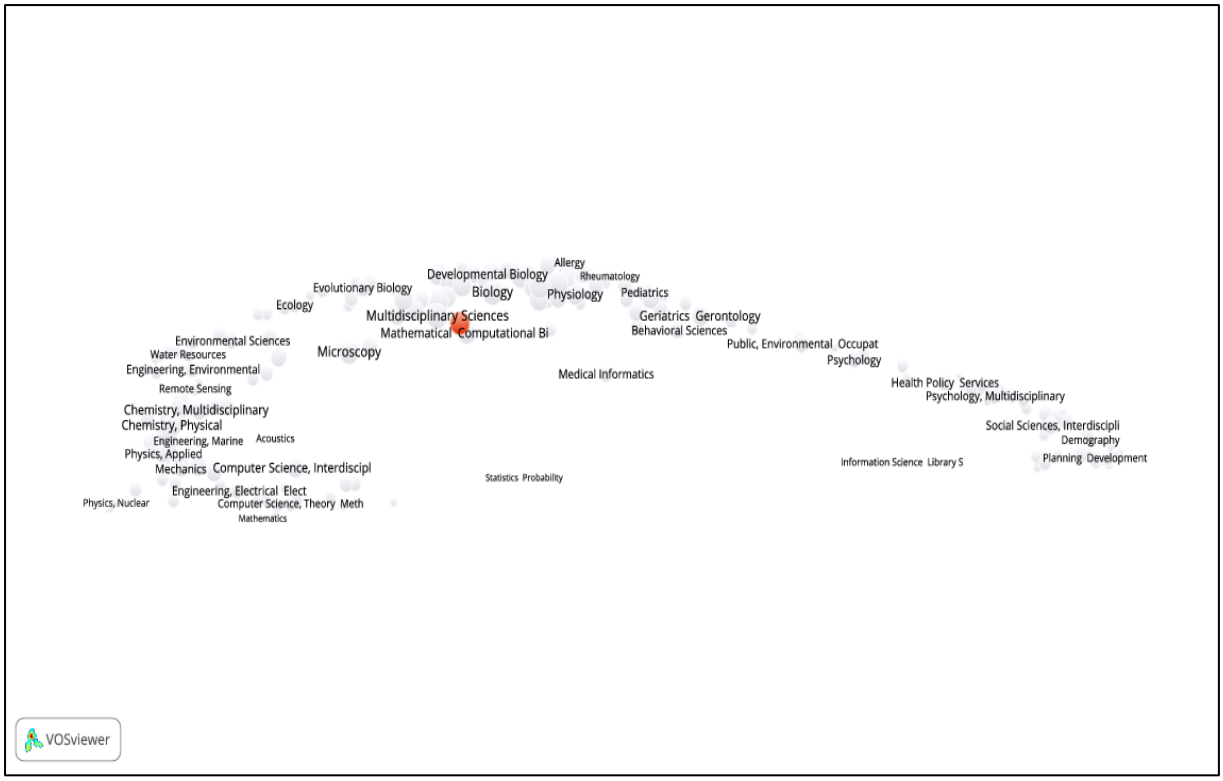


Figure 62. SAPV of the PHAR-D research group's publications in WoS SCs similarity matrix

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

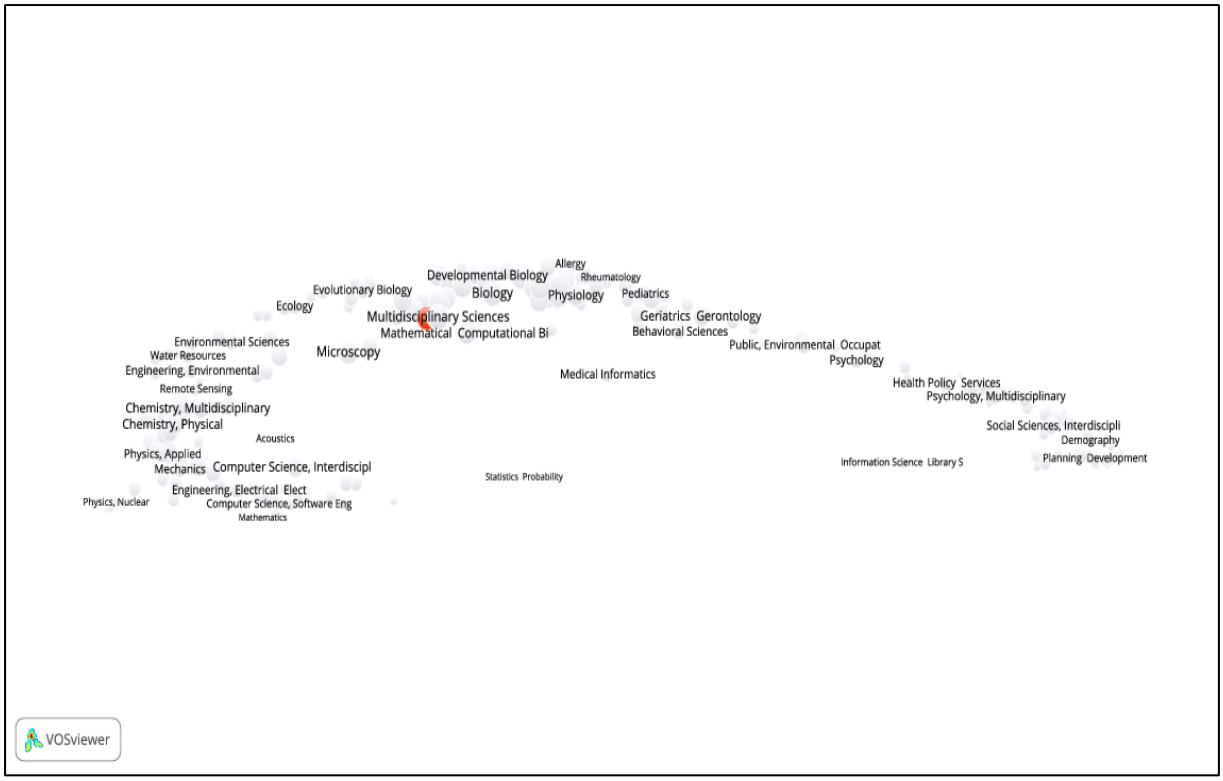


Figure 64. SAPV of the PHAR-F research group's publications in WoS SCs similarity matrix

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

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

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

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

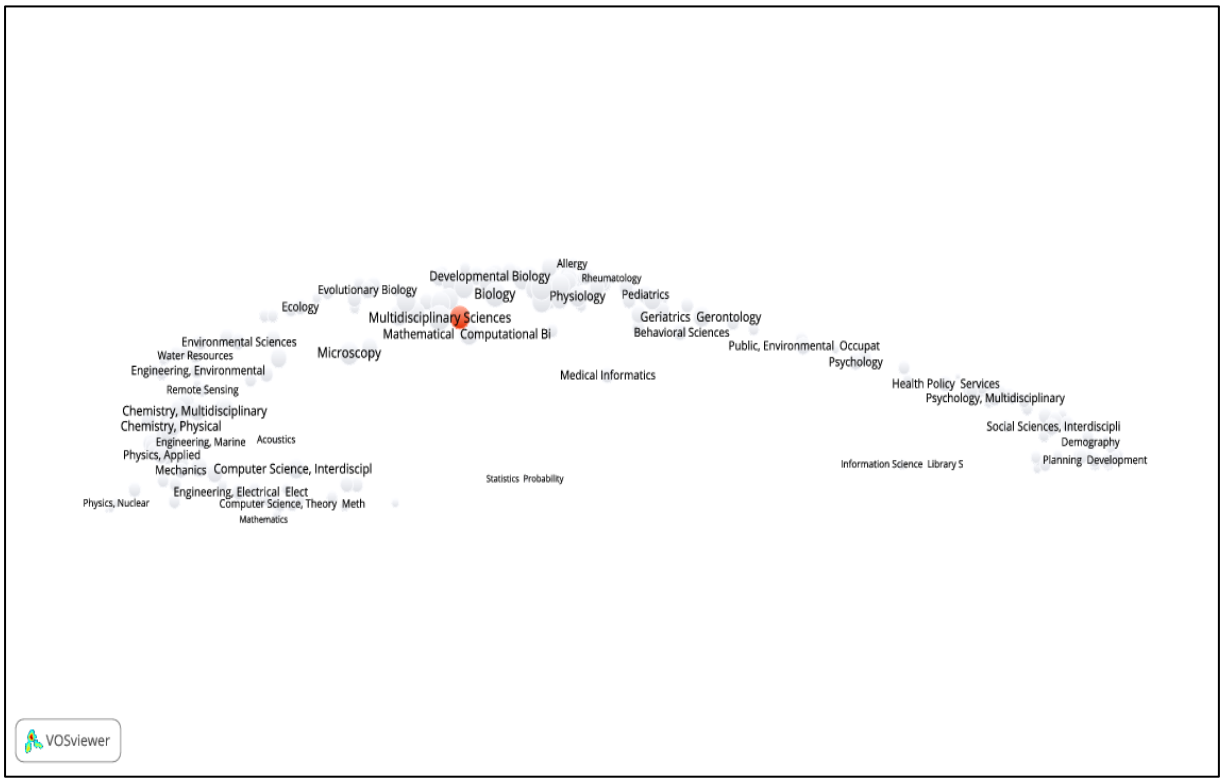


Figure 69. SAPV of the Pharmacy research group's publications in WoS SCs similarity matrix

Figure 70. SAPV of the PM1's publications in WoS SCs similarity matrix

Figure 71. SAPV of the PM2's publications in WoS SCs similarity matrix

Figure 72. SAPV of the PM3's publications in WoS SCs similarity matrix

Figure 73. SAPV of the PM4's publications in WoS SCs similarity matrix

Figure 74. SAPV of the PM5's publications in WoS SCs similarity matrix

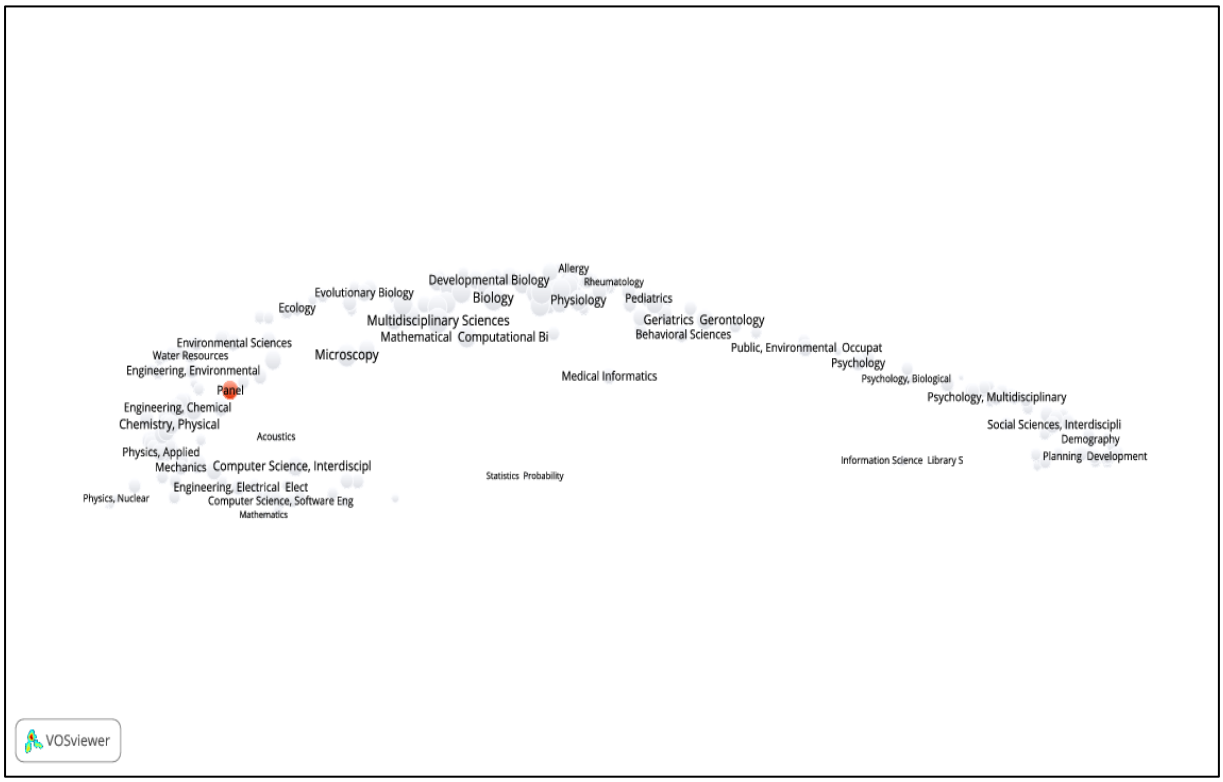


Figure 75. SAPV of the panel publications in WoS SCs similarity matrix

Appendix C

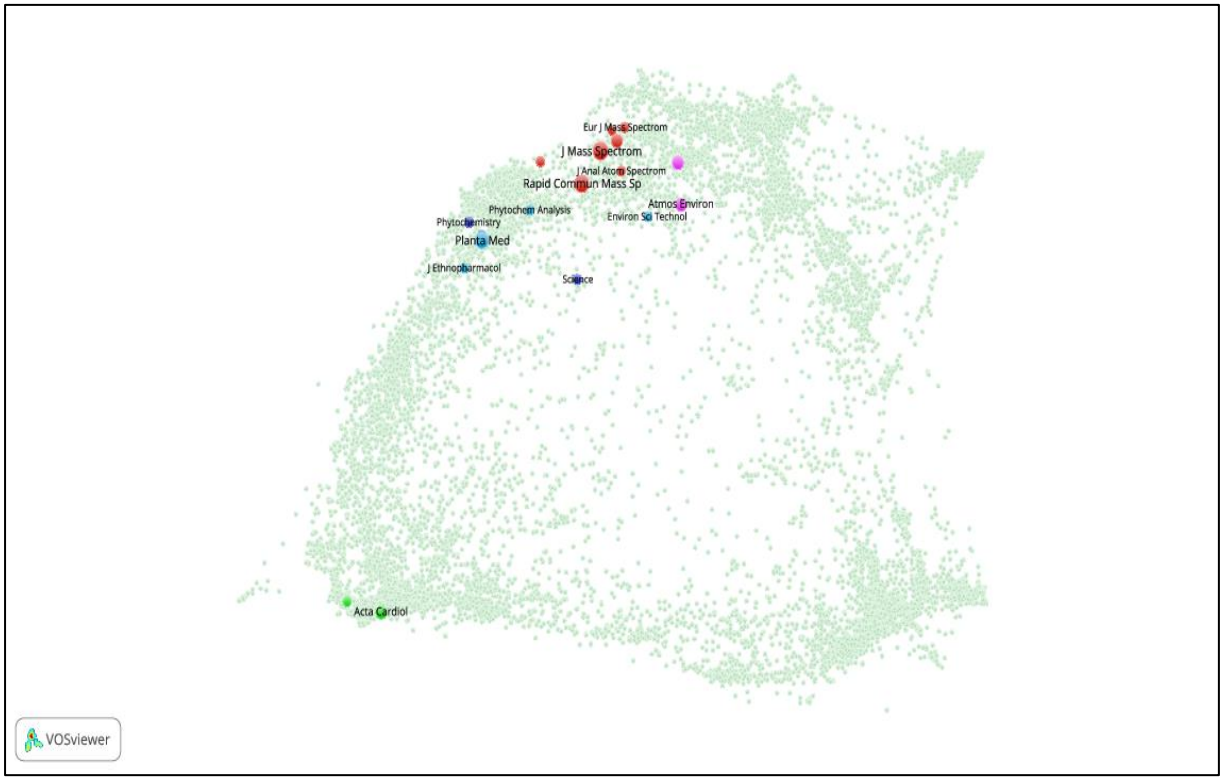


Figure 76. Journal overlay map of PHAR-A research group's publications

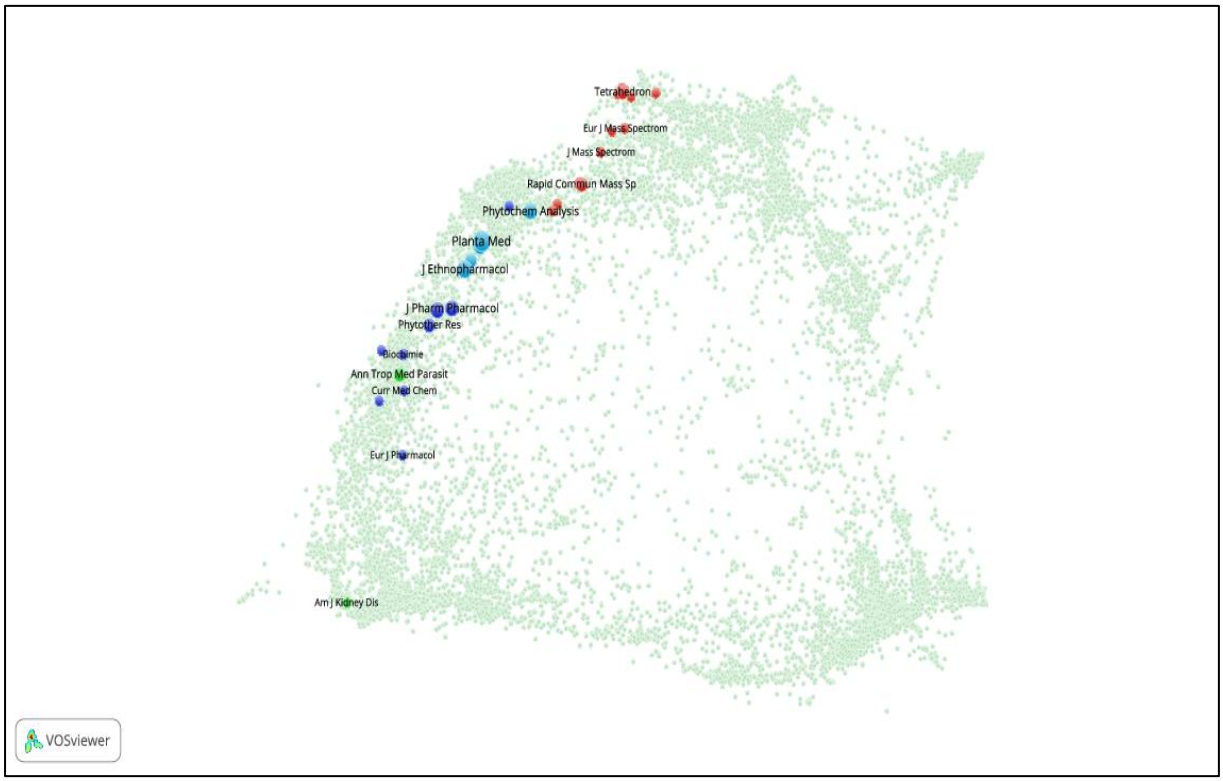


Figure 77. Journal overlay map of PHAR-B research group's publications

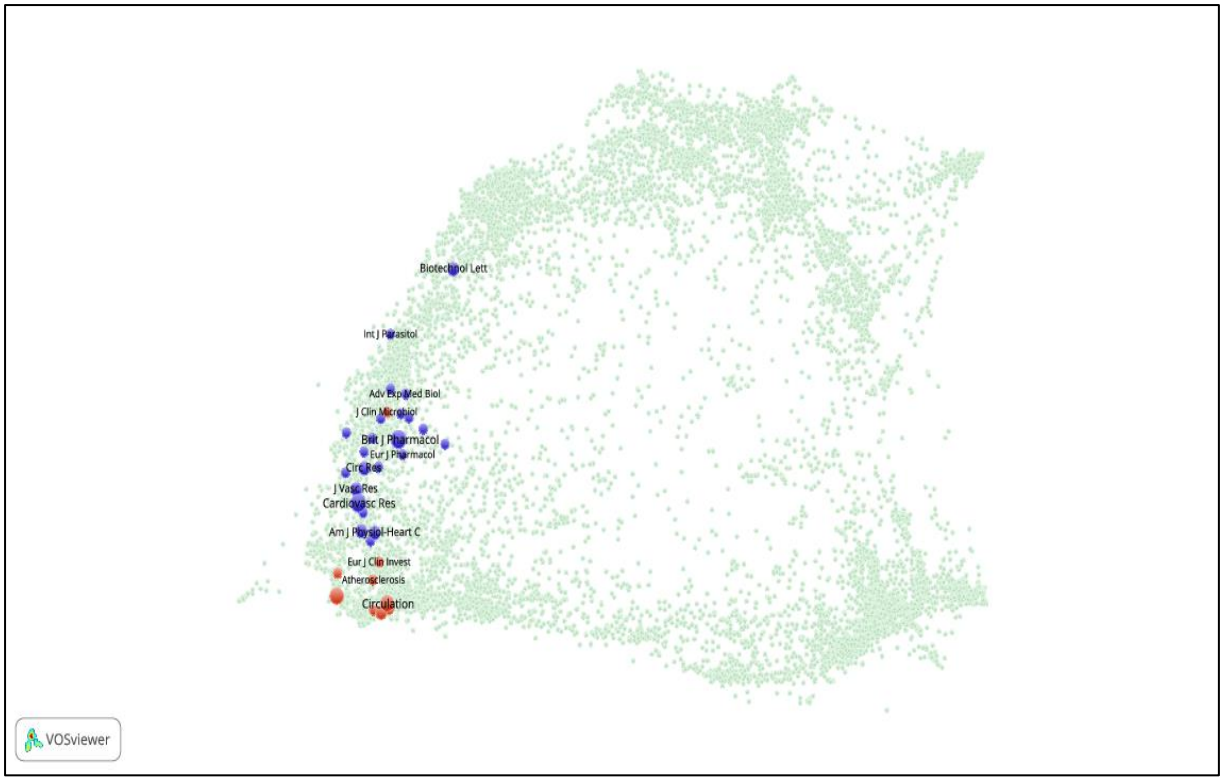


Figure 78. Journal overlay map of PHAR-C research group's publications

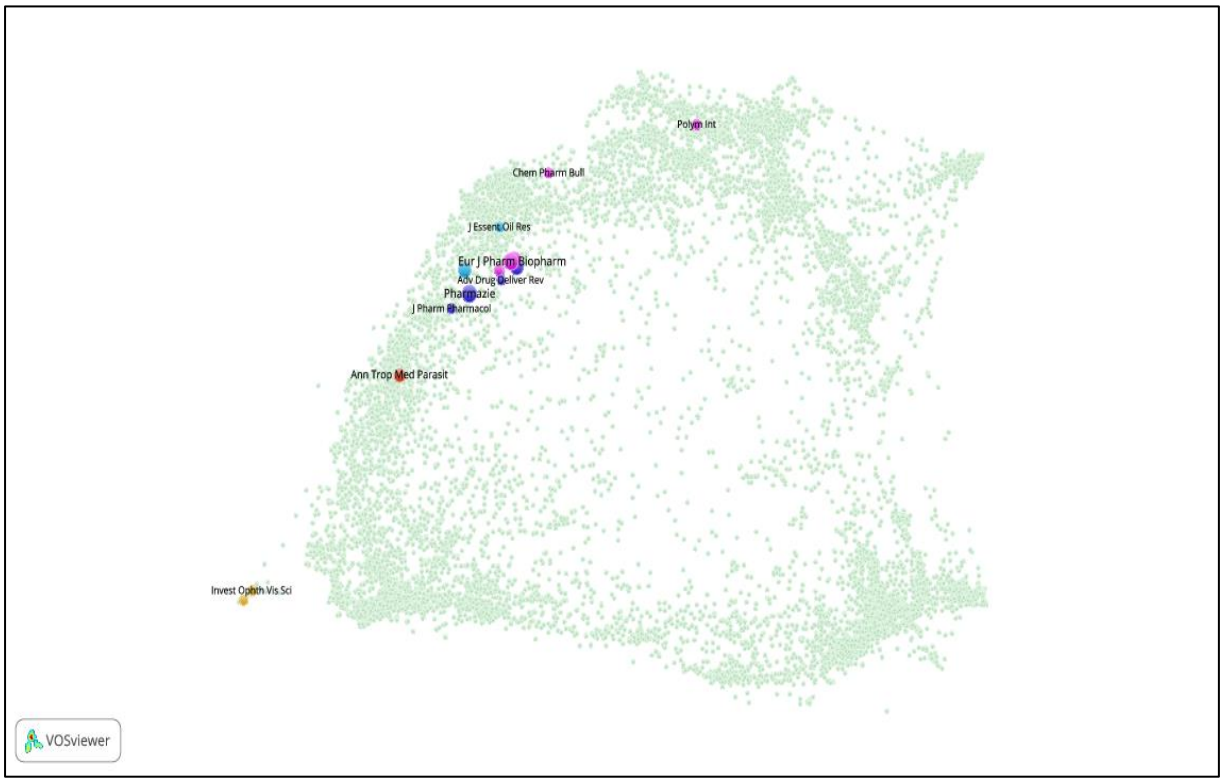


Figure 79. Journal overlay map of PHAR-D research group's publications

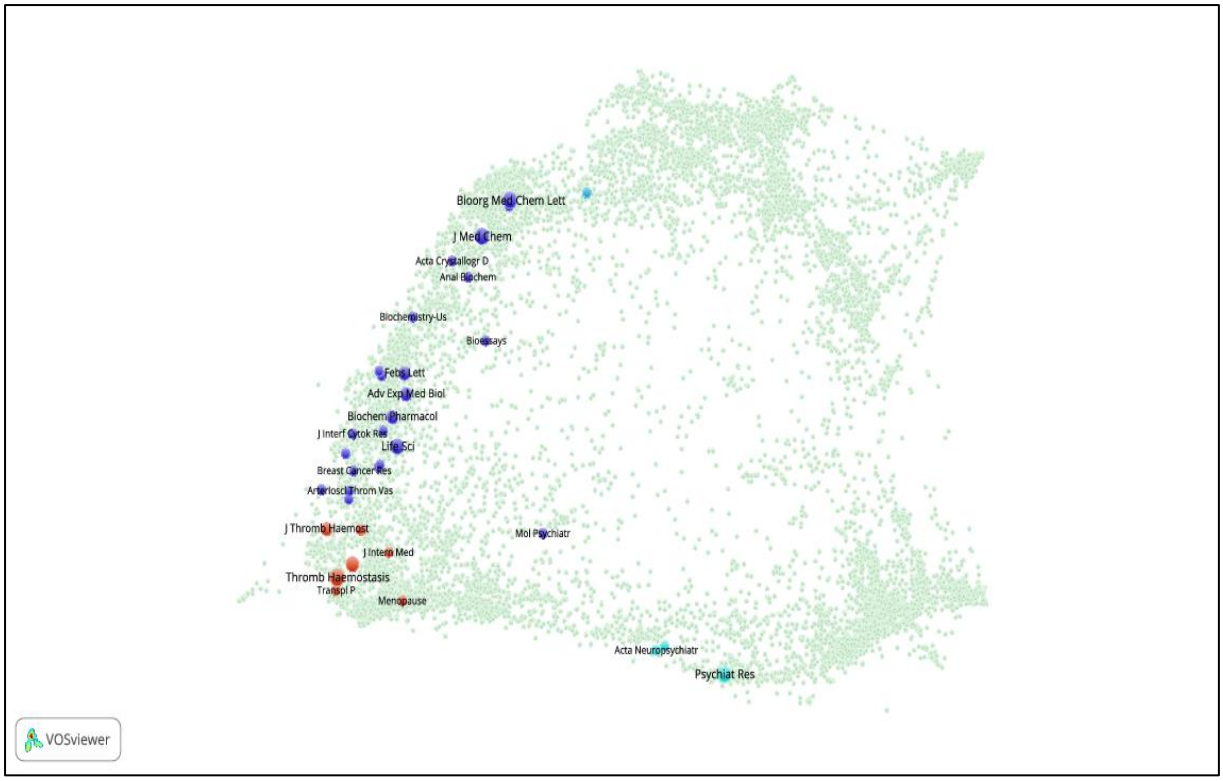


Figure 80. Journal overlay map of PHAR-E research group's publications

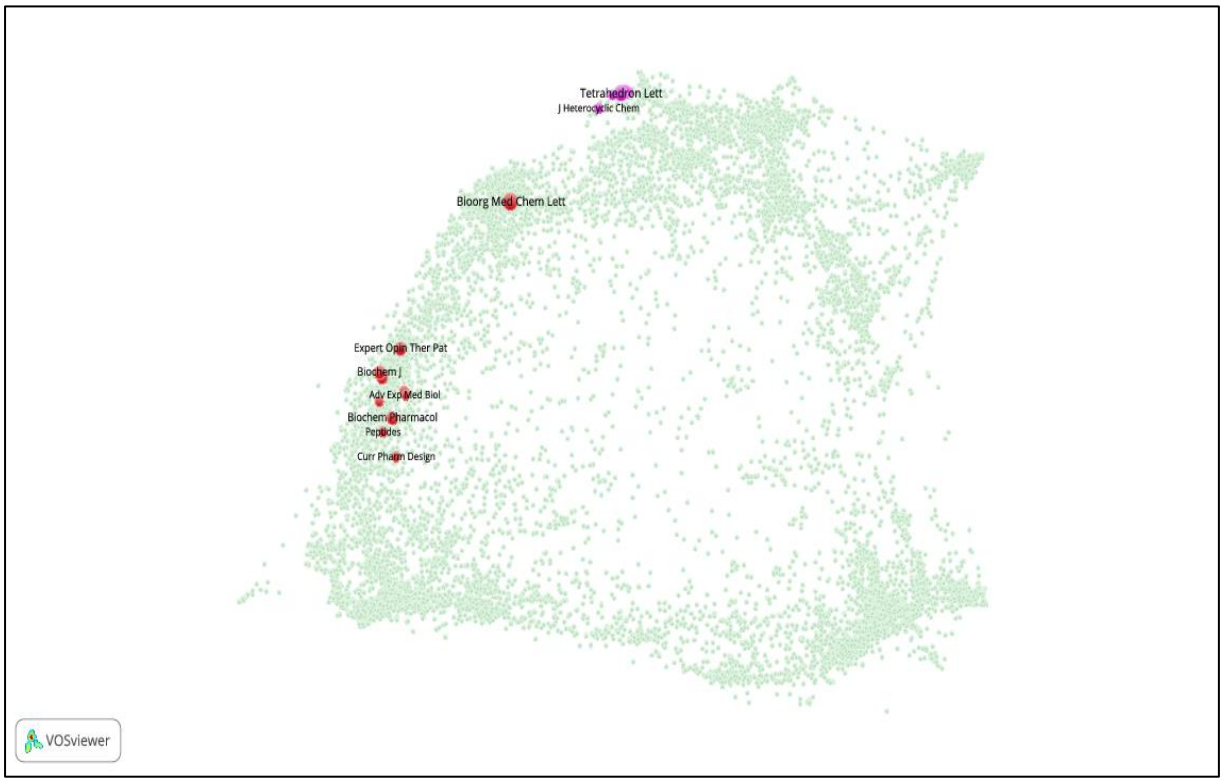


Figure 81. Journal overlay map of PHAR-F research group's publications

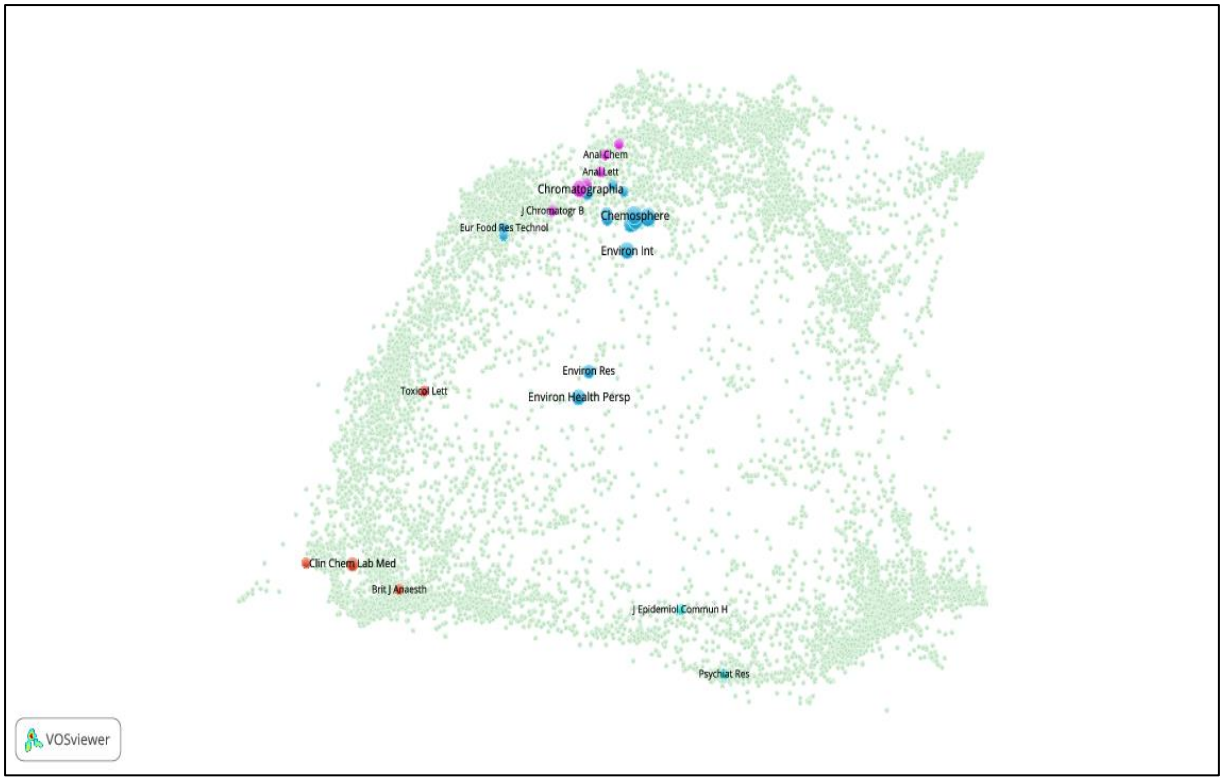


Figure 82. Journal overlay map of PHAR-G research group's publications

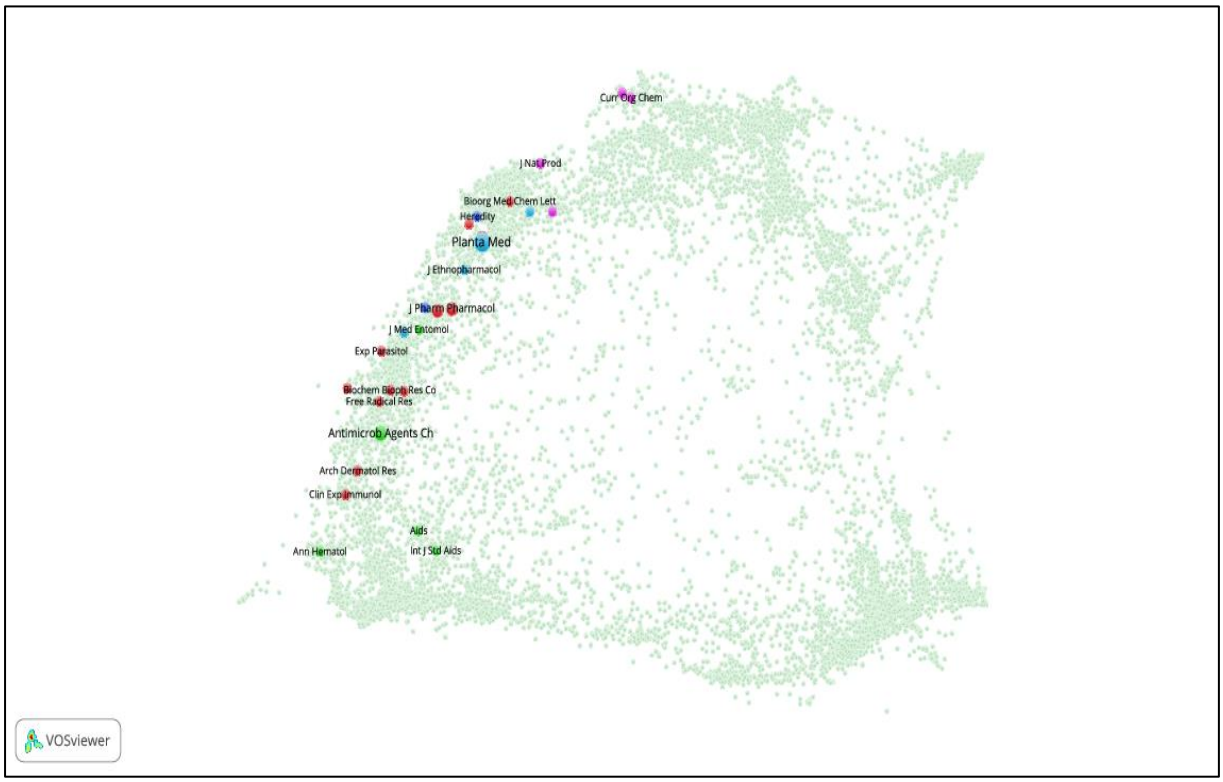


Figure 83. Journal overlay map of PHAR-H research group's publications

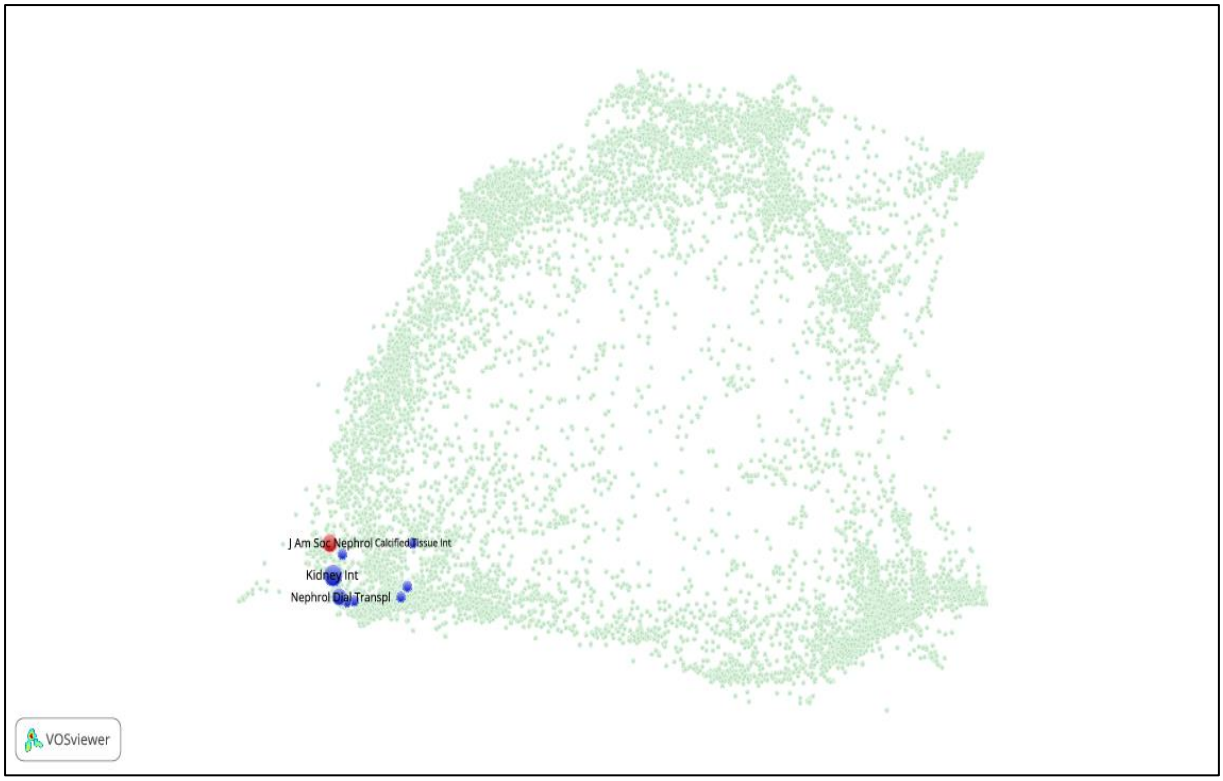


Figure 84. Journal overlay map of PHAR-I research group's publications

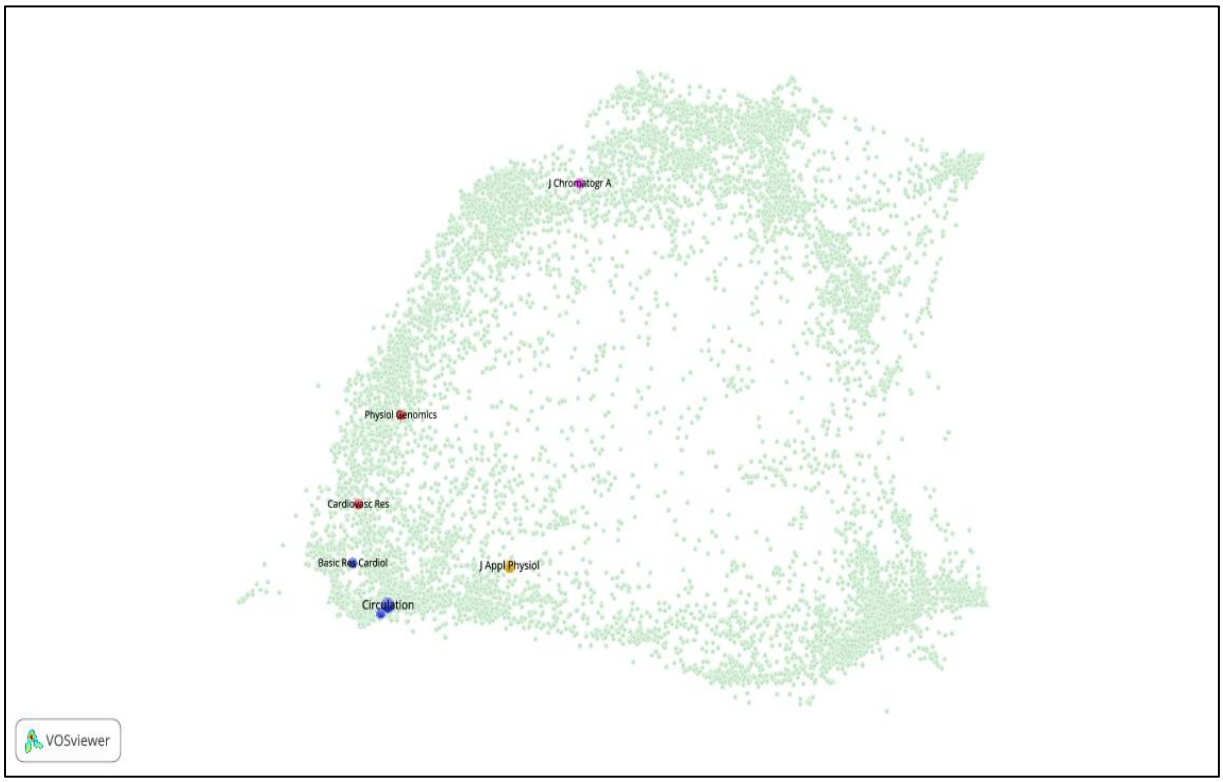


Figure 85. Journal overlay map of PHAR-j research group's publications

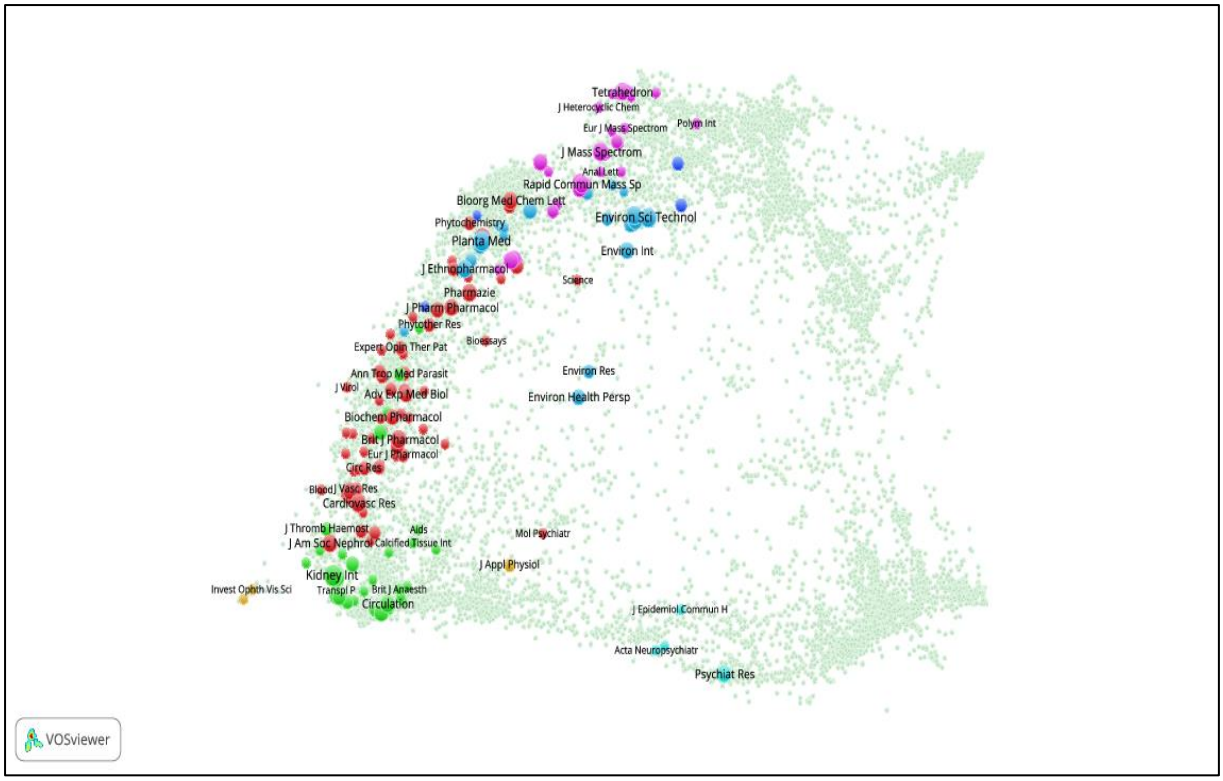


Figure 86. Journal overlay map of Pharmacy research groups' publications

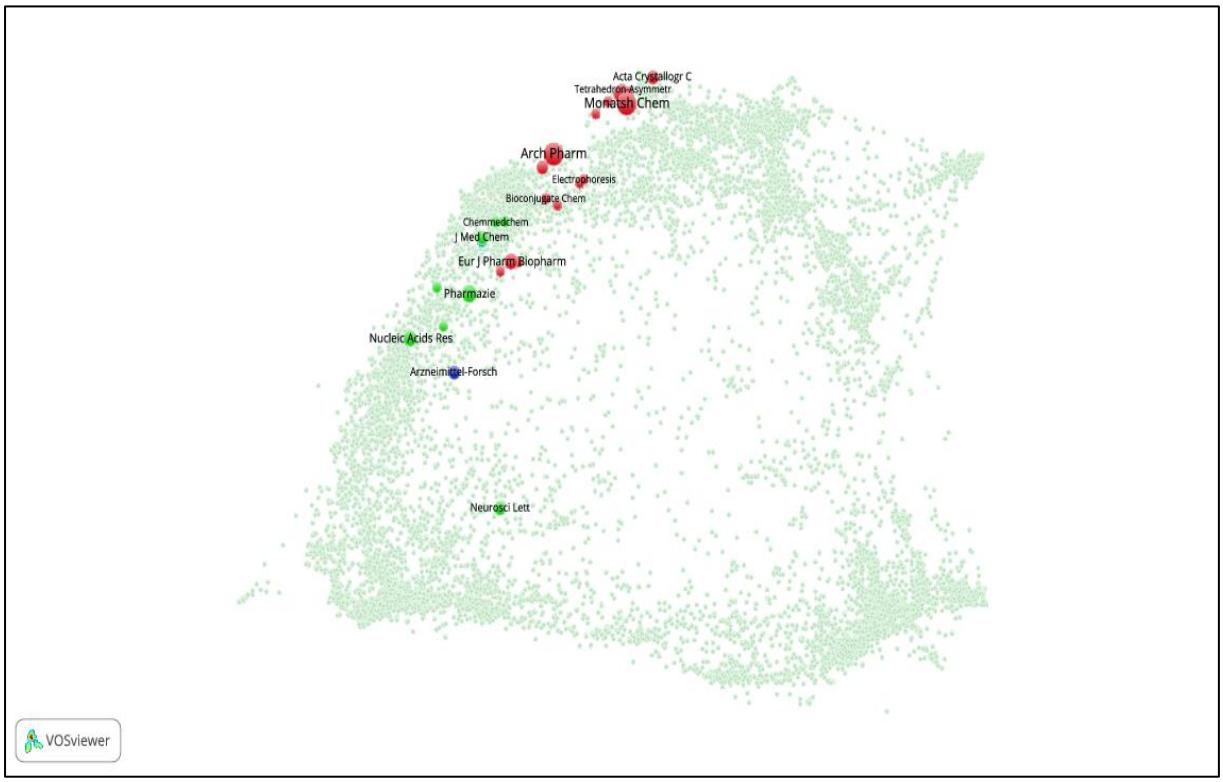


Figure 87. Journal overlay map of PM1's publications

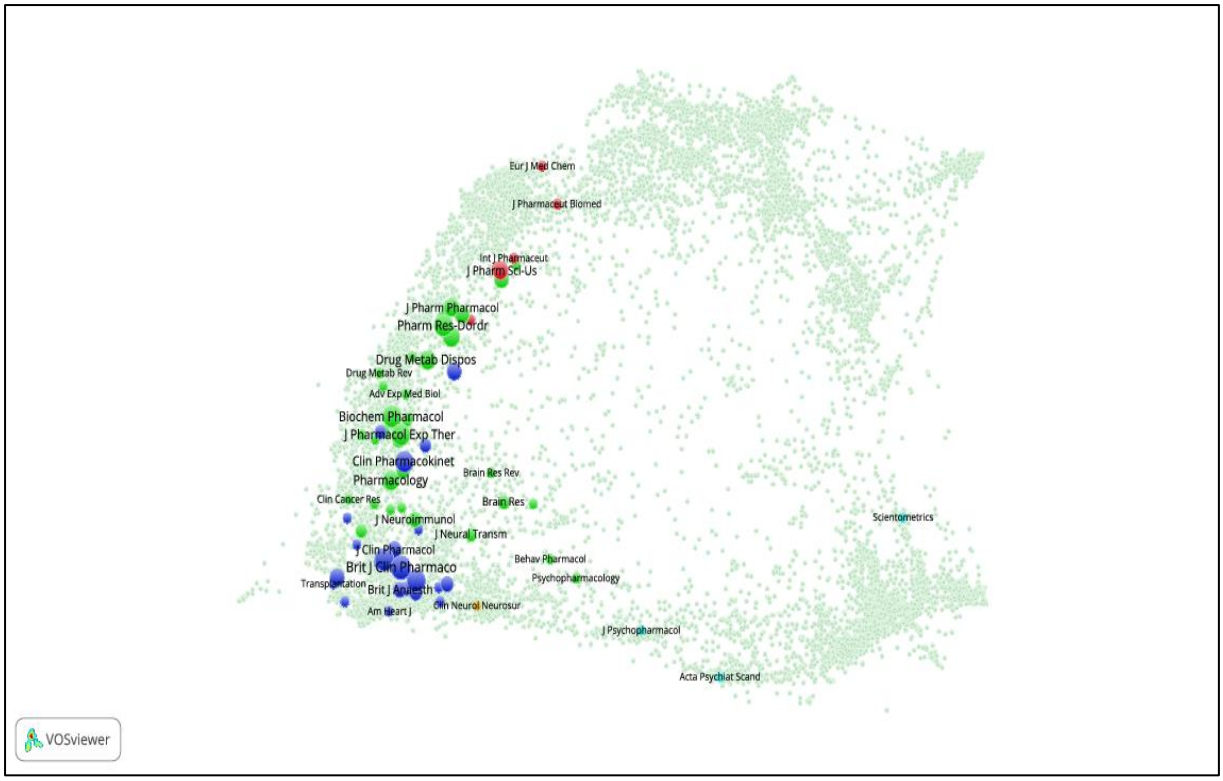


Figure 88.. Journal overlay map of PM2's publications

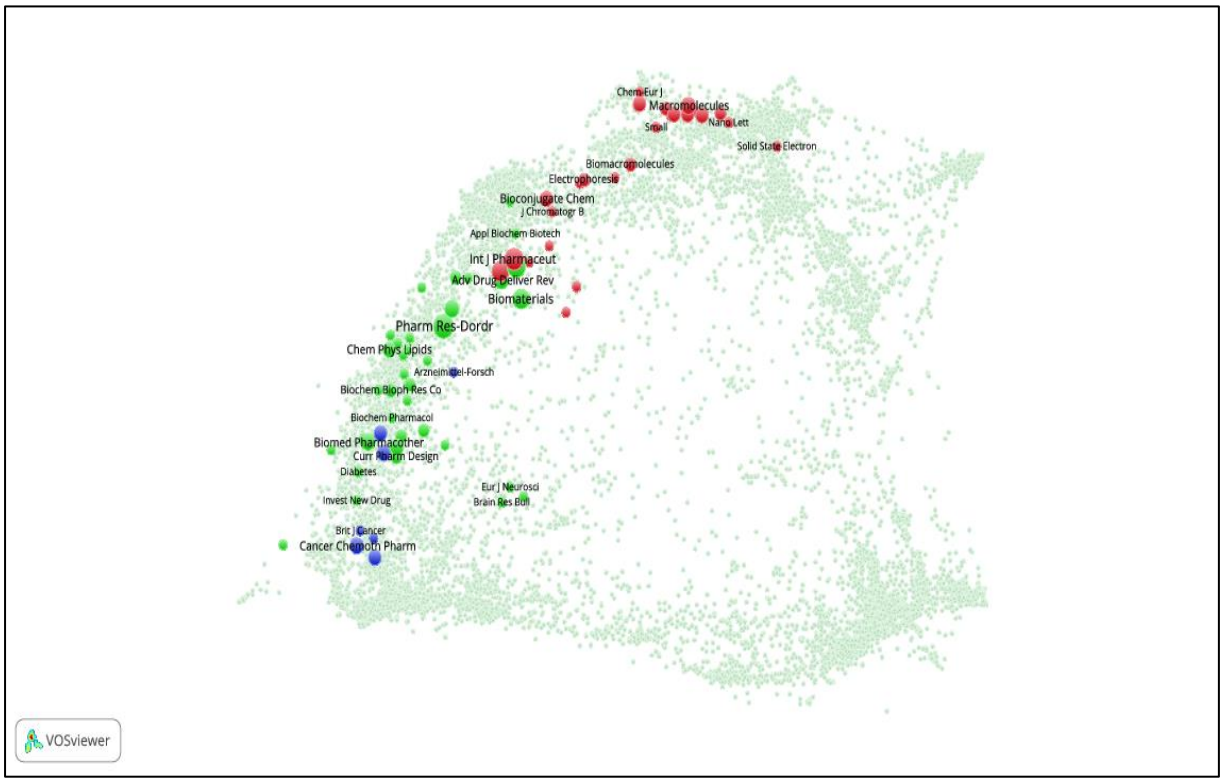


Figure 89. Journal overlay map of PM3's publications

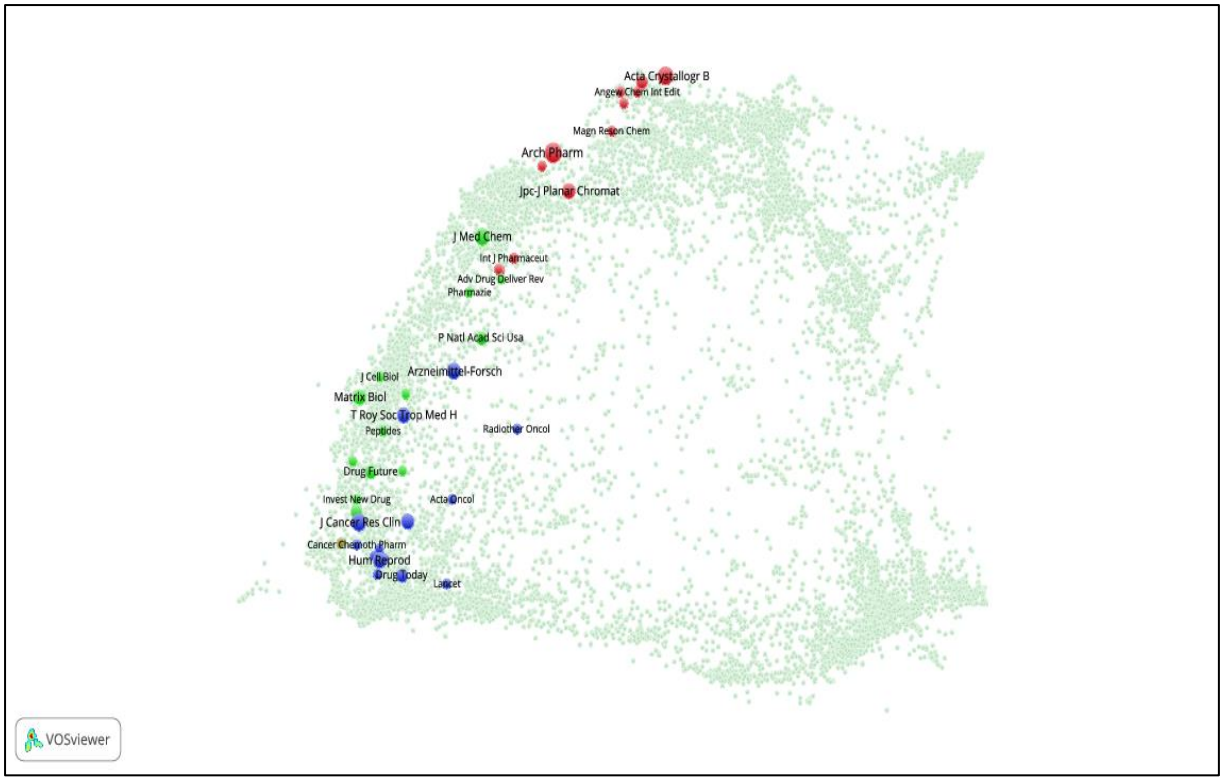


Figure 90. Journal overlay map of PM4's publications

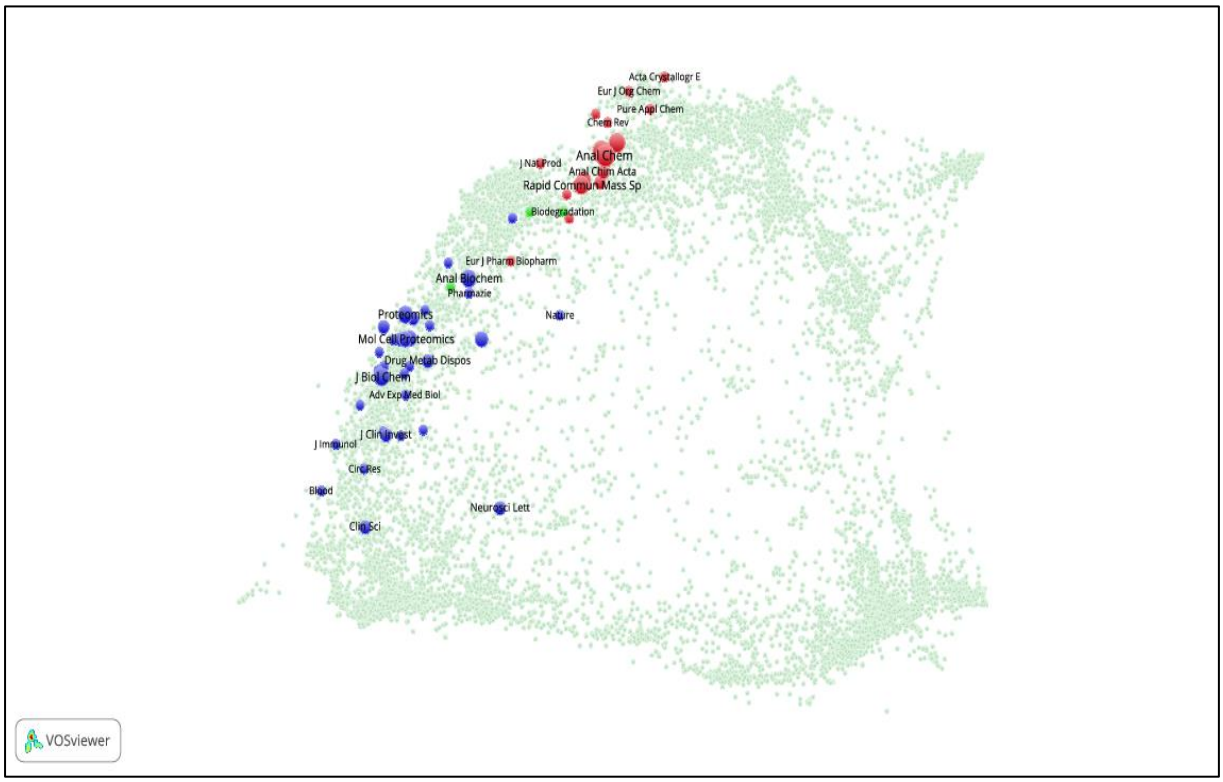


Figure 91. Journal overlay map of PM5's publications

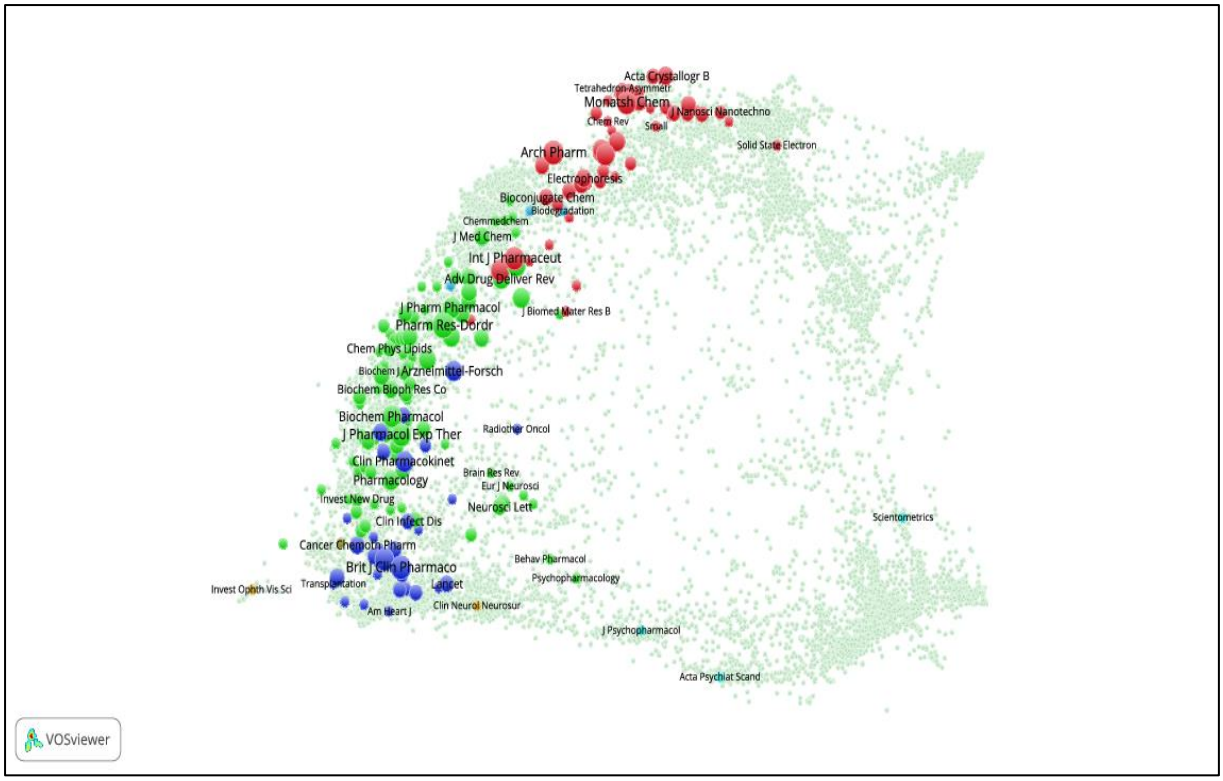


Figure 92. Journal overlay map of the panel's publications