



**Determining cognitive distance between publication portfolios of
evaluators and evaluatees in research evaluation: A case study of
Veterinary 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 evaluatees in research evaluation: Exploration of informetric methods*. Similar technical reports on Biology, Biomedical Science, Chemistry Pharmaceutical Sciences and Physics department are also available at the institutional repository of the University of Antwerp (<https://repository.uantwerpen.be>).

Contents

List of Tables	v
List of Figures.....	vi
1 Introduction.....	1
2 Cognitive distance based on Web of Science subject categories.....	3
2.1 Data collection process	3
a) Research groups data retrieval	3
b) Panel members data retrieval.....	4
2.2 Correlation between publication profiles of research groups together and panel.....	6
a) Pearson's correlation coefficient and Spearman's rank-order correlation coefficient ..	6
b) Top-Down correlation coefficient	8
2.3 Web of Science subject categories similarity matrix	9
2.4 Web of Science subject categories overlay map creation	13
2.5 Bootstrapping and confidence intervals.....	14
2.6 Barycenter method	15
a) Barycenter calculation	15
b) Euclidean distance between barycenters	16
c) Barycenter overlay map	17
2.7 Similarity-adapted publication vector method.....	20
a) Similarity-adapted publication vector calculation	20
b) Euclidean distance between similarity-adapted publication vectors	22
c) Similarity-adapted publication vector overlay map.....	23

2.8 Weighted cosine similarity method	25
3 Cognitive distance based on journals.....	28
3.1 Data collection process	28
3.2 Correlation between publication profiles of research groups together and panel.....	29
a) Pearson's correlation coefficient and Spearman's rank-order correlation coefficient	29
b) Top-Down correlation coefficient	31
3.3 Journal similarity matrix	32
3.4 Journal overlay map creation	34
3.5 Barycenter method	35
a) Barycenter calculation	35
b) Euclidean distance between barycenters	38
c) Barycenter overlay map	39
3.6 Similarity-adapted publication vector method.....	41
a) Similarity-adapted publication vector calculation	41
b) Euclidean distance between similarity-adapted publication vectors	42
3.7 Weighted cosine similarity method	44
4 Heat map.....	46
5 Programming code in Python	49
References	53
Appendix A	55
Appendix B	59
Appendix C	64

List of Tables

Table 1. Publication statistics of Veterinary Sciences research groups (2006-2013)	3
Table 2. Publication statistics of Veterinary Sciences panel members	4
Table 3. Euclidean distances between barycenter of Veterinary Sciences individual research groups, panel members, research groups together and panel using WoS SCs VOS map	17
Table 4. Euclidean distances between SAPVs of Veterinary Sciences individual groups, panel members, research groups together and panel in WoS SCs similarity matrix	23
Table 5. WCS value of Veterinary Sciences individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix.....	26
Table 6. WCD value between Veterinary Sciences individual research groups, panel members, groups and panel using WoS SCs similarity matrix.....	27
Table 7. Pearson and Spearman correlation between three methods using data from Veterinary Sciences individual research groups and panel members.....	27
Table 8. Euclidean distances between Veterinary Sciences individual research groups, panel members, research groups together and panel using the journal VOS map.....	39
Table 9. Euclidean distances between SAPV of Veterinary Sciences individual research groups, panel members, research groups together and panel using the journal similarity matrix	43
Table 10. WCS value of the Veterinary Sciences groups, panel members, panel and research groups together using the journal similarity matrix.	45
Table 11. WCD value of the Veterinary Sciences groups, panel members, panel and research groups together using the journal similarity matrix	46

List of Figures

Figure 1. Excerpt of Veterinary Sciences research groups and panel members_WoS SCs.xlsx file	5
Figure 2. Excerpt of the Veterinary Sciences panel and research groups together_WoS SCs.xlsx file..	6
Figure 3. Excerpt of the Veterinary Sciences panel and groups together_WoS SCs - joined.xlsx file...	7
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 Veterinary Sciences department.....	7
Figure 5. Excerpt of the map10.paj file	9
Figure 6. Transformation of WoS SCs similarity matrix to Kamada-Kawai map and VOS map	10
Figure 7. VOSviewer message before choosing Kamada Kawai map or VOS map data.....	11
Figure 8. Excerpt of WoS SCs Kamada-Kawai map data from map10.paj file.....	11
Figure 9. Excerpt of WoS SCs VOS map data from the map10.paj file	12
Figure 10. Excerpt of WoS SCs VOS map	12
Figure 11. VETE-B research group's publication overlay map in WoS SCs.	13
Figure 12. Barycenter coordinates of the Veterinary Sciences individual research groups, panel members, research groups together and panel using the WoS SCs VOS map.....	16
Figure 13. Excerpt of Euclidean distances matrix of barycenter of the Veterinary Sciences individual research groups, panel members, research groups together and panel using WoS SCs VOS map	17
Figure 14. Barycenter overlay map of Veterinary Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs	18
Figure 15. Barycenter map of Veterinary Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed)	19
Figure 16. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together (groups) with their confidence regions.....	19
Figure 17. Excerpt of WoS SCs similarity matrix	21
Figure 18. Excerpt of SAPV of the Veterinary Sciences individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix.....	21
Figure 19. Excerpt of pairwise Euclidean distance matrix between SAPVs of Veterinary Sciences individual research groups, panel members, research groups together and panel together using WoS SCs similarity matrix	22

Figure 20. Excerpt of VETE-B.csv file.....	24
Figure 21. Location of the SAPV of VETE-B in the WoS SCs similarity matrix	24
Figure 22. Excerpt of WCS value matrix of the Veterinary Sciences individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix	26
Figure 23. Scatter plot of the correlation between barycenter, SAPV and WCS.....	27
Figure 24. Excerpt of Veterinary Sciences research groups and panel_journal title.xlsx file	29
Figure 25. Excerpt of the Veterinary Sciences panel and research groups together_journals title.xlsx file	30
Figure 26. Excerpt of the Veterinary Sciences panel and research groups together_journals title-joined.xlsx file	31
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 Veterinary Sciences department.....	31
Figure 28. Excerpt of the journals VOS map data	33
Figure 29. Journal overlay map of the VETE-B research group.....	34
Figure 30. Excerpt of short form to full journal titles	35
Figure 31. Excerpt of Journal name change.xlsx file.....	37
Figure 32. Barycenter coordinates of the Veterinary Sciences individual research groups, panel members, research groups together, and panel using journal VOS map	37
Figure 33. Excerpt of Euclidean distances matrix of the barycenter of the Veterinary Sciences groups, panel members, research groups together and panel using the journal VOS map	38
Figure 34. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together	39
Figure 35. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together (zoomed).....	40
Figure 36. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together with their confidence regions	40
Figure 37. Excerpt of SAPV of the Veterinary Sciences research groups, research groups together, panel members and panel using journal similarity matrix.....	42
Figure 38. Excerpt of pairwise Euclidean distances matrix between the SAPV of the Veterinary Sciences individual research groups, panel members, panel and research groups together using the journal similarity matrix	43

Figure 39. Excerpt of WCS value matrix of the Veterinary Sciences individual research groups, panel members, groups and panel using the journal similarity matrix.....	45
Figure 40. Excerpt of the dissimilarities/distances between panel members and individual research groups according to each of the six approaches	47
Figure 41. Heat map with hierarchical clustering based on correlation coefficient between six methods in the Veterinary sciences department.....	48
Figure 42. WoS SCs overlay map of VETE-A research group's publications.....	55
Figure 43. WoS SCs overlay map of VETE-B research group's publications	55
Figure 44. WoS SCs overlay map of VETE-C research group's publications	56
Figure 45. WoS SCs overlay map of Veterinary research groups' publications.....	56
Figure 46. WoS SCs overlay map of PM1's publications	57
Figure 47. WoS SCs overlay map of PM2's publications	57
Figure 48. WoS SCs overlay map of PM3's publications	58
Figure 49. WoS SCs overlay map of PM4's publications	58
Figure 50. WoS SCs overlay map of panel's publications	59
Figure 51. SAPV of the VETE-A research group's publications in WoS SCs similarity matrix	59
Figure 52. SAPV of the VETE-B research group's publications in WoS SCs similarity matrix	60
Figure 53. SAPV of the VETE-C research group's publications in WoS SCs similarity matrix	60
Figure 54. SAPV of the Veterinary research group's publications in WoS SCs similarity matrix	61
Figure 55. SAPV of the PM1's publications in WoS SCs similarity matrix	61
Figure 56. SAPV of the PM2's publications in WoS SCs similarity matrix	62
Figure 57. SAPV of the PM3's publications in WoS SCs similarity matrix	62
Figure 58. SAPV of the PM4's publications in WoS SCs similarity matrix	63
Figure 59. SAPV of the panel publications in WoS SCs similarity matrix	63
Figure 60. Journal overlay map of VETE-A research group's publications.....	64
Figure 61. Journal overlay map of VETE-B research group's publications.....	64
Figure 62. Journal overlay map of VETE-C research group's publications.....	65

Figure 63. Journal overlay map of Veterinary research groups' publications	65
Figure 64. Journal overlay map of PM1's publications	66
Figure 65. Journal overlay map of PM2's publications	66
Figure 66. Journal overlay map of PM3's publications	67
Figure 67. Journal overlay map of PM4's publications	67
Figure 68. Journal overlay map of the panel's publications	68

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., Veterinary 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 methodology 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 Veterinary Sciences and the respective expert panels based on research evaluations carried out at the University of Antwerp.

In this technical report, we present the Veterinary 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's rank-order correlation coefficient between each pair of the six approaches (section 4) and the programming code for the main methods used respectively (section 5). Finally, we present overlay maps and location of similarity adapted

publication vector of Veterinary 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 2014 assessment of the three research groups of the Department of Veterinary 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 VETE-A, VETE-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 Science Citation Index Expanded (SCIE) and Social Sciences Citation Index (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 ‘VETE-A_WoS SCs.txt’ and keep both files.

Table 1. Publication statistics of Veterinary Sciences research groups (2006-2013)

Group code	Number of Publications	Number of Journals	Number of WoS SCs
VETE-A	143	102	55
VETE-B	41	33	25
VETE-C	52	21	16
All groups	231	146	61

Table 1 lists the publication profile of the Veterinary Sciences research groups during the eight years preceding their evaluation. The Veterinary Sciences research groups generated 231 publications in 146 journals. VETE A and VETE B, VETE-B and VETE-C co-authored one and four publications respectively. In total, their publications are distributed over 61 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 Veterinary Sciences panel was composed of four 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 (2014) 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 Veterinary Sciences panel members

Panel code	Number of Publications	Number of Journals	Number of WoS SCs
PM1	313	50	21
PM2	121	66	31
PM3	272	46	8
PM4	132	53	19
Panel	837	200	55

Table 2 lists the publication profile of the Veterinary Sciences panel members. The combined publication output of the Veterinary Sciences panel members consists of 837 publications, none of which is co-authored publications between panel members. The number of publications per panel member ranges from 121 to 313. In total, these publications appeared in 200 different journals and are assigned to 55 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 231
2	VETERINARY SCIENCES	42	18.182
3	REPRODUCTIVE BIOLOGY	29	12.554
4	CELL BIOLOGY	28	12.121
5	NEUROSCIENCES	27	11.688
6	AGRICULTURE DAIRY ANIMAL SCIENCE	20	8.658
7	GASTROENTEROLOGY HEPATOLOGY	17	7.359
8	BIOCHEMISTRY MOLECULAR BIOLOGY	15	6.494
9	PHYSIOLOGY	14	6.061
10	PHARMACOLOGY PHARMACY	14	6.061
11	ONCOLOGY	11	4.762
12	CLINICAL NEUROLOGY	11	4.762

◀
▶
▶
▶
VETE-A
VETE-B
VETE-C
Groups together
PM1
PM2
PM3
PM4
Panel together

Figure 1. Excerpt of Veterinary 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 (VETE-A, VETE-B etc.), the panel members’ code names, (PM1, PM2 etc.), Panel together and Groups together. The Excel file is saved as ‘Veterinary 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 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 'Veterinary 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.

1	Web of Science Categories	records	% of 231	1	Web of Science Categories	records	% of 837
2	VETERINARY SCIENCES	42	18.182	2	VETERINARY SCIENCES	455	54.361
3	REPRODUCTIVE BIOLOGY	29	12.554	3	REPRODUCTIVE BIOLOGY	202	24.134
4	CELL BIOLOGY	28	12.121	4	AGRICULTURE DAIRY ANIMAL SCIENCE	107	12.784
5	NEUROSCIENCES	27	11.688	5	ANATOMY MORPHOLOGY	67	8.005
6	AGRICULTURE DAIRY ANIMAL SCIENCE	20	8.658	6	IMMUNOLOGY	62	7.407
7	GASTROENTEROLOGY HEPATOLOGY	17	7.359	7	ORTHOPEDICS	34	4.062
8	BIOCHEMISTRY MOLECULAR BIOLOGY	15	6.494	8	DEVELOPMENTAL BIOLOGY	32	3.823
9	PHYSIOLOGY	14	6.061	9	SURGERY	28	3.345
10	PHARMACOLOGY PHARMACY	14	6.061	10	RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING	26	3.106
11	ONCOLOGY	11	4.762	11	ANDROLOGY	22	2.628
12	CLINICAL NEUROLOGY	11	4.762	12	ZOOLOGY	20	2.389
13	CARDIAC CARDIOVASCULAR SYSTEMS	9	3.896	13	GENETICS HEREDITY	19	2.27
14	PATHOLOGY	8	3.463	14	ONCOLOGY	18	2.151
15	ANATOMY MORPHOLOGY	8	3.463	15	CELL BIOLOGY	16	1.912
16	OBSTETRICS GYNECOLOGY	7	3.03	16	BIOCHEMISTRY MOLECULAR BIOLOGY	16	1.912
17	ZOOLOGY	6	2.597	17	PARASITOLOGY	14	1.673
18	MICROBIOLOGY	6	2.597	18	ENGINEERING BIOMEDICAL	8	0.956
19	MEDICINE RESEARCH EXPERIMENTAL	6	2.597	19	ENDOCRINOLOGY METABOLISM	7	0.836
20	ENVIRONMENTAL SCIENCES	6	2.597	20	SPORT SCIENCES	6	0.717
21	ENDOCRINOLOGY METABOLISM	6	2.597	21	RHEUMATOLOGY	6	0.717
22	PERIPHERAL VASCULAR DISEASE	5	2.165	22	OTORHINOLARYNGOLOGY	6	0.717
23	NUTRITION DIETETICS	5	2.165	23	CLINICAL NEUROLOGY	6	0.717
24	MICROSCOPY	5	2.165	24	MULTIDISCIPLINARY SCIENCES	5	0.597
25	IMMUNOLOGY	5	2.165	25	MICROBIOLOGY	5	0.597
26	GENETICS HEREDITY	5	2.165	26	BIOPHYSICS	5	0.597
27	FOOD SCIENCE TECHNOLOGY	5	2.165	27	NUTRITION DIETETICS	4	0.478
28	DEVELOPMENTAL BIOLOGY	5	2.165	28	EMERGENCY MEDICINE	4	0.478
29	RESPIRATORY SYSTEM	4	1.732	29	BIOTECHNOLOGY APPLIED MICROBIOLOGY	4	0.478
30	PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALT	4	1.732	30	BIOLOGY	4	0.478
31	PARASITOLOGY	4	1.732	31	ALLERGY	4	0.478
32	MULTIDISCIPLINARY SCIENCES	4	1.732	32	TROPICAL MEDICINE	3	0.358

Figure 2. Excerpt of the Veterinary 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 "Veterinary Sciences panel and research groups together_WoS SCs.xlsx"
```

This produces a new Excel file called ‘Veterinary 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). We find positive correlation coefficient ($r = 0.70$, $p = 0.30$). Figure 4 shows a log-log plot of the number of publications per WoS SCs for the Veterinary Sciences panel and research groups together.

A	B	C	D
	Web of Science Categories	records_x	records_y
0	VETERINARY SCIENCES	455	42
1	REPRODUCTIVE BIOLOGY	202	29
2	CELL BIOLOGY	16	28
3	NEUROSCIENCES	0	27
4	AGRICULTURE DAIRY ANIMAL SCIENCE	107	20
5	GASTROENTEROLOGY HEPATOLOGY	2	17
6	BIOCHEMISTRY MOLECULAR BIOLOGY	16	15
7	PHYSIOLOGY	2	14
8	PHARMACOLOGY PHARMACY	0	14
9	ONCOLOGY	18	11
10	CLINICAL NEUROLOGY	6	11
11	CARDIAC CARDIOVASCULAR SYSTEMS	0	9
12	PATHOLOGY	3	8
13	ANATOMY MORPHOLOGY	67	8
14	OBSTETRICS GYNECOLOGY	3	7

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

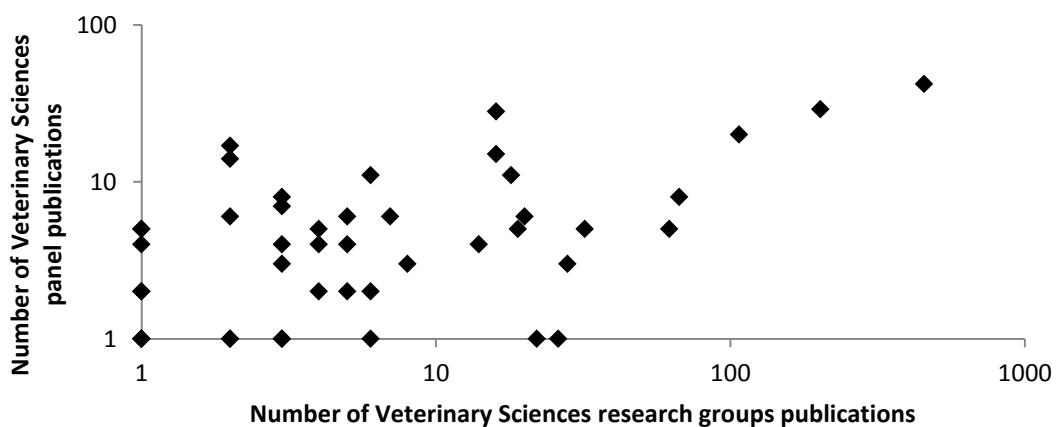


Figure 4. Log-log plot of the number of publications (log-log scale) per WoS SC for the panel (vertical axis) and research groups together (horizontal axis) of the Veterinary Sciences department

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:

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

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:

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

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 ‘Veterinary 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 "Veterinary Sciences panel and research groups together_WoS SCs-joined.xlsx"
```

The outcome shows that the top-down correlation between Veterinary Sciences research groups together and the panel's profile in the WoS SCs is moderate (0.52).

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

2.3 Web of Science subject categories similarity matrix

We download the global map of science based on WoS SCs data made available at <http://www.leydesdorff.net/overlaytoolkit/map10.paj>. These authors (Leydesdorff & Rafols, 2009; Rafols, Porter, & Leydesdorff, 2010; Leydesdorff, Carley, & Rafols, 2013) created a matrix of citing to cited WoS SCs based on the SCIE and SSCI, which was subsequently normalized in the citing direction. Only cosine values > 0.15 were retained. The result is a symmetric $N \times N$ 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.



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

Figure 5. Excerpt of the map10.paj file

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

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

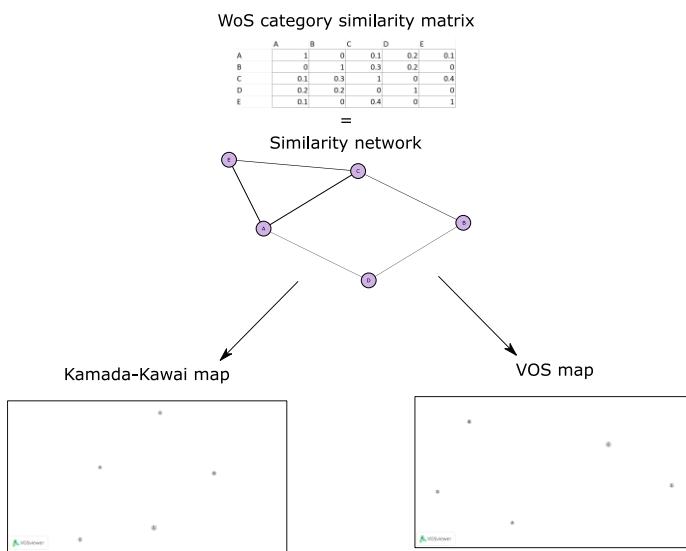


Figure 6. Transformation of WoS SCs similarity matrix to Kamada-Kawai map and VOS map

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

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

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

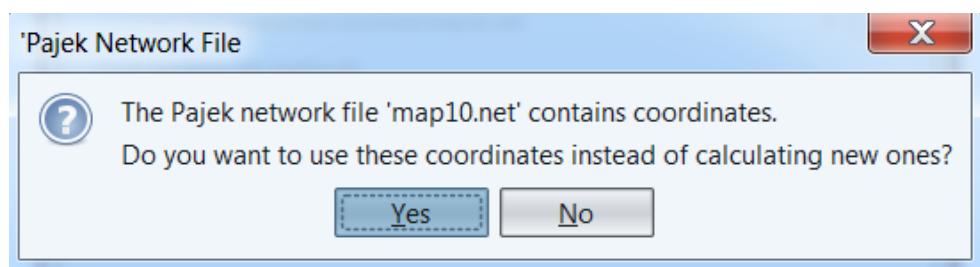


Figure 7. VOSviewer message before choosing Kamada Kawai map or VOS map data

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

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	-0.6842	-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 the 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 than 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 generates 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 VETE-B research group’s publications overlay map in WoS SCs.

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

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

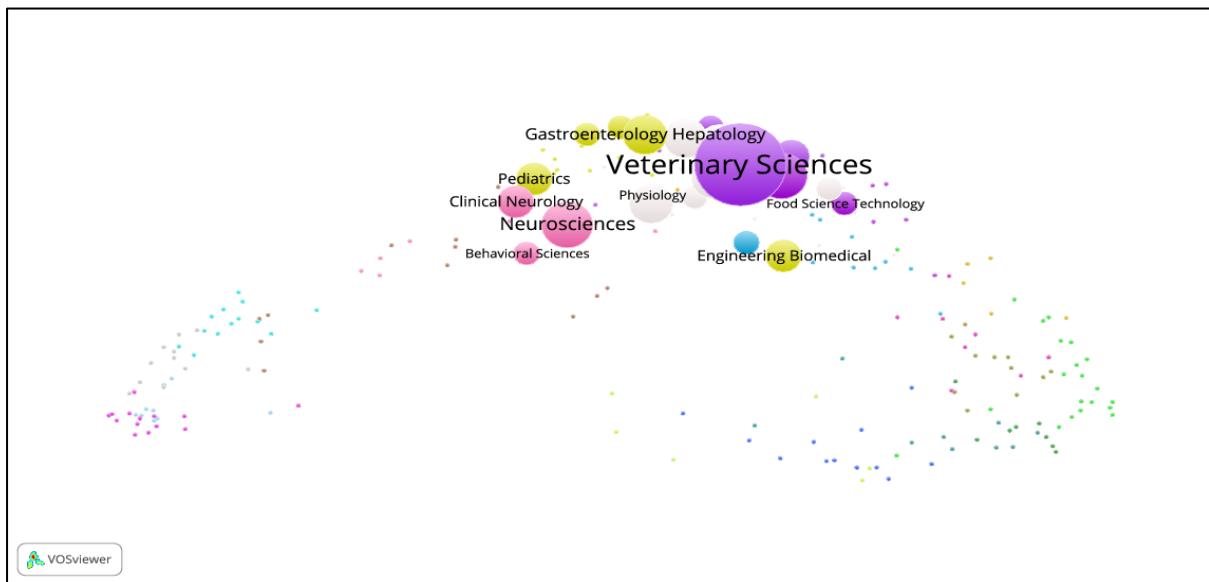


Figure 11. VETE-B research group’s publication overlay map in WoS SCs.

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 the 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) determines 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/faff8dc090b67a783b85d488f88952ba>).

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

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

Here, $L_{j,1}$ and $L_{j,2}$ 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 (‘Veterinary 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" "Veterinary Sciences
research groups and panel_WoS SCs.xlsx"
```

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

	A	B	C
1		x	y
2	Groups	0.091037	0.366106
3	PM1	0.177587	0.41017
4	PM2	0.123059	0.419574
5	PM3	0.173915	0.398628
6	PM4	-0.06624	0.356255
7	Panel	0.132225	0.400908
8	VETE-A	0.046782	0.367721
9	VETE-B	0.072995	0.373413
10	VETE-C	0.206248	0.358347

Figure 12. Barycenter coordinates of the Veterinary 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 $C = (C_1, C_2)$ and $D = (D_1, D_2)$ is calculated as follows:

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

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).

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.

	A	B	C	D	E	F	G	H	I	J
1		Groups	PM1	PM2	PM3	PM4	Panel	VETE-A	VETE-B	VETE-C
2	Groups	0	0.097122	0.062324	0.089031	0.157584	0.053922	0.044284	0.019465	0.115473
3	PM1	0.097122	0	0.055333	0.012112	0.249716	0.046299	0.137521	0.110863	0.05922
4	PM2	0.062324	0.055333	0	0.055	0.199608	0.020795	0.092234	0.068098	0.103291
5	PM3	0.089031	0.012112	0.055	0	0.243863	0.041752	0.130836	0.104022	0.051652
6	PM4	0.157584	0.249716	0.199608	0.243863	0	0.203425	0.113601	0.140287	0.272496
7	Panel	0.053922	0.046299	0.020795	0.041752	0.203425	0	0.091662	0.0653	0.085387
8	VETE-A	0.044284	0.137521	0.092234	0.130836	0.113601	0.091662	0	0.026824	0.159742
9	VETE-B	0.019465	0.110863	0.068098	0.104022	0.140287	0.0653	0.026824	0	0.134102
10	VETE-C	0.115473	0.05922	0.103291	0.051652	0.272496	0.085387	0.159742	0.134102	0

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

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.

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

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.054	0.092	0.065	0.085
PM1	0.097	0.138	0.111	0.059
PM2	0.062	<u>0.092</u>	<u>0.068</u>	<u>0.103</u>
PM3	0.089	0.131	0.104	0.052
PM4	0.158	0.114	0.140	0.272

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.624 (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).

c) Barycenter overlay map

We take the ‘WoS SC_VOS_map.xlsx’ (Figure 10) file and manually input the Veterinary 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.

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 Veterinary 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 Veterinary 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.

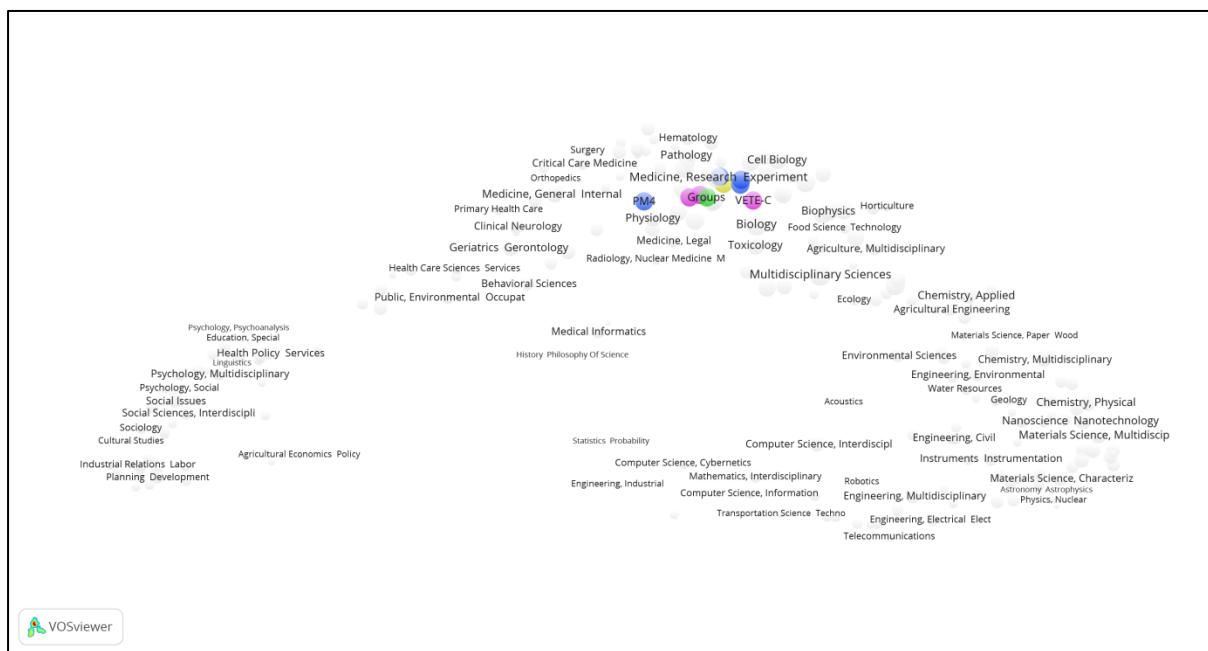


Figure 14. Barycenter overlay map of Veterinary Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs

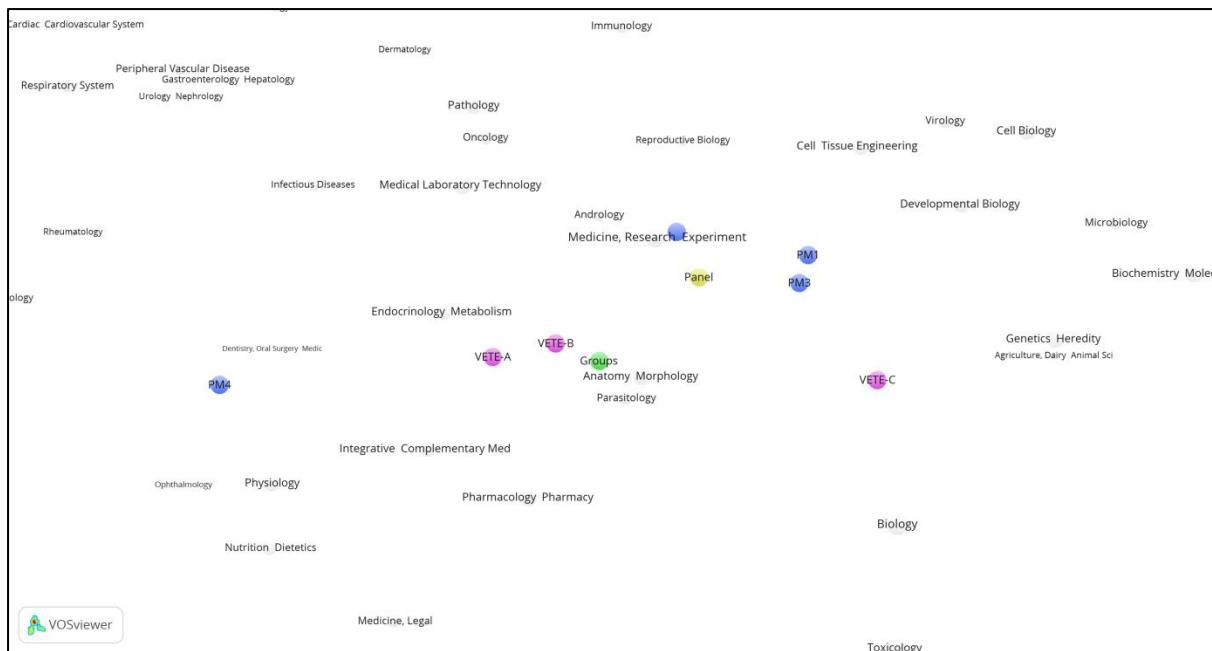


Figure 15. Barycenter map of Veterinary Sciences individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed)

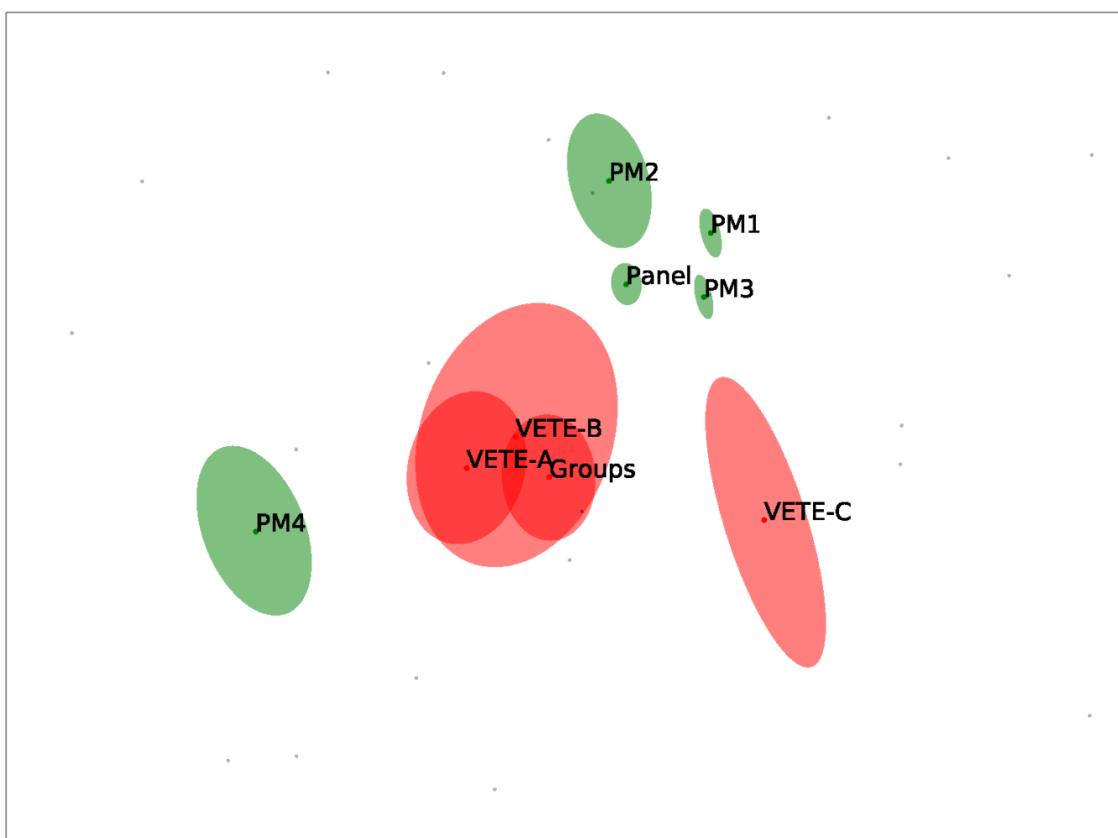


Figure 16. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together (groups) with their confidence regions

The confidence region is an ellipse that covers 95% of the bootstrapped barycenters. The larger the confidence region, the less stable the barycenter is. Although the CI of the distance between two barycenters and their confidence regions are related, the two should not be conflated. In particular, we stress that overlapping confidence regions as seen in Figure 16 (figure with overlapping regions in it) does not correspond to overlap between CIs for distances. The maps were plotted using Matplotlib (<http://matplotlib.org>). First, the base map was plotted using the pre-existing coordinates. Next, the barycenters were added as slightly larger red or green points. Finally, a partially translucent confidence region (ellipse) was calculated and superimposed on the map. Calculation of the confidence region was done using an implementation by Kington (2014). We briefly outline what elements determine the location and placement of such a confidence ellipse. The center of the ellipse is simply the mean of all bootstrapped barycenters. The width and height of the ellipse (or its axes) depend on the variance in the cloud of points. Finally, the orientation of the ellipse is obtained from the largest eigenvector.

2.7 Similarity-adapted publication vector method

a) *Similarity-adapted publication vector calculation*

A similarity-adapted publication vector (SAPV) is determined as the vector $C = (C_1, C_2, \dots, C_N)$, where:

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

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 $S * M$, 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.

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	B	C	D	E	F	G	H	I	J	K	L	M	N
	Acoustics	Al Economical	Environmental	Engin	Dairy	Anire, Multidi-	Agronomy	Allergy	my	Morp	Andrology	esthesiology	anthropolog	area Studie
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	0.185
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 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 Veterinary 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" "Veterinary 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 ‘Veterinary Sciences research groups and panel_WoS SCs-sa-vectors.xlsx’ (Figure 18).

	A	B	C	D	E	F	G	H	I	J	K	L
	Groups	Acoustics	Agricultur	Agricultur	Agricultur	Agricultur	Agronom	Allergy	Anatomy	Andrology	Anesthesi	Anthropol
2	Groups	0.000362	0	0.004713	0.010439	0.008053	0.003053	0.005005	0.021982	0.01661	0.005375	0
3	PM1	2.8E-05	0	0.003091	0.037933	0.012358	0.001595	0.004454	0.027437	0.038502	0.000724	0
4	PM2	0	0	0.003998	0.010314	0.006977	0.00222	0.019958	0.018224	0.013076	0.001962	0
5	PM3	0	0	0.00122	0.041946	0.015773	0.001834	0.012379	0.021224	0.019832	0.000135	0
6	PM4	0.003979	0	0.000286	0.003775	0.003405	0.000641	0.000627	0.029221	0.010862	0.011847	0
7	Panel	0.000714	0	0.002415	0.0272	0.010371	0.001597	0.008392	0.02472	0.024968	0.002817	0
8	VETE-A	0.000268	0	0.004168	0.003517	0.005817	0.003085	0.004945	0.021945	0.013389	0.006291	0
9	VETE-B	0.00041	0	0.003513	0.015454	0.009343	0.001735	0.00578	0.023024	0.015791	0.006176	0
10	VETE-C	0.000687	0	0.007687	0.036052	0.016267	0.00371	0.004543	0.021698	0.031734	0.001026	0

Figure 18. Excerpt of SAPV of the Veterinary 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. The Euclidean distance between vectors a and b in \mathbf{R}^N is:

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

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 (Figure 19) in an output file ‘Veterinary Sciences research groups and panel_WoS SCs-sa-vectors.xlsx’.

	A	B	C	D	E	F	G	H	I	J	
1		Groups	PM1	PM2	PM3	PM4	Panel	VETE-A	VETE-B	VETE-C	
2	Groups	0	0.068942	0.042845	0.092707	0.058476	0.04305107	0.01557	0.013553	0.063222	
3	PM1	0.0689425	0	0.075759	0.063205	0.110539	0.03520353	0.083332	0.065564	0.028007	
4	PM2	0.0428445	0.075759	0	0.078285	0.083447	0.04541553	0.048167	0.041034	0.071664	
5	PM3	0.0927069	0.063205	0.078285	0	0.131865	0.05639262	0.105717	0.084843	0.061636	
6	PM4	0.0584762	0.110539	0.083447	0.131865	0	0.08480752	0.052942	0.058376	0.106137	
7	Panel	0.0430511	0.035204	0.045416	0.056393	0.084808	0	0.057323	0.036811	0.035382	
8	VETE-A	0.0155703	0.083332	0.048167	0.105717	0.052942	0.05732288	0	0.024672	0.07864	
9	VETE-B	0.0135528	0.065564	0.041034	0.084843	0.058376	0.03681111	0.024672	0	0.06068	
10	VETE-C	0.0632223	0.028007	0.071664	0.061636	0.106137	0.03538196	0.07864	0.06068	0	

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

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

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

Groups	VETE-A	VETE-B	VETE-C
Panel	0.043	0.057	0.037
PM1	0.069	0.083	0.028
PM2	0.043	<u>0.048</u>	<u>0.041</u>
PM3	0.093	0.106	0.085
PM4	0.058	0.053	0.058

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.040 (SD 0.009). The number in the row of this panel member is indicated in bold and underlined. Distances whose confidence intervals overlap with that of the shortest distance are in bold (same column).

c) Similarity-adapted publication vector overlay map

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

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

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

A	B	C	D	E	F
id	label	x	y	weight	cluster
37	Environmental Sciences	-0.7867	0.0784	16.2845	1
38	Psychology, Clinical	1.1541	-0.0221	10.8045	1
39	Biology	-0.1041	0.1811	36.7052	1
40	Agricultural Economics Policy	1.2902	-0.1991	5.484	1
41	Paleontology	-0.8934	0.0788	7.9912	1
42	Engineering, Civil	-0.9297	-0.0504	13.544	1
43	Immunology	0.0429	0.2384	18.6074	1
44	VETE-B	-0.0063	0.1894	20	3
45	Computer Science, Hardware Architecture	-0.6956	-0.2628	10.257	1
46	Polymer Science	-0.9535	-0.0929	14.2607	1
47	Public Administration	1.4279	-0.1748	9.366	1
48	Remote Sensing	-0.8821	-0.0241	9.165	1
49	Communication	1.3241	-0.1209	7.09	1
50	Metallurgy Metallurgical Engineering	-1.0234	-0.1403	15.1693	1
51	Hospitality, Leisure, Sport Tourism	1.3987	-0.1709	6.561	1
52	Zoology	-0.2823	0.1635	15.7178	1

Figure 20. Excerpt of VETE-B.csv file

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

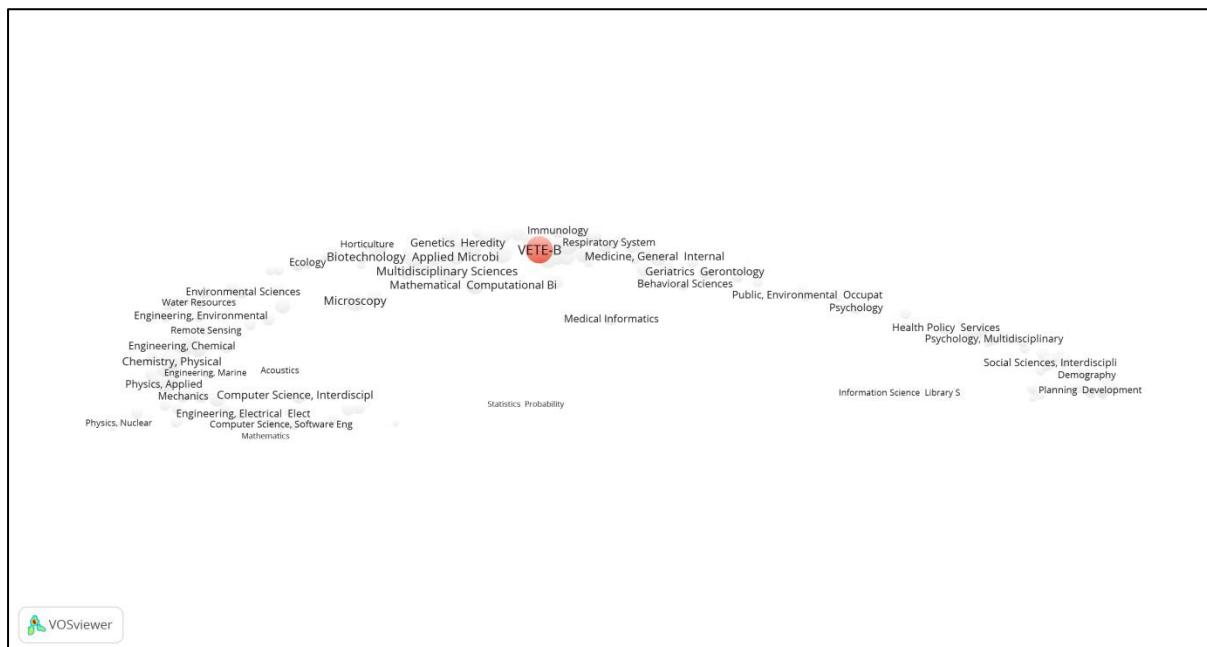


Figure 21. Location of the SAPV of VETE-B in the WoS SCs similarity matrix

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

$$\begin{aligned}
 & \frac{\sum_{i=1}^N M_i^k (\sum_{j=1}^N R_j^m s_{ji})}{\sqrt{(\sum_{i=1}^N M_i^k (\sum_{j=1}^N M_j^k s_{ji})) \cdot (\sum_{i=1}^N R_i^m (\sum_{j=1}^N R_j^m s_{ji}))}} \\
 & = \frac{(M^k)^t * S * R^m}{\sqrt{(M^k)^t * S * M^k} \cdot \sqrt{(R^m)^t * S * R^m}}
 \end{aligned} \tag{7}$$

The numerator is nothing but the matrix multiplication: $(M^k)^t * S * R^m$, 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: $(M^k)^t * S * M^k$ and $(R^m)^t * S * R^m$. 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 (‘Veterinary 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" "Veterinary Sciences research groups and panel_WoS SCs.xlsx"
```

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

	A	B	C	D	E	F	G	H	I	J
1		Groups	PM1	PM2	PM3	PM4	Panel	VETE-A	VETE-B	VETE-C
2	Groups	1	0.725879	0.798593	0.573541	0.632715	0.806492	0.939766	0.959025	0.707878
3	PM1	0.725879	1	0.612041	0.788922	0.301719	0.948878	0.460147	0.783193	0.981881
4	PM2	0.798593	0.612041	1	0.66922	0.402484	0.773179	0.72194	0.817908	0.601805
5	PM3	0.573541	0.788922	0.66922	1	0.172419	0.892399	0.320748	0.731693	0.815353
6	PM4	0.632715	0.301719	0.402484	0.172419	1	0.434149	0.674096	0.615678	0.262761
7	Panel	0.806492	0.948878	0.773179	0.892399	0.434149	1	0.571804	0.890404	0.940231
8	VETE-A	0.939766	0.460147	0.72194	0.320748	0.674096	0.571804	1	0.841716	0.427097
9	VETE-B	0.959025	0.783193	0.817908	0.731693	0.615678	0.890404	0.841716	1	0.774815
10	VETE-C	0.707878	0.981881	0.601805	0.815353	0.262761	0.940231	0.427097	0.774815	1

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

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

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.806	0.572	0.890	0.940
PM1	0.726	0.460	0.783	0.982
PM2	0.799	0.722	0.818	0.602
PM3	0.574	0.321	0.732	0.815
PM4	0.633	0.674	0.616	0.263

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).

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.

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

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.194	0.428	0.110	0.060
PM1	0.274	0.540	0.217	0.018
PM2	0.201	0.278	0.182	0.398
PM3	0.426	0.679	0.268	0.185
PM4	0.367	0.326	0.384	0.737

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 Veterinary Sciences individual research groups and panel members

Pearson Spearman	Barycenter	SAPV	WCS
Barycenter	1.00	0.70	0.83
SAPV	0.61	1.00	0.87
WCS	0.83	0.75	1.00

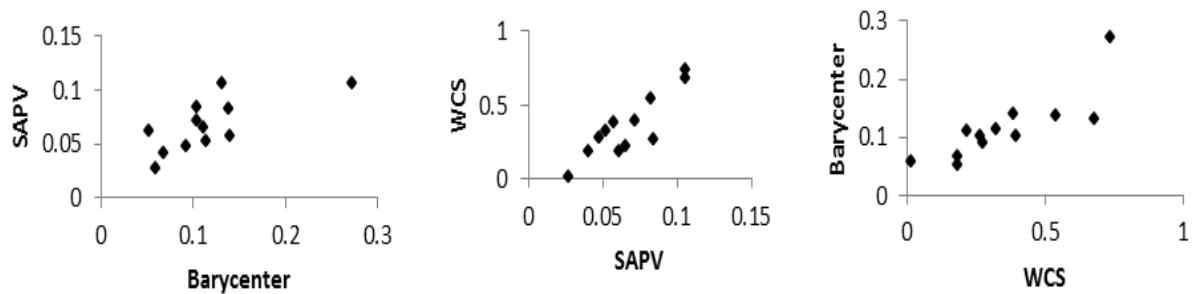


Figure 23. Scatter plot of the correlation between barycenter, SAPV and WCS

Since our barycenter method (see section 2.6) and SAPV method (see section 2.7) are distance-based rather than similarity-based, we use $1 - \text{WCS}$ 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.

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.

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 all methods are moderate to strongly correlated.

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 'VETE-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
Source Titles	records	% of 231
THERIOGENOLOGY	11	4.762
REPRODUCTION IN DOMESTIC ANIMALS	10	4.329
NEUROGASTROENTEROLOGY AND MOTILITY	7	3.03
HISTOCHEMISTRY AND CELL BIOLOGY	5	2.165
VLAAMS DIERGENEESKUNDIG TIJDSCHRIFT	4	1.732
PLOS ONE	4	1.732
JOURNAL OF DAIRY SCIENCE	4	1.732
JOURNAL OF CELLULAR AND MOLECULAR MEDICINE	4	1.732
CHEMOSPHERE	4	1.732
CANCER EPIDEMIOLOGY BIOMARKERS PREVENTION	4	1.732

◀ ▶ ⏪ Groups ⏩ VETE-A VETE-B VETE-C Panel VETE PM1 VETE PM2 VETE PM3 VETE PM4

Figure 24. Excerpt of Veterinary 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 (VETE-A, VETE-B, VETE-C, etc.), the panel members’ code names, (PM1, PM2, PM3, etc.), Panel members together and Groups together. The Excel file is saved as ‘Veterinary 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 ‘Veterinary 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).

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

It produces a new Excel file named ‘Veterinary 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).

	A	B	C
1	Source Titles	records	% of 231
2	THERIOGENOLOGY	11	4.762
3	REPRODUCTION IN DOMESTIC ANIMALS	10	4.329
4	NEUROGASTROENTEROLOGY AND MOTILITY	7	3.03
5	HISTOCHEMISTRY AND CELL BIOLOGY	5	2.165
6	VLAAMS DIERGENEESKUNDIG TIJDSCHRIFT	4	1.732
7	PLOS ONE	4	1.732
8	JOURNAL OF DAIRY SCIENCE	4	1.732
9	JOURNAL OF CELLULAR AND MOLECULAR MEDICINE	4	1.732
10	CHEMOSPHERE	4	1.732
11	CANCER EPIDEMIOLOGY BIOMARKERS PREVENTION	4	1.732
12	AUTONOMIC NEUROSCIENCE BASIC CLINICAL	4	1.732
13	AMERICAN JOURNAL OF RESPIRATORY CELL AND MOLECUL	4	1.732
14	AMERICAN JOURNAL OF PHYSIOLOGY GASTROINTESTINAL ,	4	1.732
15	LIVESTOCK SCIENCE	3	1.299
16	JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERA	3	1.299
17	JOURNAL OF CLINICAL PATHOLOGY	3	1.299
18	HUMAN REPRODUCTION	3	1.299
19	EUROPEAN JOURNAL OF CANCER PREVENTION	3	1.299
20	EQUINE VETERINARY JOURNAL	3	1.299
21	BRITISH JOURNAL OF PHARMACOLOGY	3	1.299
22	ANATOMICAL RECORD ADVANCES IN INTEGRATIVE ANATOI	3	1.299
23	VETERINARY MICROBIOLOGY	2	0.866
24	VETERINARY JOURNAL	2	0.866
25	REPRODUCTION FERTILITY AND DEVELOPMENT	2	0.866
26	PARASITOLOGY	2	0.866
27	NEUROSCIENCE LETTERS	2	0.866
28	JOURNAL OF NEUROSCIENCE	2	0.866
29	JOURNAL OF HISTOCHEMISTRY CYTOCHEMISTRY	2	0.866
30	JOURNAL OF HEPATOLOGY	2	0.866
31	JOURNAL OF CLINICAL MICROBIOLOGY	2	0.866
32	JOURNAL OF APPLIED PHYSIOLOGY	2	0.866

	A	B	C
1	Source Titles	records	% of 837
2	THERIOGENOLOGY	71	8.483
3	REPRODUCTION IN DOMESTIC ANIMALS	47	5.615
4	ANIMAL REPRODUCTION SCIENCE	40	4.779
5	BERLINER UND MUNCHENER TIERARZTLICHE WOCHENSCHRIFT	38	4.54
6	KLEINTIERPRAXIS	30	3.584
7	VETERINARY IMMUNOLOGY AND IMMUNOPATHOLOGY	26	3.106
8	TIERARZTLICHE PRAXIS AUSGABE KLEINTIERE HEIMTIERE	26	3.106
9	JOURNAL OF VETERINARY MEDICINE SERIES A ZENTRALBLATT	26	3.106
10	PRAKTIISCHE TIERARZT	19	2.27
11	DEUTSCHE TIERARZTLICHE WOCHENSCHRIFT	18	2.151
12	ACTA ANATOMICA	16	1.912
13	JOURNAL OF VETERINARY MEDICINE SERIES A PHYSIOLOG	15	1.792
14	INTERNATIONAL JOURNAL OF ANDROLOGY	15	1.792
15	TIERARZTLICHE PRAXIS AUSGABE KLEINTIERE HEIMTIERE	13	1.553
16	JOURNAL OF ANATOMY	13	1.553
17	ACTA VETERINARIA SCANDINAVICA	13	1.553
18	REPRODUCTION FERTILITY AND DEVELOPMENT	11	1.314
19	ORTHOPADE	11	1.314
20	AMERICAN JOURNAL OF VETERINARY RESEARCH	11	1.314
21	ZYGOTE	8	0.956
22	SURGICAL AND RADIOLOGIC ANATOMY	8	0.956
23	ANTICANCER RESEARCH	8	0.956
24	ZEITSCHRIFT FUR ORTHOPADIE UND IHRE GRENZGEBIETE	7	0.836
25	RESEARCH IN VETERINARY SCIENCE	7	0.836
26	JOURNAL OF VETERINARY INTERNAL MEDICINE	7	0.836
27	JOURNAL OF SMALL ANIMAL PRACTICE	7	0.836
28	ANNALS OF ANATOMY ANATOMISCHER ANZEIGER	7	0.836
29	VETERINARY RECORD	6	0.717
30	MAGYAR ALLATORVOSOK LAPJA	6	0.717
31	IMMUNOLOGY	6	0.717
32	CLINICAL AND EXPERIMENTAL IMMUNOLOGY	6	0.717

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

We calculate correlation using SPSS and obtain correlation coefficient value ($r = 0.29$, $p = -0.66$). A log-log plot of the number of publications per journal for the Veterinary Sciences panel and research groups together are shown in Figure 27.

A	B	C	D
	Source Titles	records_x	records_y
1	0 THERIOGENOLOGY	11	71
2	1 REPRODUCTION IN DOMESTIC ANIMALS	10	47
3	2 NEUROGASTROENTEROLOGY AND MOTILITY	7	0
4	3 HISTOCHEMISTRY AND CELL BIOLOGY	5	0
5	4 VLAAMS DIERGENEESKUNDIG TIJDSCHRIFT	4	0
6	5 PLOS ONE	4	4
7	6 JOURNAL OF DAIRY SCIENCE	4	0
8	7 JOURNAL OF CELLULAR AND MOLECULAR MEDICINE	4	0
9	8 CHEMOSPHERE	4	0
10	9 CANCER EPIDEMIOLOGY BIOMARKERS PREVENTION	4	0
11	10 AUTONOMIC NEUROSCIENCE BASIC CLINICAL	4	0
12	11 AMERICAN JOURNAL OF RESPIRATORY CELL AND MO	4	0
13	12 AMERICAN JOURNAL OF PHYSIOLOGY GASTROINTESTI	4	0
14	13 LIVESTOCK SCIENCE	3	2
15	14 JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL TH	3	0
16	15 JOURNAL OF CLINICAL PATHOLOGY	3	0
17	16 HUMAN REPRODUCTION	3	0
18	17 EUROPEAN JOURNAL OF CANCER PREVENTION	3	0
19	18 EQUINE VETERINARY JOURNAL	3	3

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

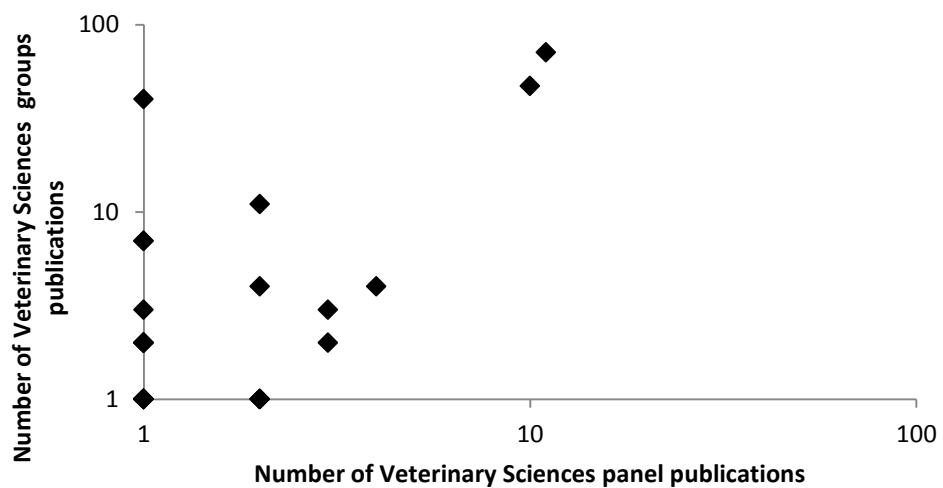


Figure 27. Log-log plot of the number of publications (log-log scale) per journals for the panel (horizontal axis) and research groups together (vertical axis) of the Veterinary Sciences department.

b) Top-Down correlation coefficient

In some cases, the panel has published in a journal where the research groups have not or vice versa , i.e. there are many zeroes on both sides. Since traditional correlation coefficients like

Pearson's and Spearman's are not well-suited to zero-inflated data (i.e., data with a large amounts of zeroes), we adopt the top-down correlation coefficient (Iman & Conover, 1987). This correlation coefficient was found to be an adequate rank correlation coefficient for zero-inflated data (Huson, 2007). For a full description of the top-down correlation coefficient we refer to Iman and Conover (1987). This coefficient places emphasis on the higher ranked data by computing the correlation using Savage scores derived from the ranked data.

We reuse the formula 1 and 2 (details in section 2.2b) and the python script “calc_topdowncorr.py” (all core logic is in topdowncorr.py, see section 5). We reuse the ‘Excerpt of the Chemistry 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 "Veterinary Sciences research groups and panel_journals title-joined.xlsx"
```

The outcome shows that the top-down correlation between Veterinary Sciences research groups together and the panel based on the journals in which they publish is very low (0.06).

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 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 Leydesdorff, Rafols, & Chen (2013) for details). The resulting journal similarity matrix can be considered as an adjacency matrix, and thus is equivalent to a weighted network where similar journals are linked and link weights increase with similarity strength.

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

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

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

Figure 28. Excerpt of the journals VOS map data

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

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

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>.

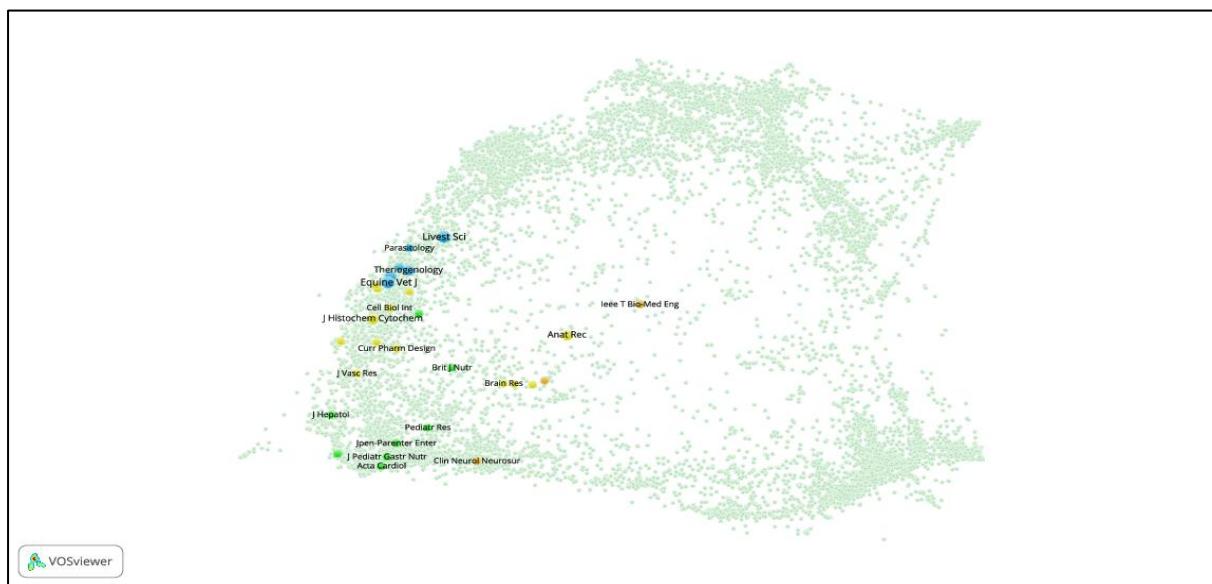


Figure 29. Journal overlay map of the VETE-B research group

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 VETE-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 $C = (C_1, C_2)$, where

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

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

Based on formula 3, Python script ‘journals-barycenter.py’ is used. This script takes ‘Journals_VOS_map.xlsx’ (Figure 28) and ‘Veterinary 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" "Veterinary Sciences  
research groups and panel journals title.xlsx"
```

	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

At this point, we notice that our program indicates that the journal titles in our Veterinary 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.

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.

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

Figure 31. Excerpt of Journal name change.xlsx file

We keep the ‘Veterinary 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" "Veterinary Sciences
research groups and panel journals title.xlsx"
```

	A	B	C
1		x	y
2	Groups	-0.73315	-0.0918
3	Panel	-0.74421	-0.00028
4	PM1	-0.76975	0.082076
5	PM2	-0.80556	-0.04265
6	PM3	-0.79554	0.093096
7	PM4	-0.53822	-0.32711
8	VETE-A	-0.76515	-0.178
9	VETE-B	-0.6995	-0.0619
10	VETE-C	-0.65431	0.12724

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

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

b) Euclidean distance 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	J
1		Groups	Panel	PM1	PM2	PM3	PM4	VETE-A	VETE-B	VETE-C
2	Groups	0	0.092177	0.177682	0.087506	0.195133	0.305567	0.091956	0.045011	0.232792
3	Panel	0.092177	0	0.08623	0.074554	0.106557	0.386325	0.178949	0.076131	0.156026
4	PM1	0.177682	0.08623	0	0.129766	0.02804	0.47015	0.260121	0.160205	0.123963
5	PM2	0.087506	0.074554	0.129766	0	0.136119	0.390363	0.141252	0.107788	0.227461
6	PM3	0.195133	0.106557	0.02804	0.136119	0	0.492732	0.272798	0.18234	0.145295
7	PM4	0.305567	0.386325	0.47015	0.390363	0.492732	0	0.271534	0.310397	0.468946
8	VETE-A	0.091956	0.178949	0.260121	0.141252	0.272798	0.271534	0	0.133378	0.324745
9	VETE-B	0.045011	0.076131	0.160205	0.107788	0.18234	0.310397	0.133378	0	0.194466
10	VETE-C	0.232792	0.156026	0.123963	0.227461	0.145295	0.468946	0.324745	0.194466	0

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

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.

Table 8. Euclidean distances between Veterinary Sciences individual research groups, panel members, research groups together and panel using the journal VOS map

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.092	0.179	0.076	0.156
PM1	0.178	0.260	0.160	0.124
PM2	0.088	0.141	0.108	0.227
PM3	0.195	0.273	0.182	0.145

For each research group we determined the panel member at the shortest distance. Average shortest distance 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).

c) Barycenter overlay map

We take the Journal level_VOS map (Figure 28) file and manually add the Veterinary Sciences individual groups, panel members, research groups together and panel's coordinates (Figure 32) after the 10,673 journals title.

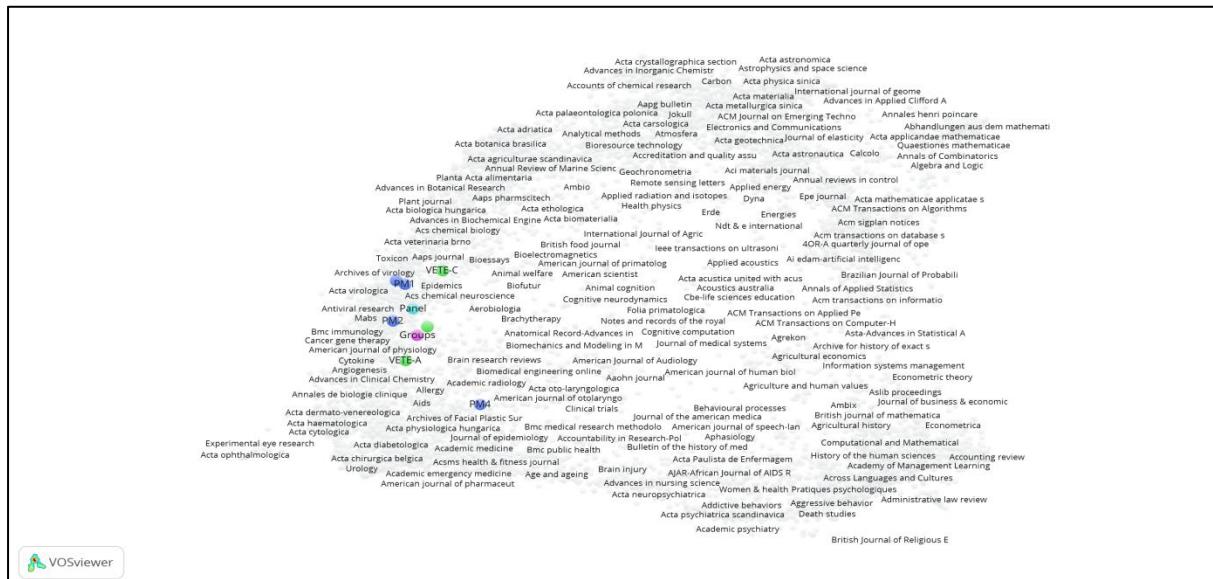


Figure 34. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together

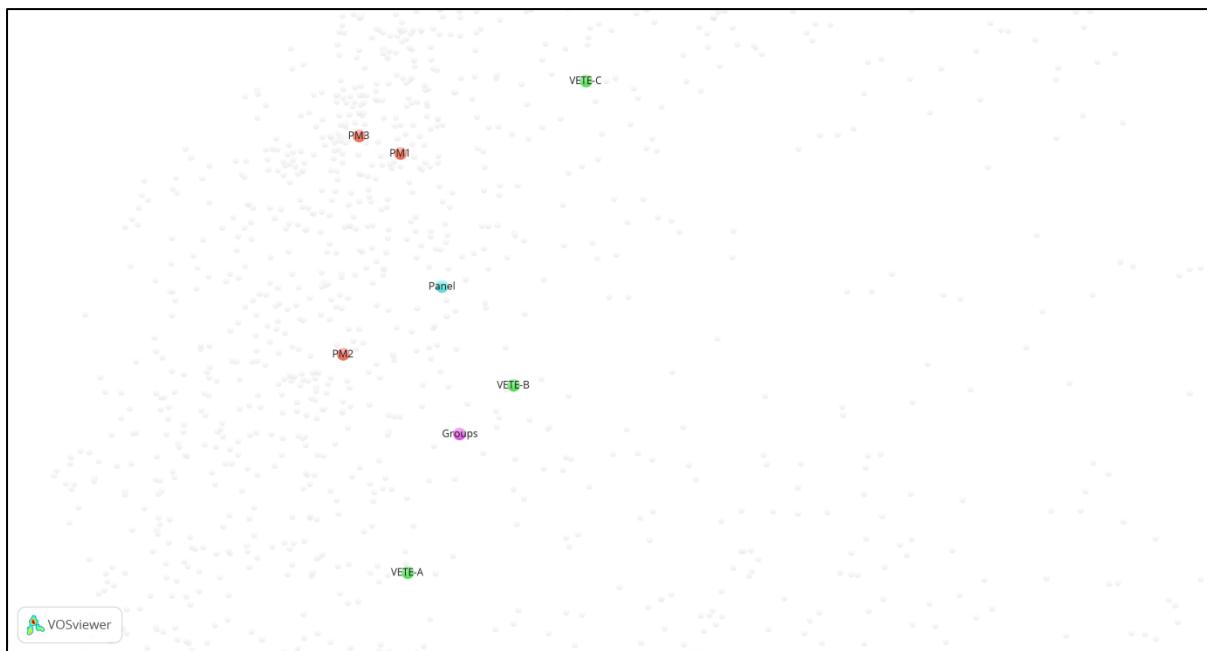


Figure 35. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together (zoomed)

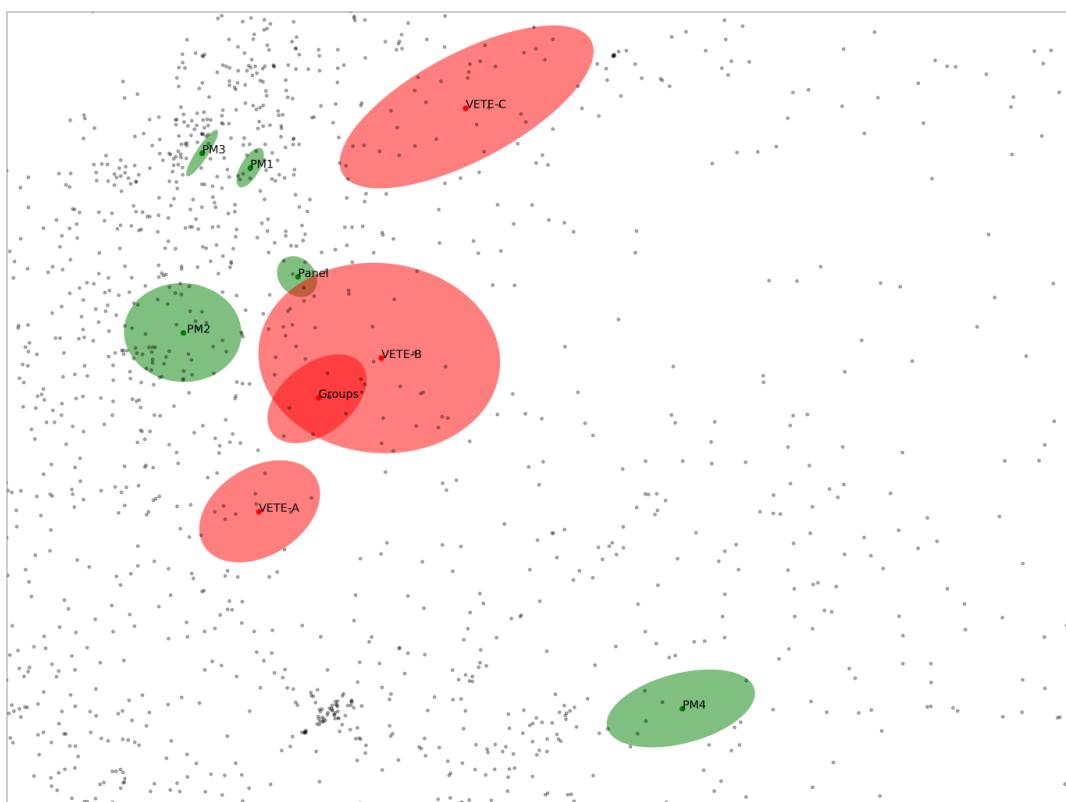


Figure 36. Barycenter overlay map of Veterinary Sciences panel, panel members (PM), research groups and research groups together with their confidence regions

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 Veterinary 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.

We also create the barycenter overlap map of Veterinary 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.

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 $C = (C_1, C_2, \dots, C_N)$, 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 matrix and the weights (number of publications) per journals (‘Veterinary 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 ‘Veterinary 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 "Veterinary Sciences research groups and panel_journals title.xlsx"
```

	A	B	C	D	E	F	G	H	I	J
1		Groups	Panel	PM1	PM2	PM3	PM4	VETE-A	VETE-B	VETE-C
2	Groups	0	0.0921768	0.1776823	0.08750552	0.195132677	0.3055672	0.0919564	0.0450111	0.232792
3	Panel	0.0921768	0	0.0862305	0.07455438	0.10655665	0.3863252	0.1789494	0.0761307	0.156026
4	PM1	0.1776823	0.0862305	0	0.12976609	0.028039721	0.4701502	0.260121	0.1602047	0.123963
5	PM2	0.0875055	0.0745544	0.1297661	0	0.136118703	0.3903635	0.1412525	0.1077882	0.227461
6	PM3	0.1951327	0.1065566	0.0280397	0.1361187	0	0.4927325	0.2727977	0.1823398	0.145295
7	PM4	0.3055672	0.3863252	0.4701502	0.39036349	0.492732483	0	0.2715335	0.3103966	0.468946
8	VETE-A	0.0919564	0.1789494	0.260121	0.14125246	0.272797716	0.2715335	0	0.1333779	0.324745
9	VETE-B	0.0450111	0.0761307	0.1602047	0.10778819	0.182339837	0.3103966	0.1333779	0	0.194466
10	VETE-C	0.2327919	0.1560264	0.1239629	0.22746106	0.145295118	0.4689463	0.3247451	0.1944656	0

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

This program calculates the SAPV (Figure 37) of each entity and stores the results in an output file named ‘Veterinary 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>).

From the calculated matrix of pairwise Euclidean distances between SAPVs of Veterinary 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.

	A	B	C	D	E	F	G	H	I	J
1		Groups	Panel	VETE PM1	VETE PM2	VETE PM3	VETE PM4	VETE-A	VETE-B	VETE-C
2	Groups	0	0.00651645	0.010954966	0.004879	0.014570972	0.00957278	0.001832	0.001866	0.010162
3	Panel	0.006516	0	0.007005243	0.006763	0.009657732	0.0114728	0.008141	0.005464	0.006445
4	PM1	0.010955	0.00700524	0	0.012047	0.01382914	0.01587799	0.012581	0.010418	0.004864
5	PM2	0.004879	0.00676348	0.01204653	0	0.01377888	0.01152712	0.005287	0.004643	0.011478
6	PM3	0.014571	0.00965773	0.01382914	0.013779	0	0.01798679	0.015815	0.013401	0.012557
7	PM4	0.009573	0.0114728	0.015877994	0.011527	0.01798679	0	0.009558	0.009562	0.014907
8	VETE-A	0.001832	0.00814079	0.012581102	0.005287	0.015815196	0.00955792	0	0.003297	0.011954
9	VETE-B	0.001866	0.00546354	0.010418216	0.004643	0.013400732	0.00956204	0.003297	0	0.009545
10	VETE-C	0.010162	0.00644516	0.004863826	0.011478	0.012556893	0.01490668	0.011954	0.009545	0

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

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

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.007	0.008	0.005	0.006
PM1	0.011	0.013	0.010	0.005
PM2	0.005	0.005	0.005	0.011
PM3	0.015	0.016	0.013	0.013
PM4	0.010	0.010	0.010	0.015

For each research group we determined the panel member at the shortest distance. Average shortest distance is 0.79 (SD 0.14). The number in the row of this panel member is indicated in bold and underlined. Distances whose confidence intervals overlap with that of the shortest distance are in bold (same column).

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

3.7 Weighted cosine similarity method

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

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

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

The numerator is nothing but the matrix multiplication: $\left(M^k \right)^t * S * R^m$, where t denotes matrix transposition, S is the journal similarity matrix, 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 ‘Veterinary 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 (‘Veterinary Sciences research groups and panel_journals title.xlsx’, Figure 24), and calculates SAPVs for all entities, as well as the pairwise distances between them.

	A	B	C	D	E	F	G	H	I	J
1		Groups	Panel	PM1	PM2	PM3	PM4	VETE-A	VETE-B	VETE-C
2	Groups	1	0.710268	0.55838	0.704693	0.372441	0.395037	0.936184	0.925337	0.584211
3	Panel	0.710268	1	0.843162	0.581043	0.712117	0.329861	0.464628	0.754017	0.843688
4	PM1	0.55838	0.843162	1	0.270427	0.311729	0.122543	0.267998	0.534294	0.941294
5	PM2	0.704693	0.581043	0.270427	1	0.434508	0.244016	0.690035	0.729261	0.286055
6	PM3	0.372441	0.712117	0.311729	0.434508	1	0.100972	0.231971	0.495799	0.399338
7	PM4	0.395037	0.329861	0.122543	0.244016	0.100972	1	0.409522	0.414427	0.114501
8	VETE-A	0.936184	0.464628	0.267998	0.690035	0.231971	0.409522	1	0.823895	0.273231
9	VETE-B	0.925337	0.754017	0.534294	0.729261	0.495799	0.414427	0.823895	1	0.556979
10	VETE-C	0.584211	0.843688	0.941294	0.286055	0.399338	0.114501	0.273231	0.556979	1

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

We run the program as:

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

This program calculates the WCS value in an output file as ‘Veterinary Sciences research groups and panel _ journals title-cosine.xlsx’ (Figure 39). From the calculated WCS value matrix of Veterinary 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.

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

	Groups	VETE-A	VETE-B	VETE-C
Panel	0.710	0.465	0.754	0.844
PM1	0.558	0.268	0.534	0.941
PM2	0.705	0.690	0.729	0.286
PM3	0.372	0.232	0.496	0.399
PM4	0.395	0.410	0.414	0.115

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).

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.

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

Groups	VETE-A	VETE-B	VETE-C
Panel	0.290	0.535	0.246
PM1	0.442	0.732	0.466
PM2	0.295	0.31	0.271
PM3	0.628	0.768	0.504
PM4	0.605	0.59	0.586
			0.885

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

Since the barycenter (see section 3.5) and SAPV (see section 3.6) approaches are distance-based rather than similarity-based, we use $1 - \text{WCS}$ 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.

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 $1 - \text{WCS}$. We create an MS Excel file (Figure 40) containing:

- i) Euclidean distances between barycenters of the Veterinary Sciences individual research groups and panel members at the level of WoS SCs,
- ii) Euclidean distances between barycenters of the Veterinary Sciences individual research groups and panel members at the level of journals,
- iii) Euclidean distances between SAPVs of the Veterinary Sciences individual research groups and panel members at the level of WoS SCs,
- iv) Euclidean distances between SAPVs of the Veterinary Sciences individual research groups and panel members at the level of journals,
- v) WCS value of Veterinary Sciences individual research groups and panel members at the level of WoS SCs,

	A	B	C	D	E	F
1	Barycenter_VETE_WoS SCs	Barycenter_VETE_Journals	SAPV_VETE_WoS SCs	SAPV_VETE_Journals	WCS_VETE_WoS SCs	WCS_VETE_Journals
2	0.138	0.260	0.083	0.013	0.540	0.732
3	0.111	0.160	0.066	0.010	0.217	0.466
4	0.059	0.124	0.028	0.005	0.018	0.059
5	0.092	0.141	0.048	0.005	0.278	0.310
6	0.068	0.108	0.041	0.005	0.182	0.271
7	0.103	0.227	0.072	0.011	0.398	0.714
8	0.131	0.273	0.106	0.016	0.679	0.768
9	0.104	0.182	0.085	0.013	0.268	0.504
10	0.052	0.145	0.062	0.013	0.185	0.601
11	0.114	0.272	0.053	0.010	0.326	0.590
12	0.140	0.310	0.058	0.010	0.384	0.586
13	0.272	0.469	0.106	0.015	0.737	0.885

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

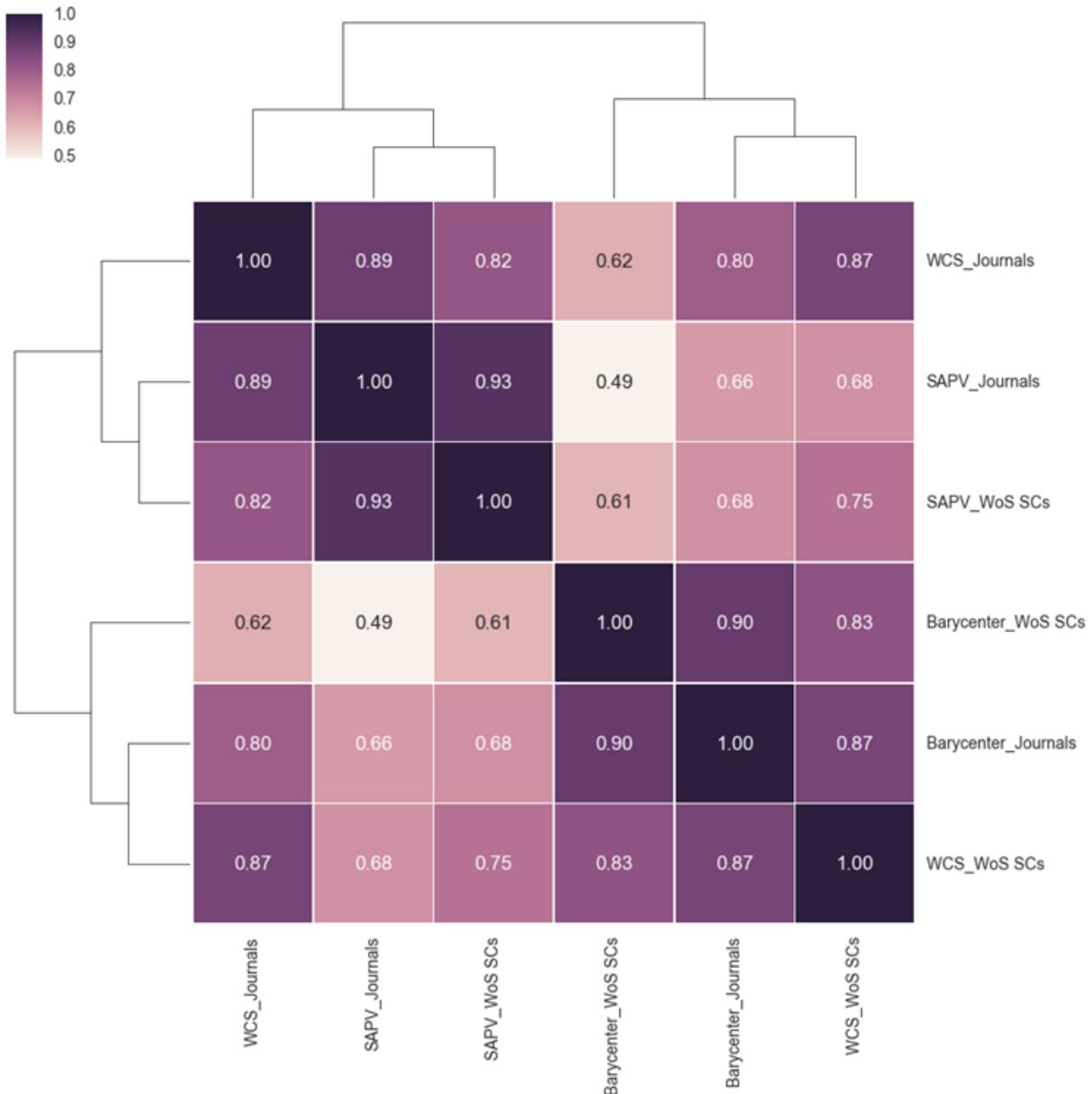


Figure 41. Heat map with hierarchical clustering based on correlation coefficient between six methods in the Veterinary sciences department

WCS value of Veterinary Sciences individual research groups and panel members at the level of journals. 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.

The heat map with hierarchical clustering (Figure 41) shows that correlations between the two levels of aggregation based on barycenter ($\rho = 0.90$), SAPV ($\rho = 0.93$) and WCS ($\rho = 0.87$) are strong. The correlations between the barycenter methods on the one hand and the

SAPV and WCS 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 analysis 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 evaluatees: 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

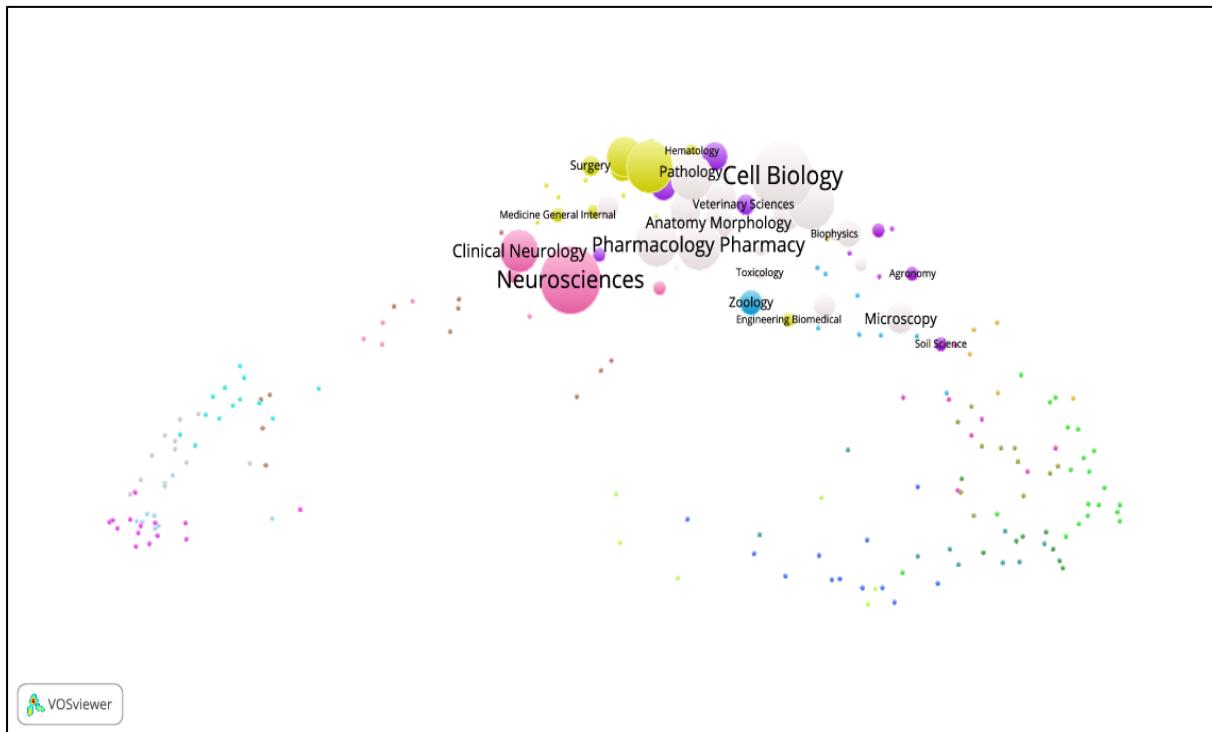


Figure 42. WoS SCs overlay map of VETE-A research group's publications

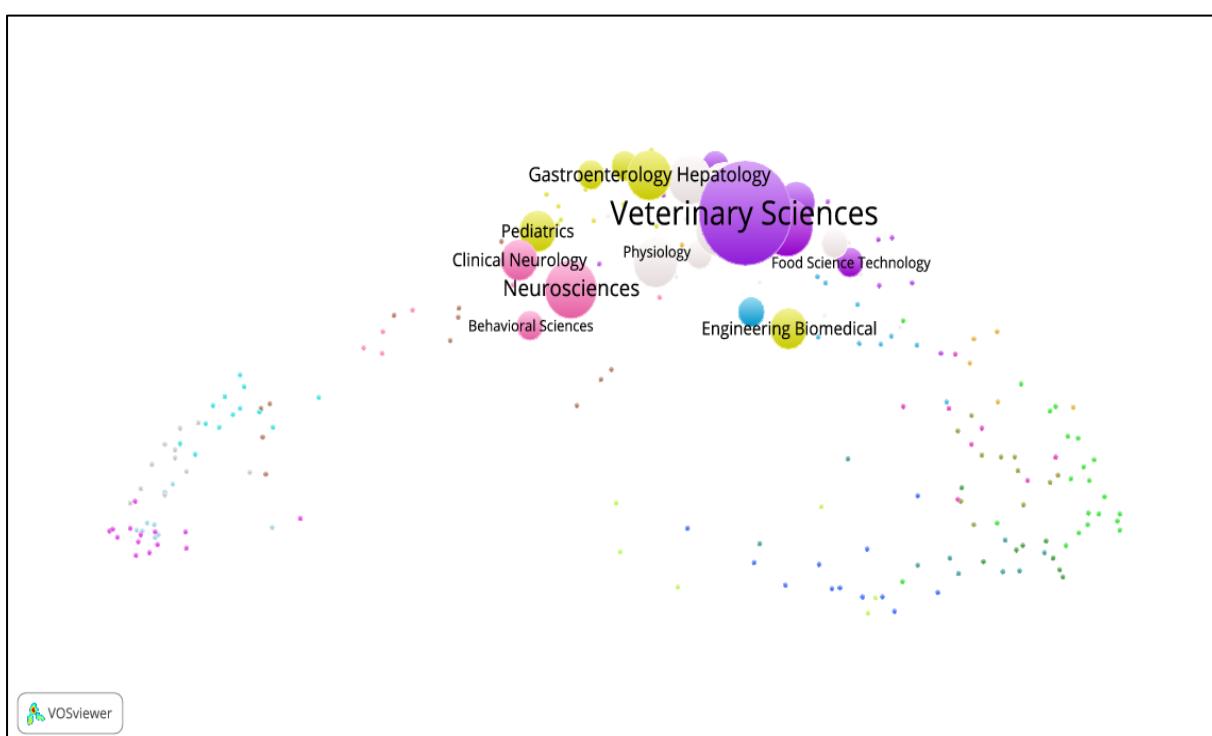


Figure 43. WoS SCs overlay map of VETE-B research group's publications

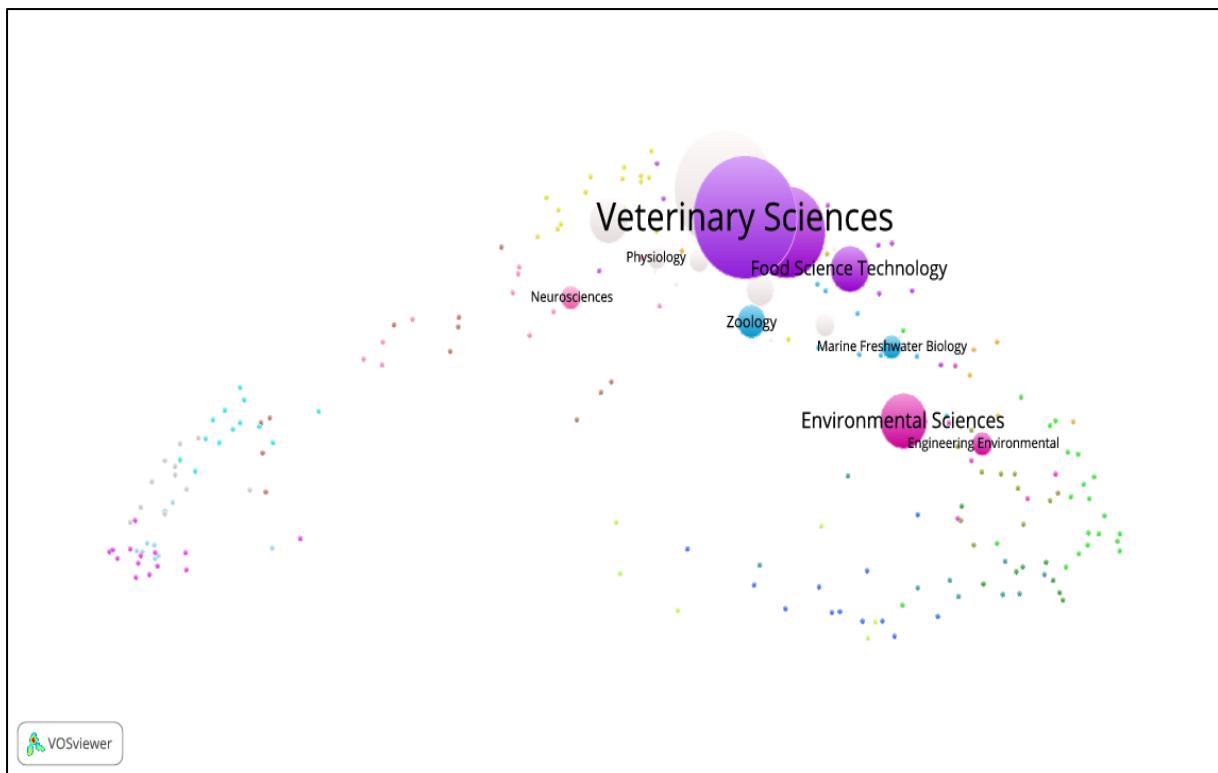


Figure 44. WoS SCs overlay map of VETE-C research group's publications

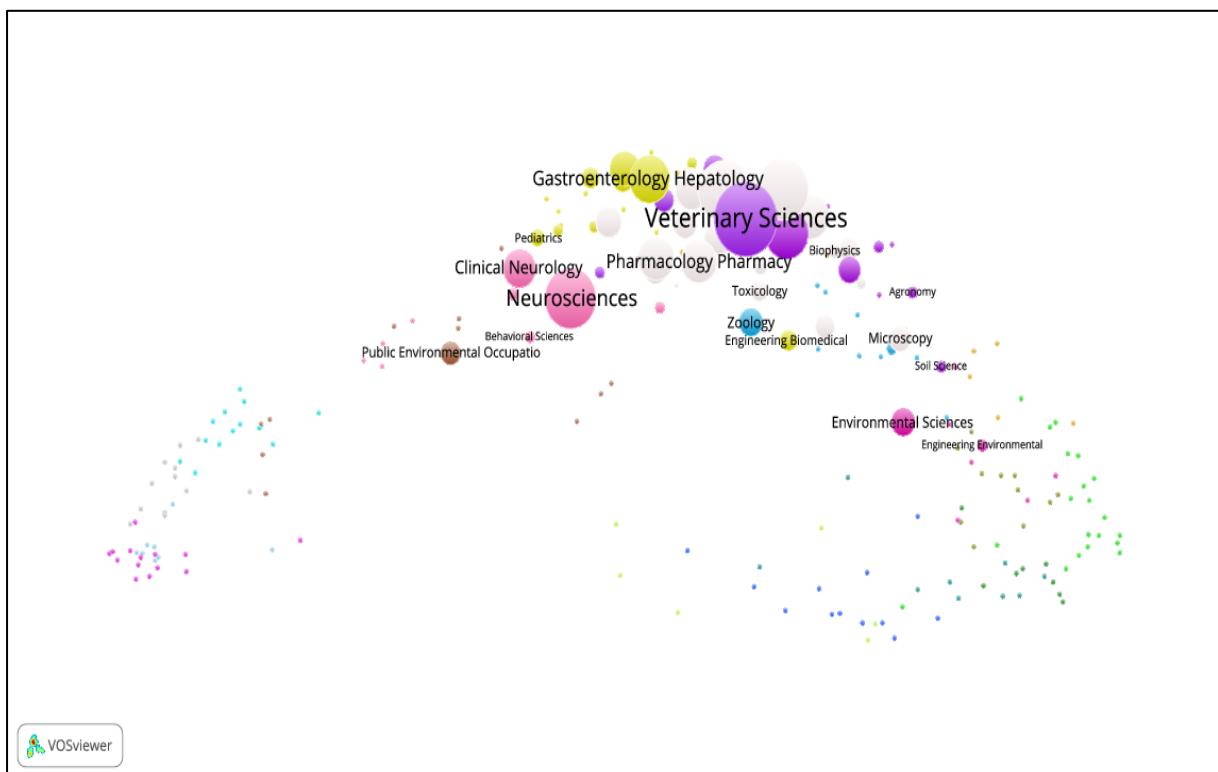


Figure 45. WoS SCs overlay map of Veterinary research groups' publications

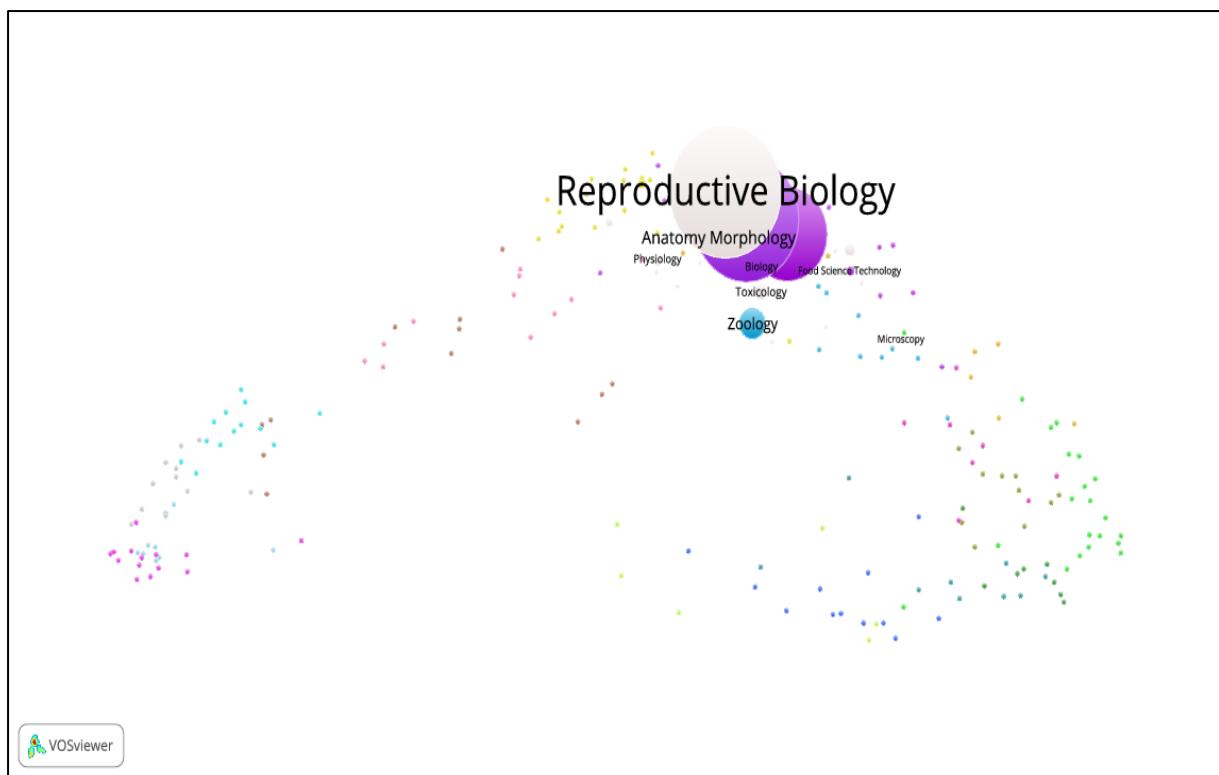


Figure 46. WoS SCs overlay map of PM1's publications

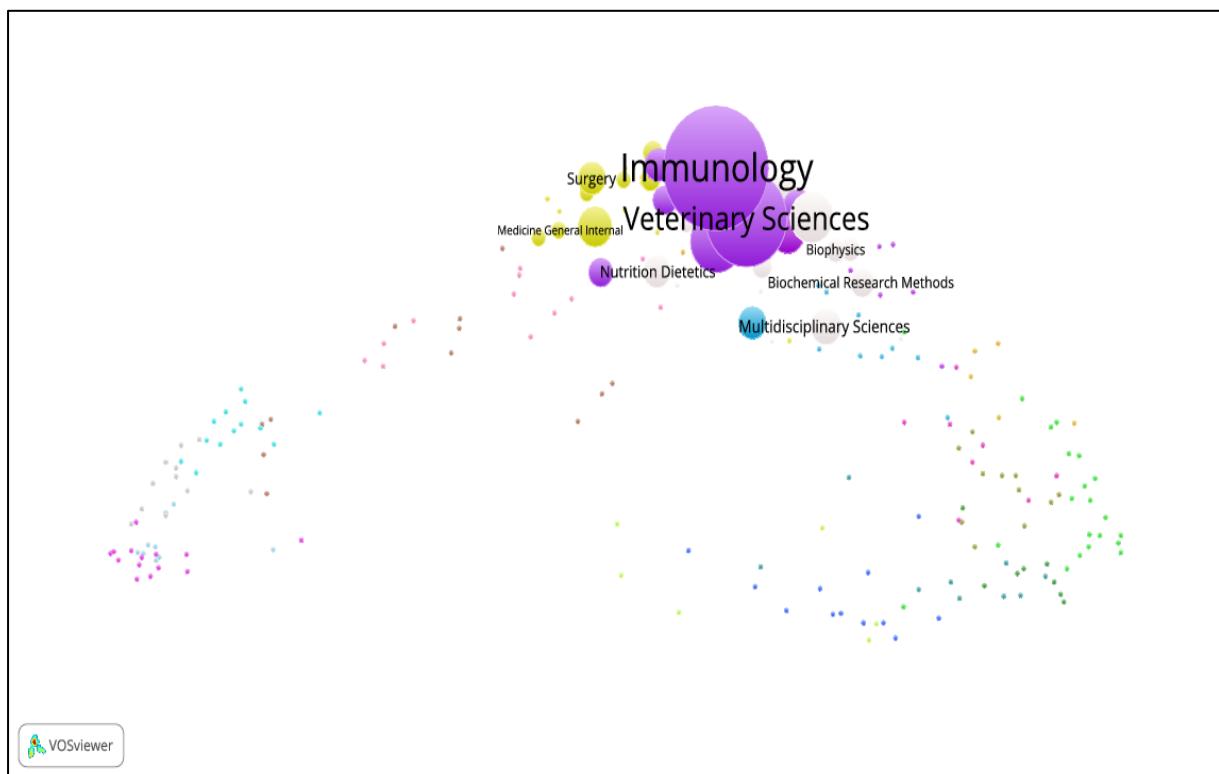


Figure 47. WoS SCs overlay map of PM2's publications

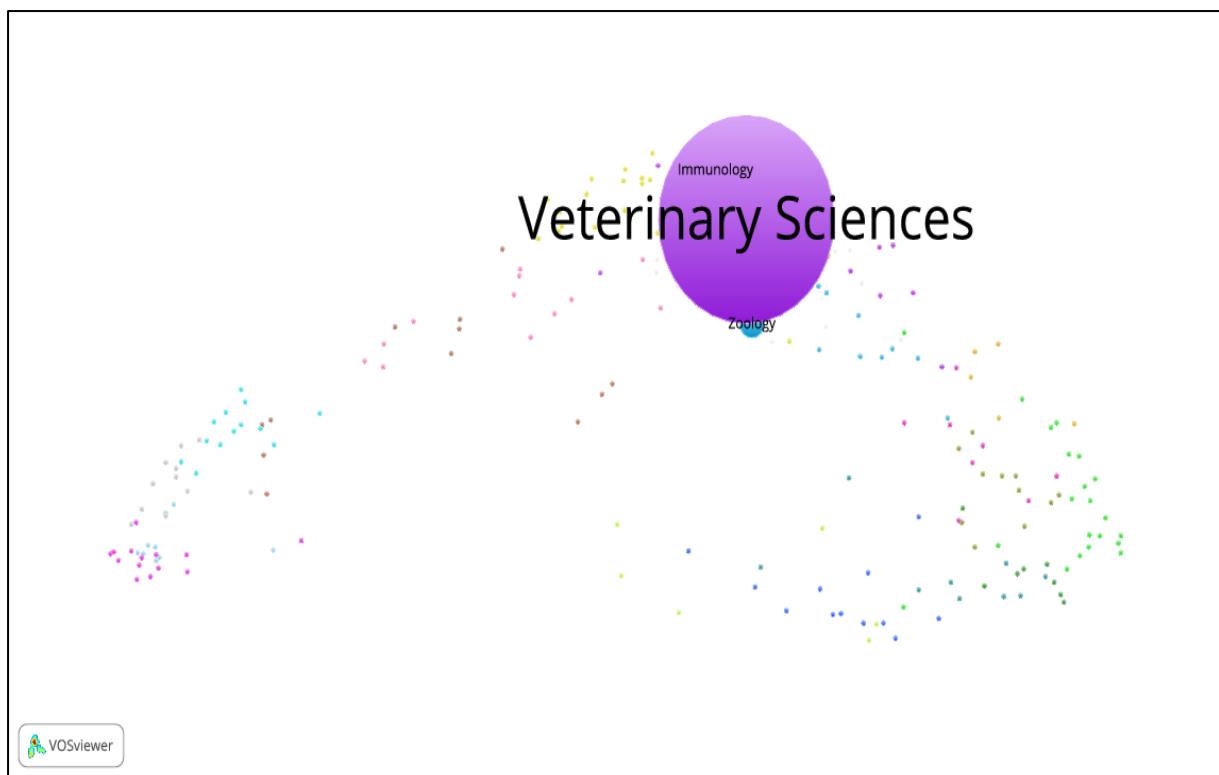


Figure 48. WoS SCs overlay map of PM3's publications

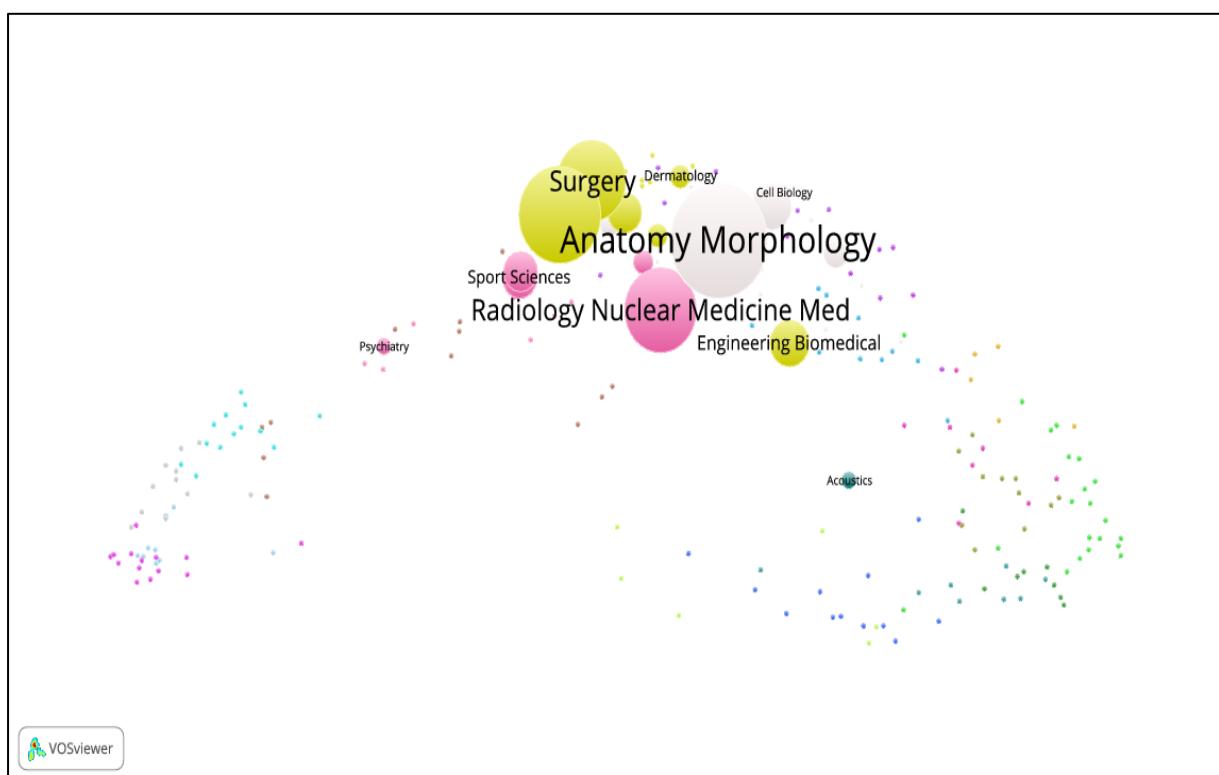


Figure 49. WoS SCs overlay map of PM4's publications

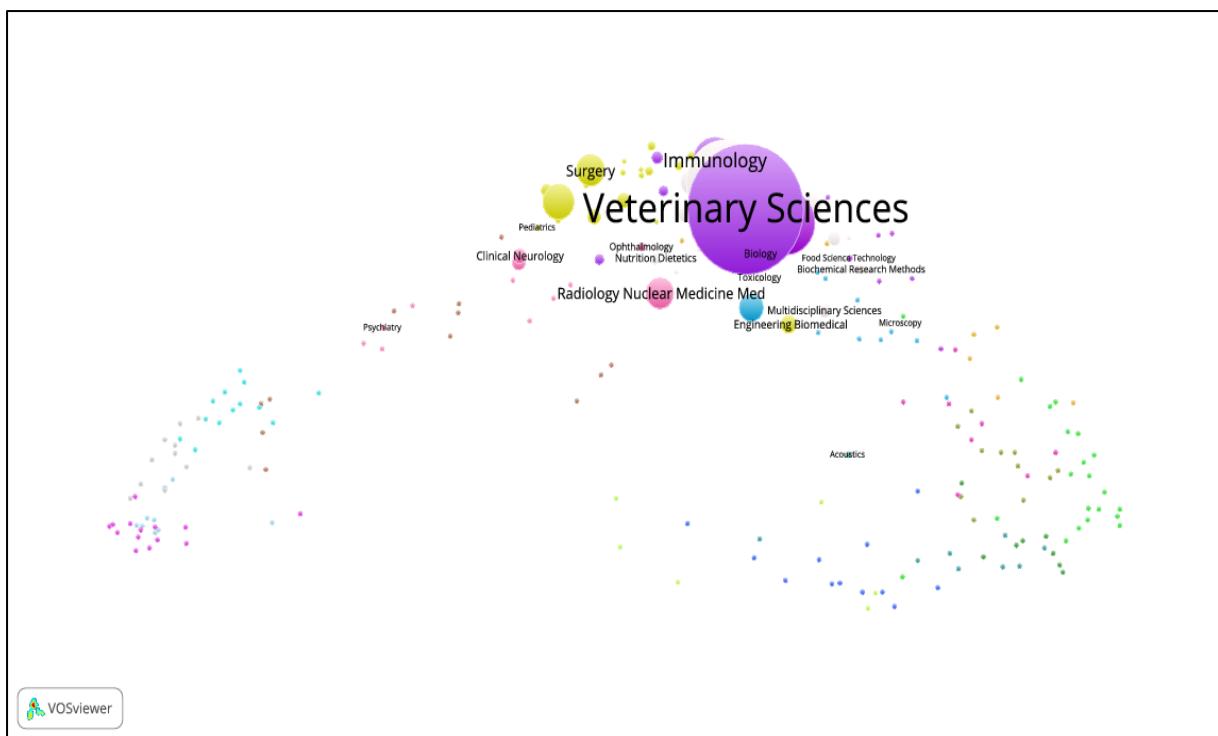


Figure 50. WoS SCs overlay map of panel's publications

Appendix B

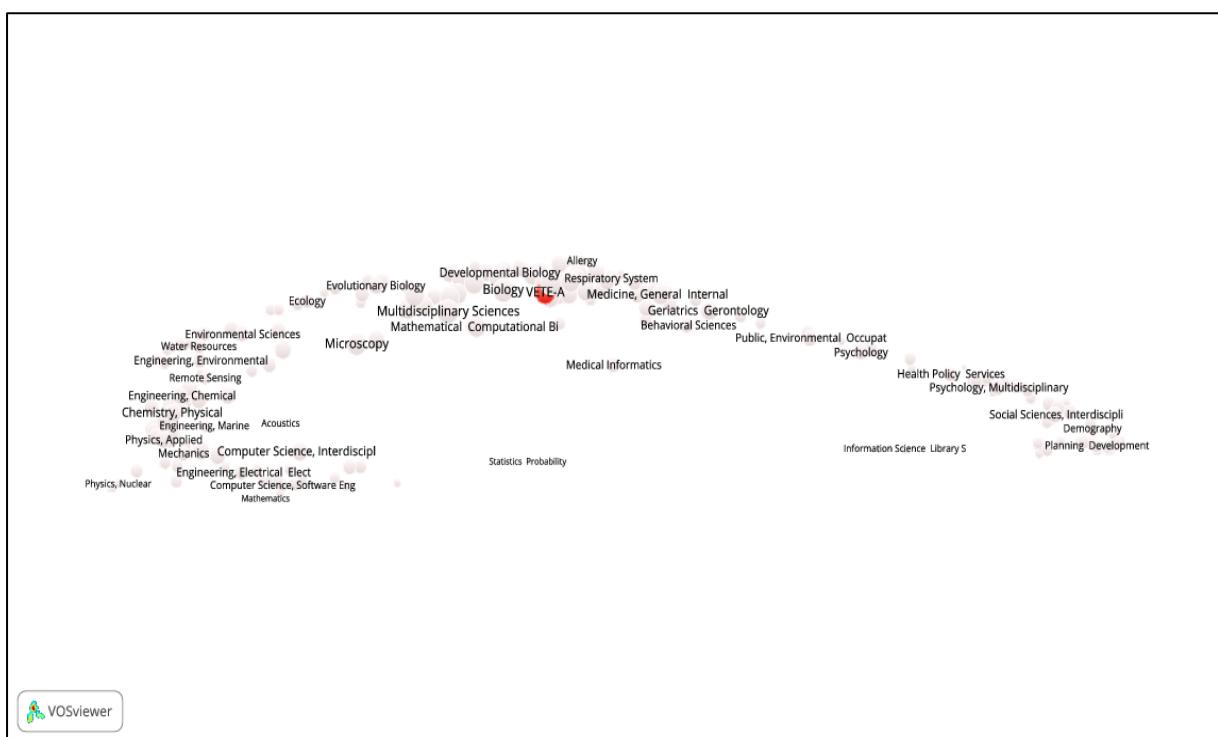


Figure 51. SAPV of the VETE-A research group's publications in WoS SCs similarity matrix

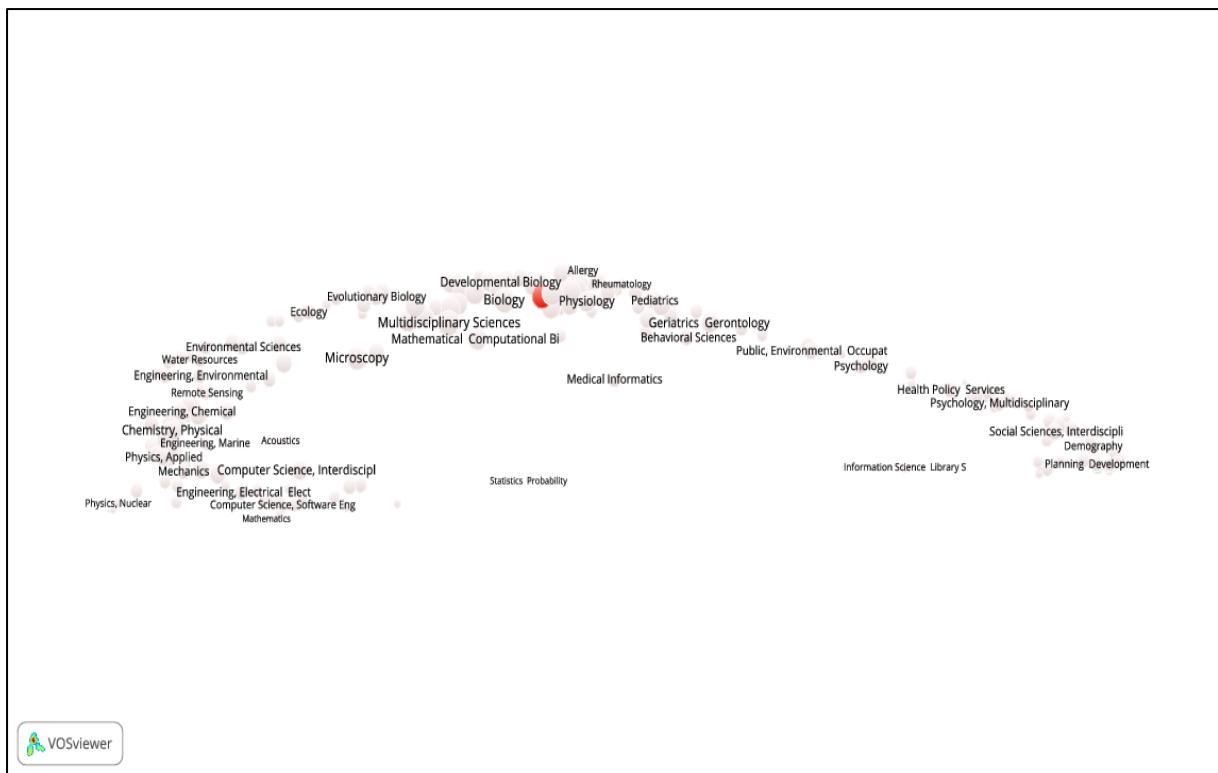


Figure 52. SAPV of the VETE-B research group's publications in WoS SCs similarity matrix

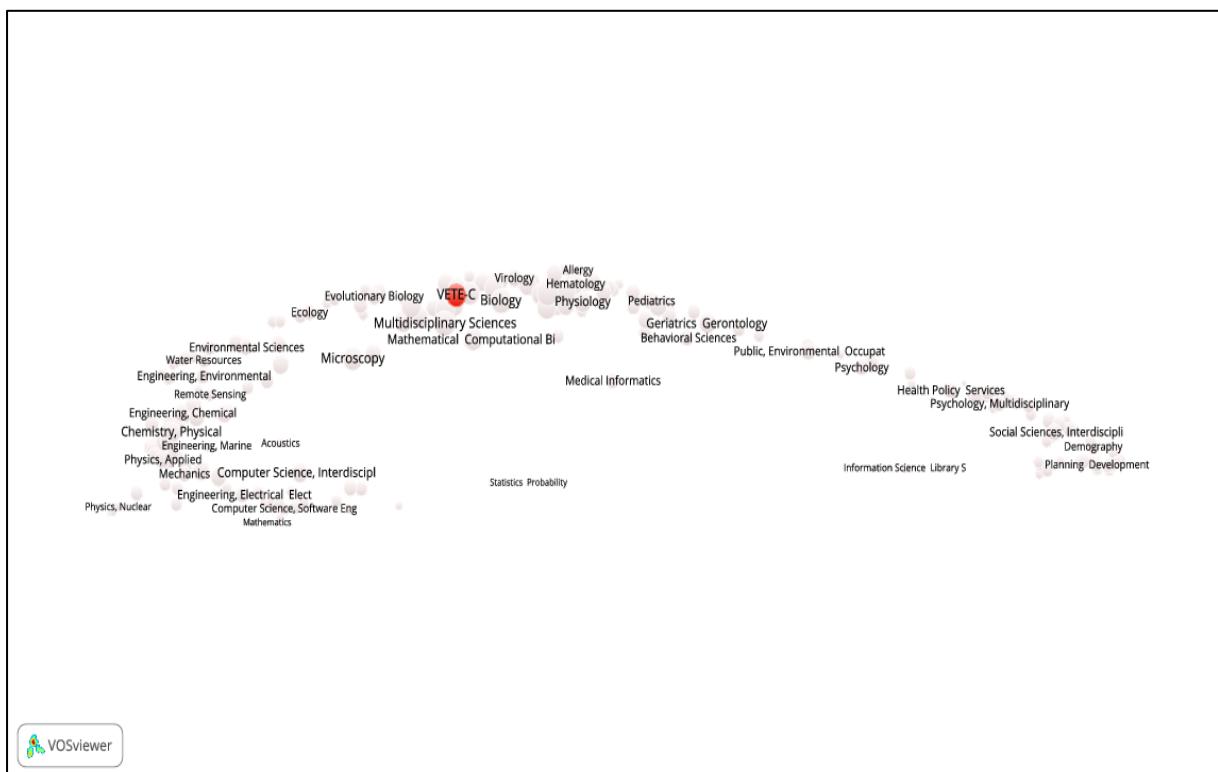


Figure 53. SAPV of the VETE-C research group's publications in WoS SCs similarity matrix

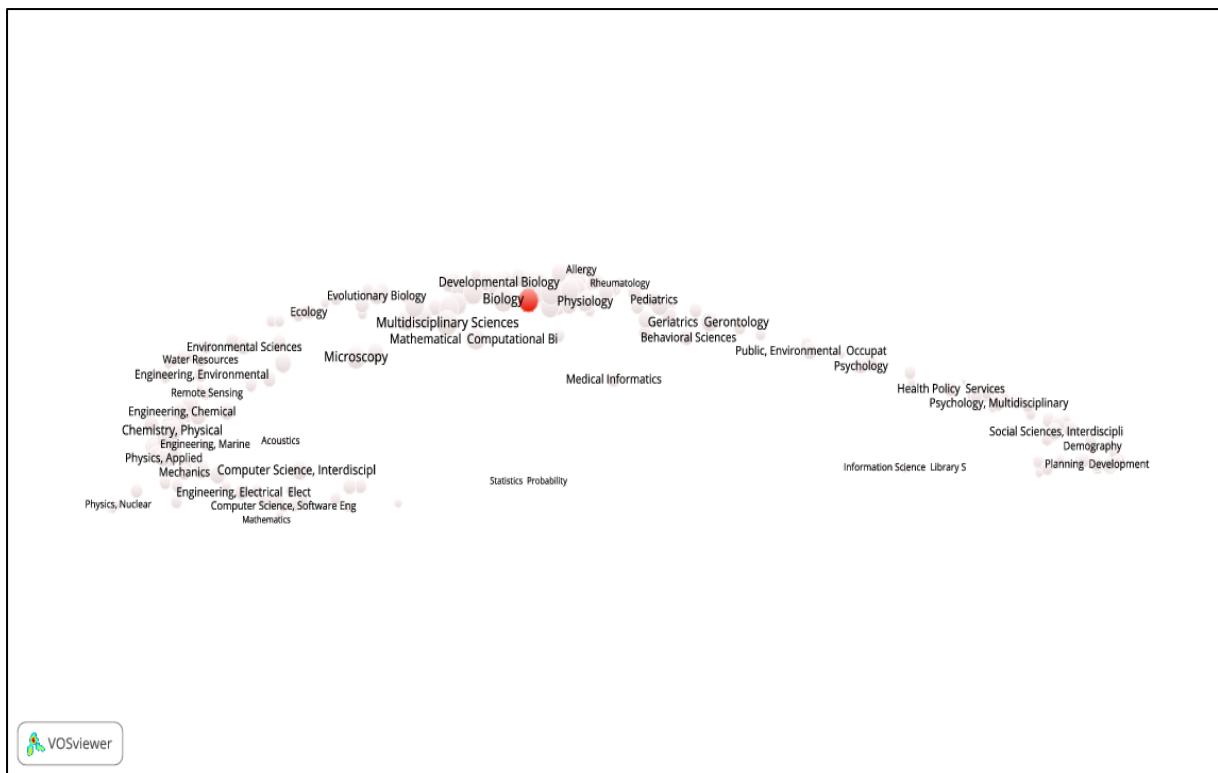


Figure 54. SAPV of the Veterinary research group's publications in WoS SCs similarity matrix

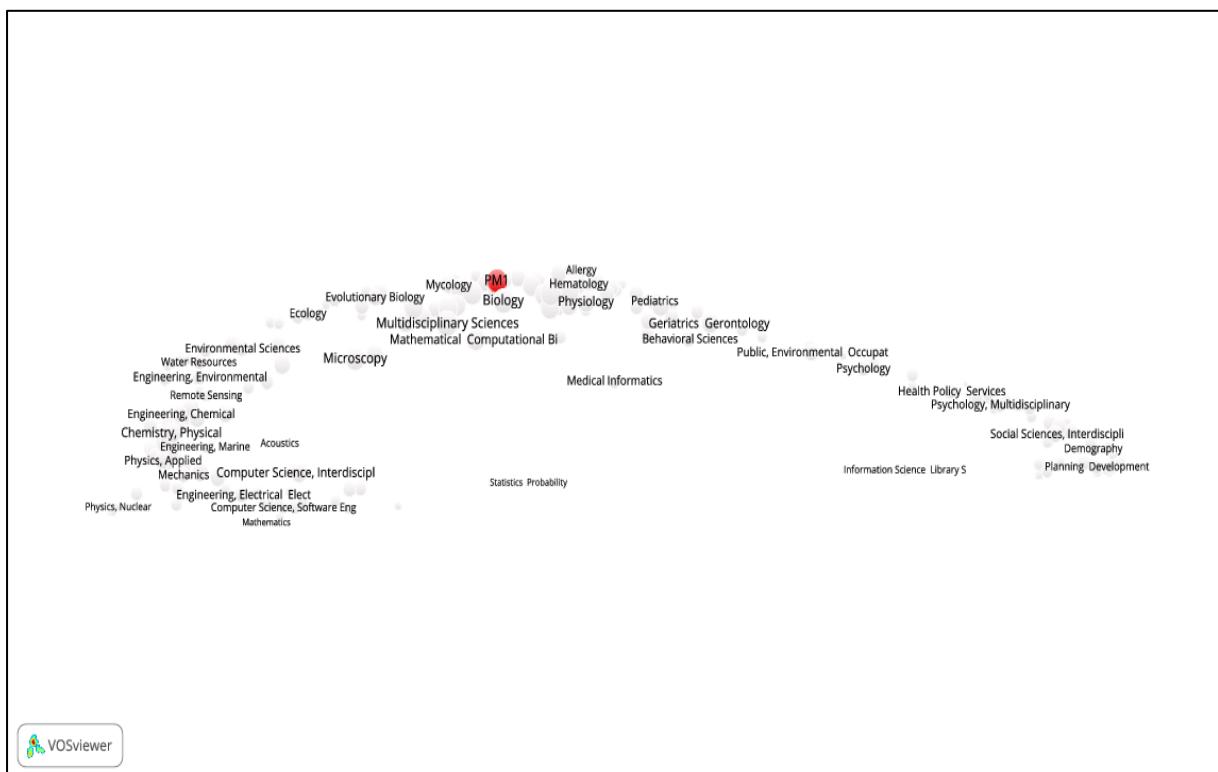


Figure 55. SAPV of the PM1's publications in WoS SCs similarity matrix

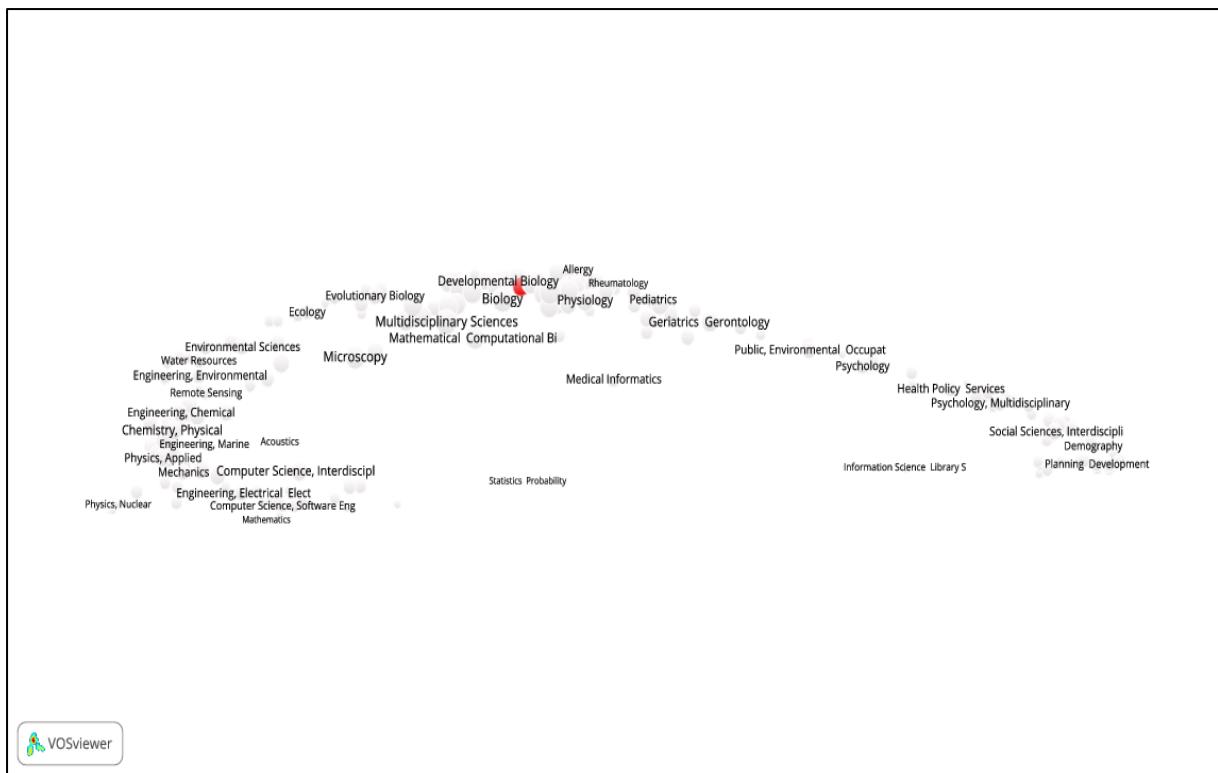


Figure 56. SAPV of the PM2's publications in WoS SCs similarity matrix

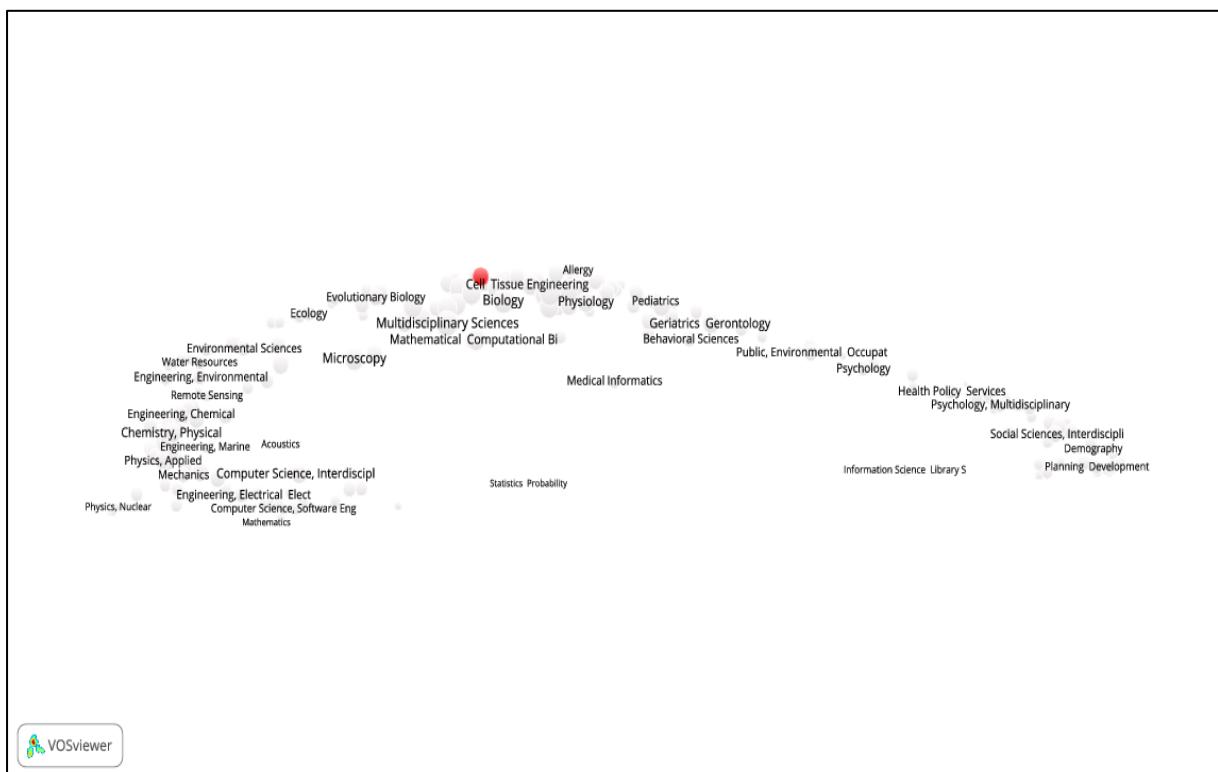


Figure 57. SAPV of the PM3's publications in WoS SCs similarity matrix



Figure 58. SAPV of the PM4's publications in WoS SCs similarity matrix

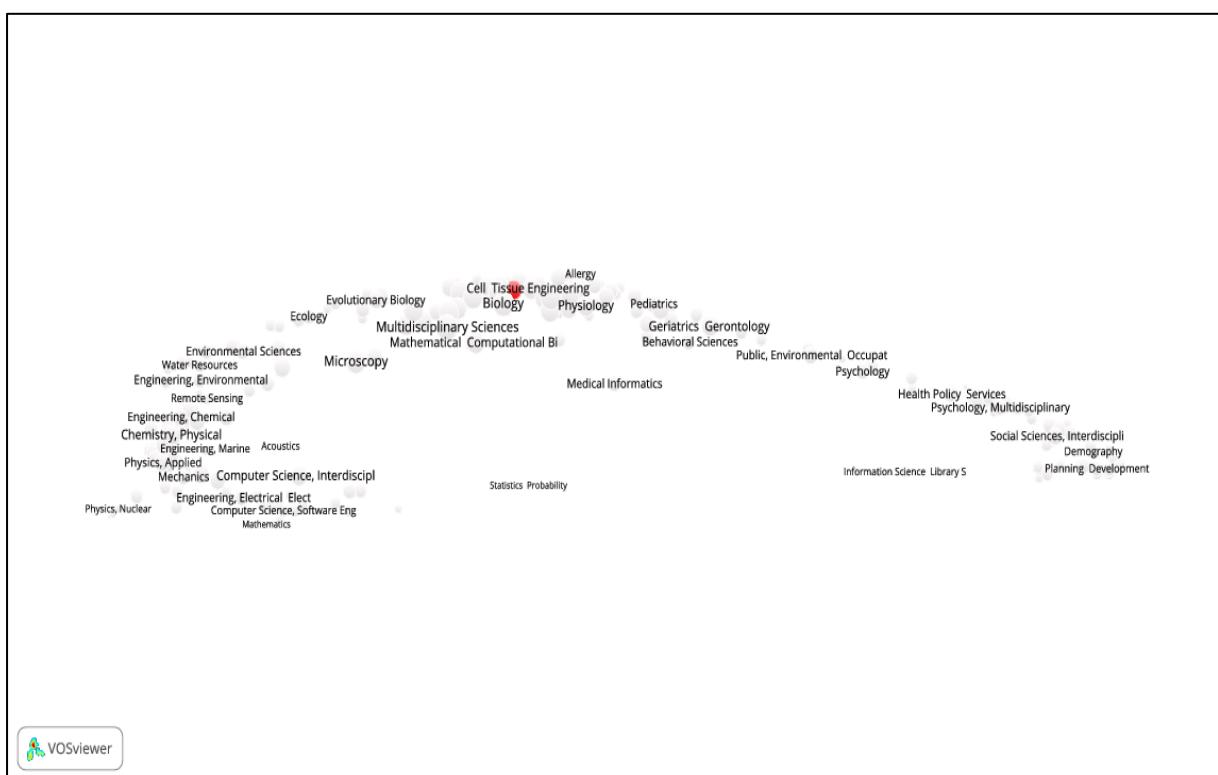


Figure 59. SAPV of the panel publications in WoS SCs similarity matrix

Appendix C

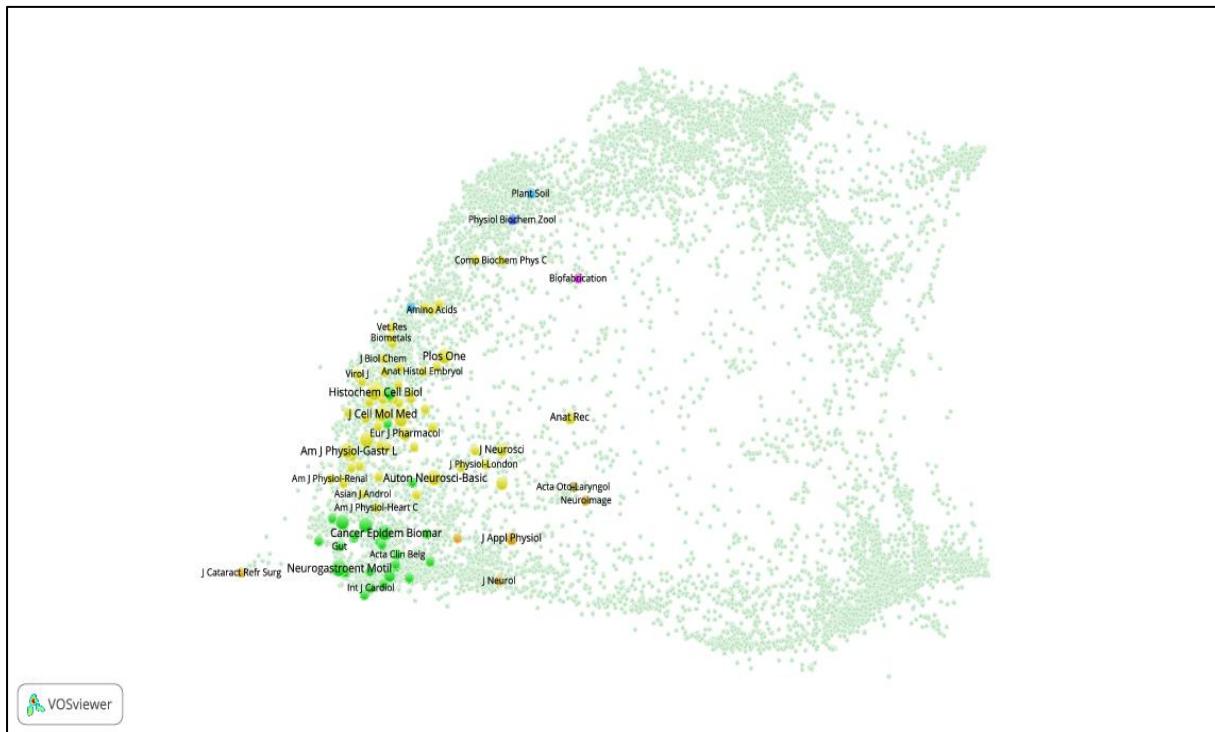


Figure 60. Journal overlay map of VETE-A research group's publications

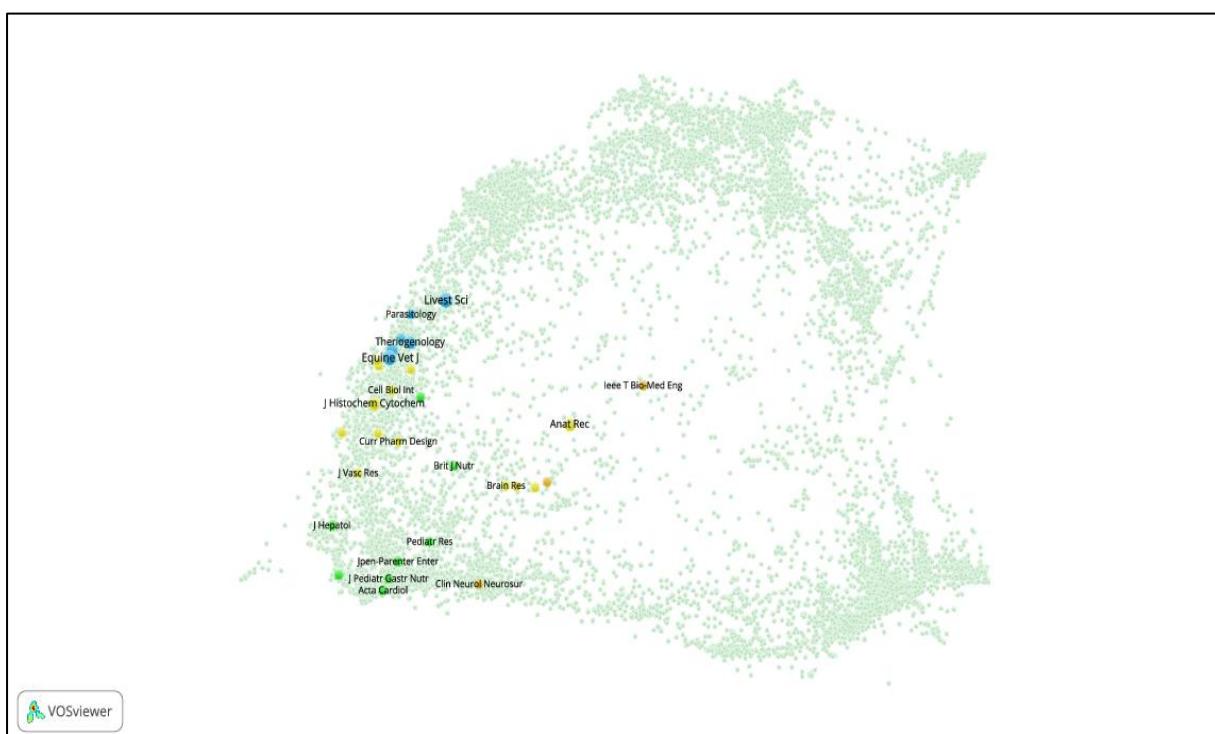


Figure 61. Journal overlay map of VETE-B research group's publications

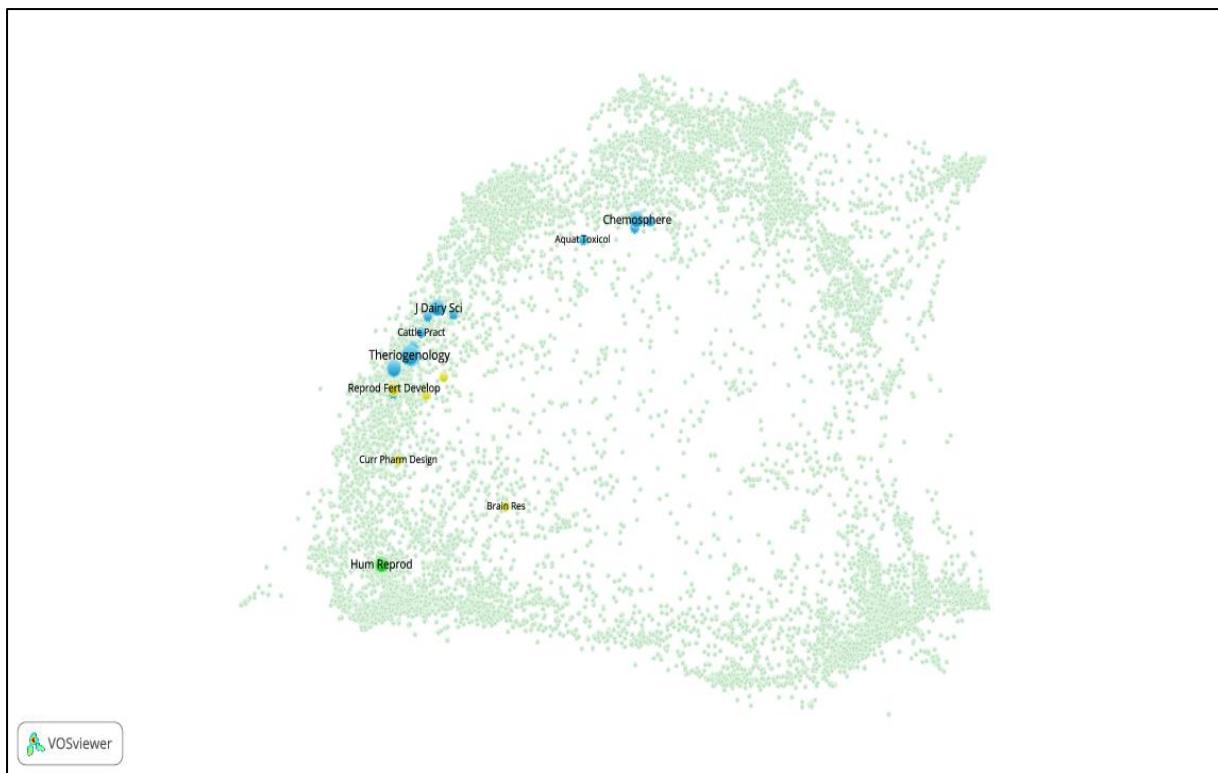


Figure 62. Journal overlay map of VETE-C research group's publications

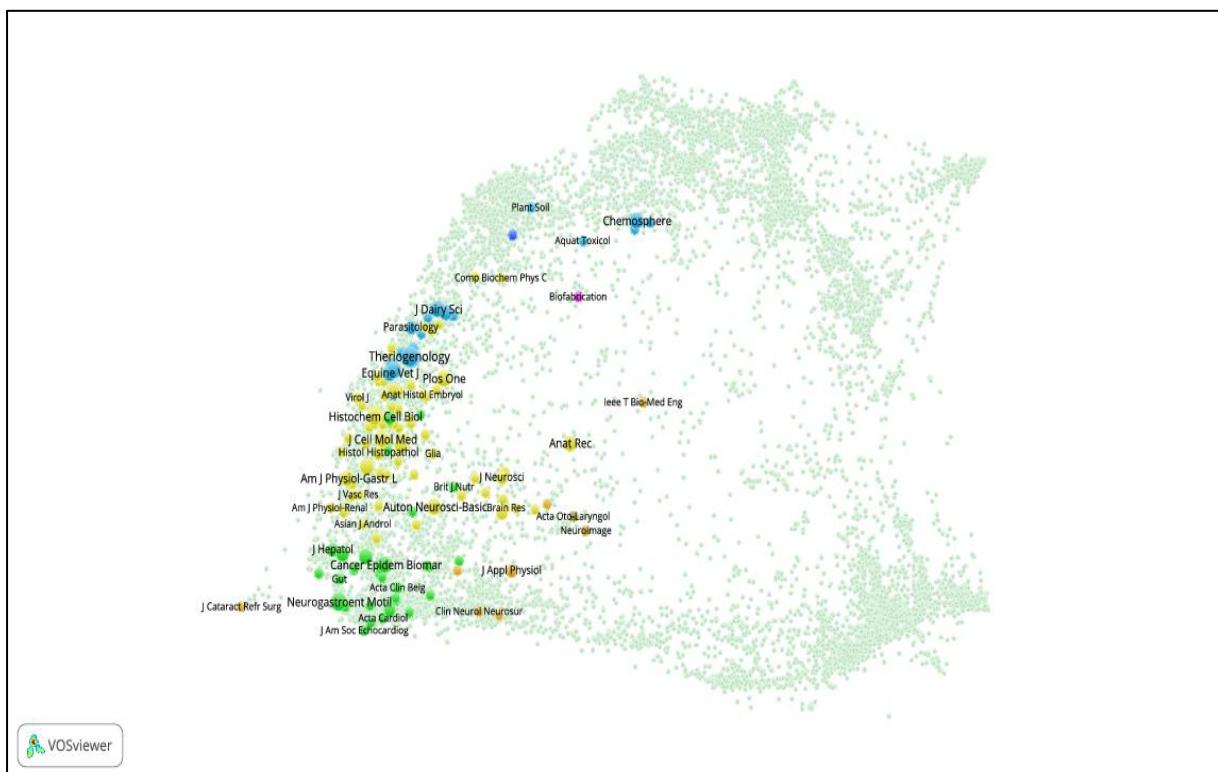


Figure 63. Journal overlay map of Veterinary research groups' publications

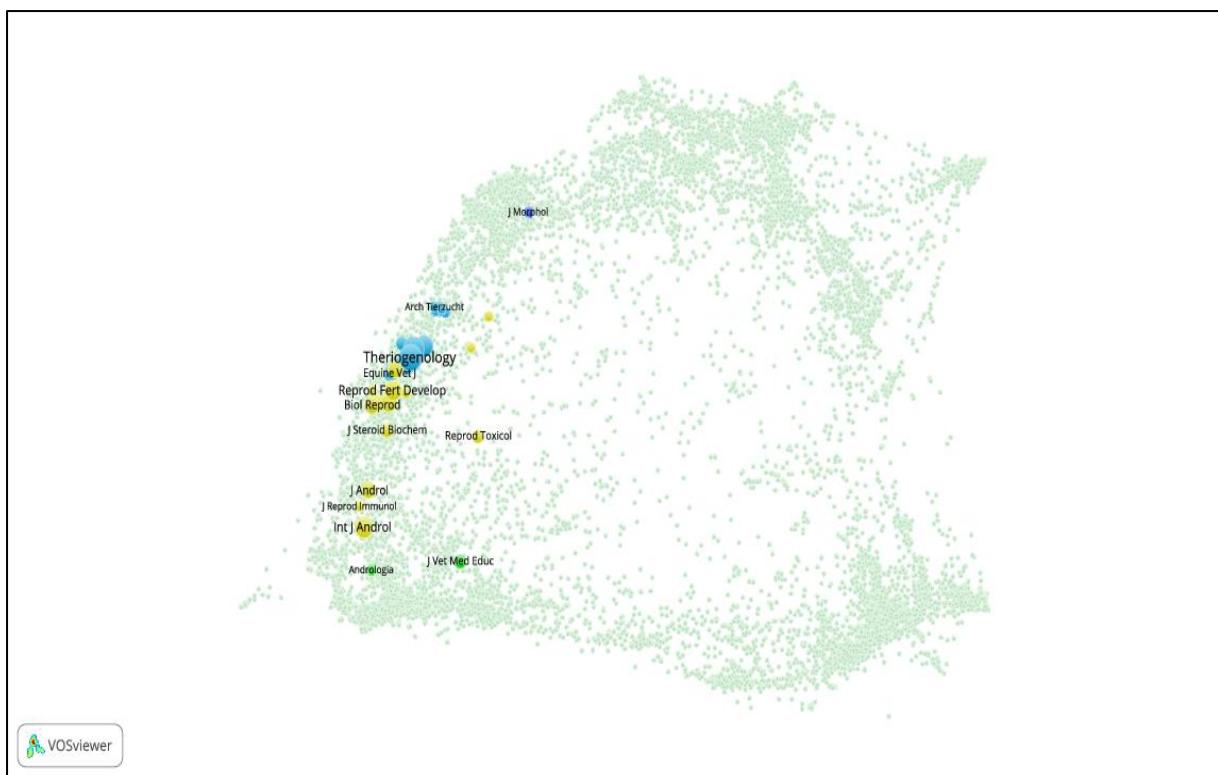


Figure 64. Journal overlay map of PM1's publications

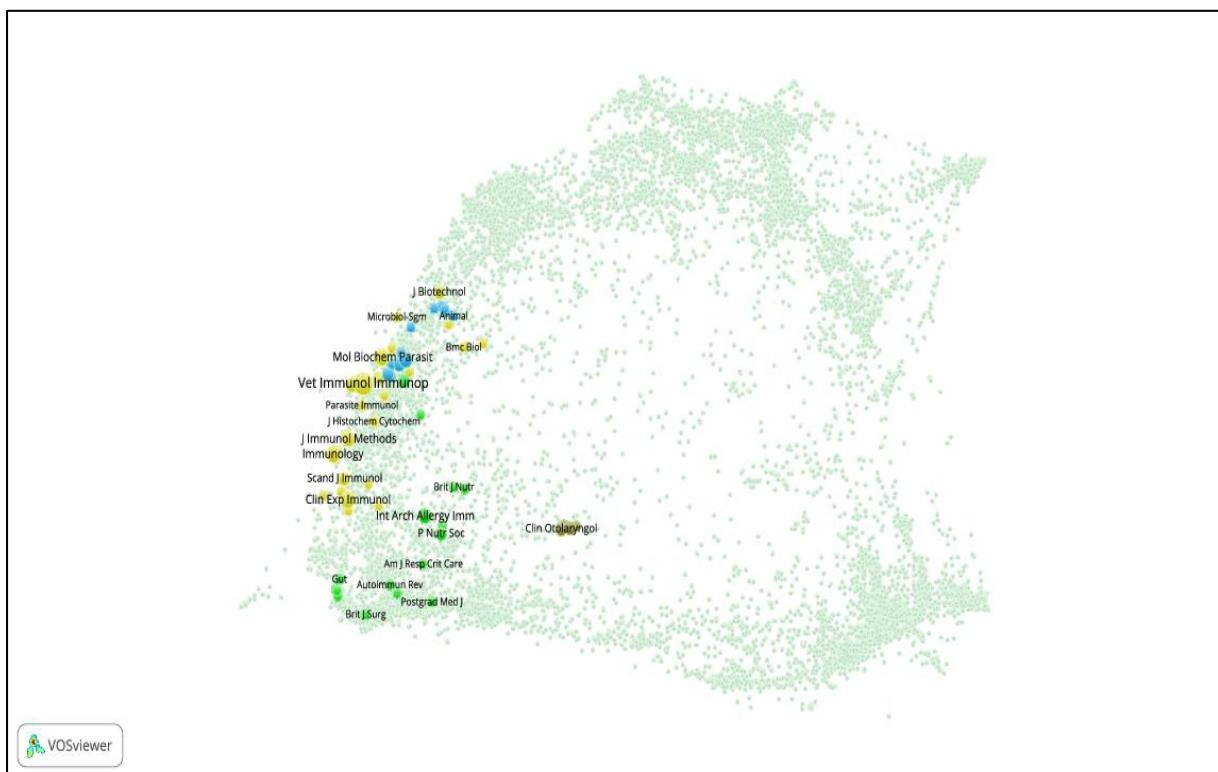


Figure 65. Journal overlay map of PM2's publications

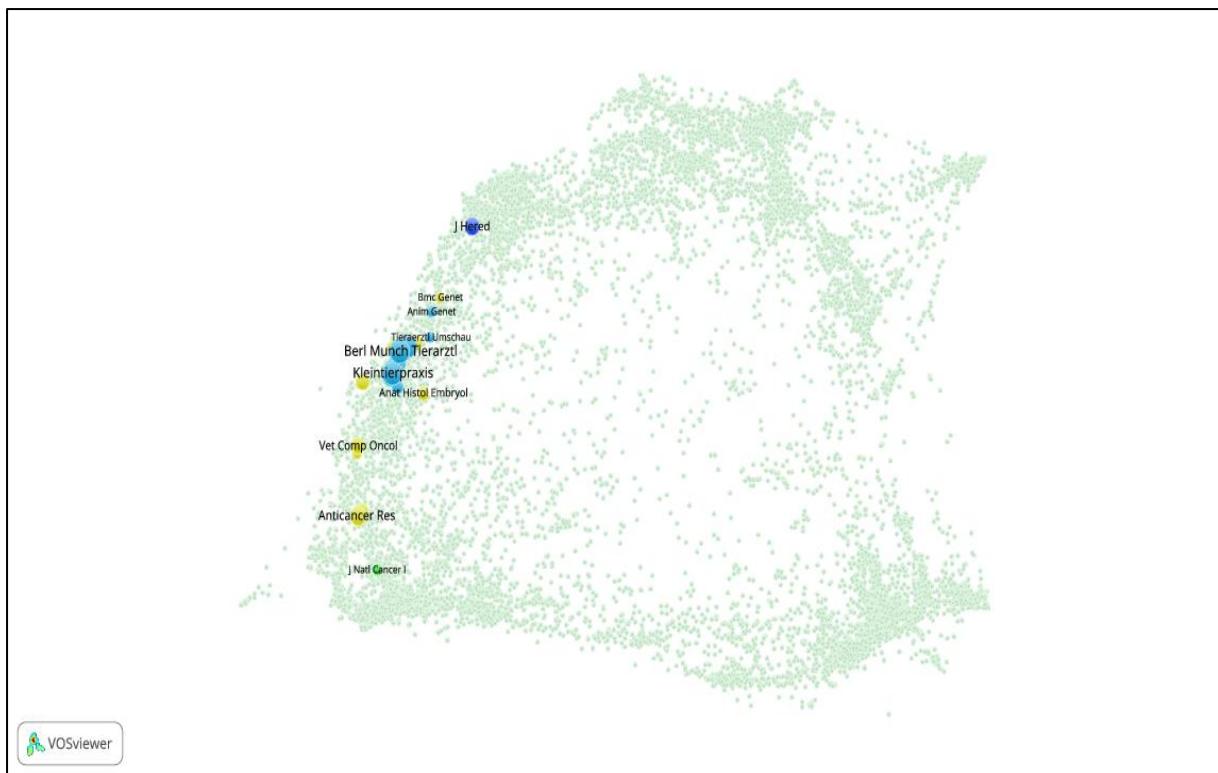


Figure 66. Journal overlay map of PM3's publications

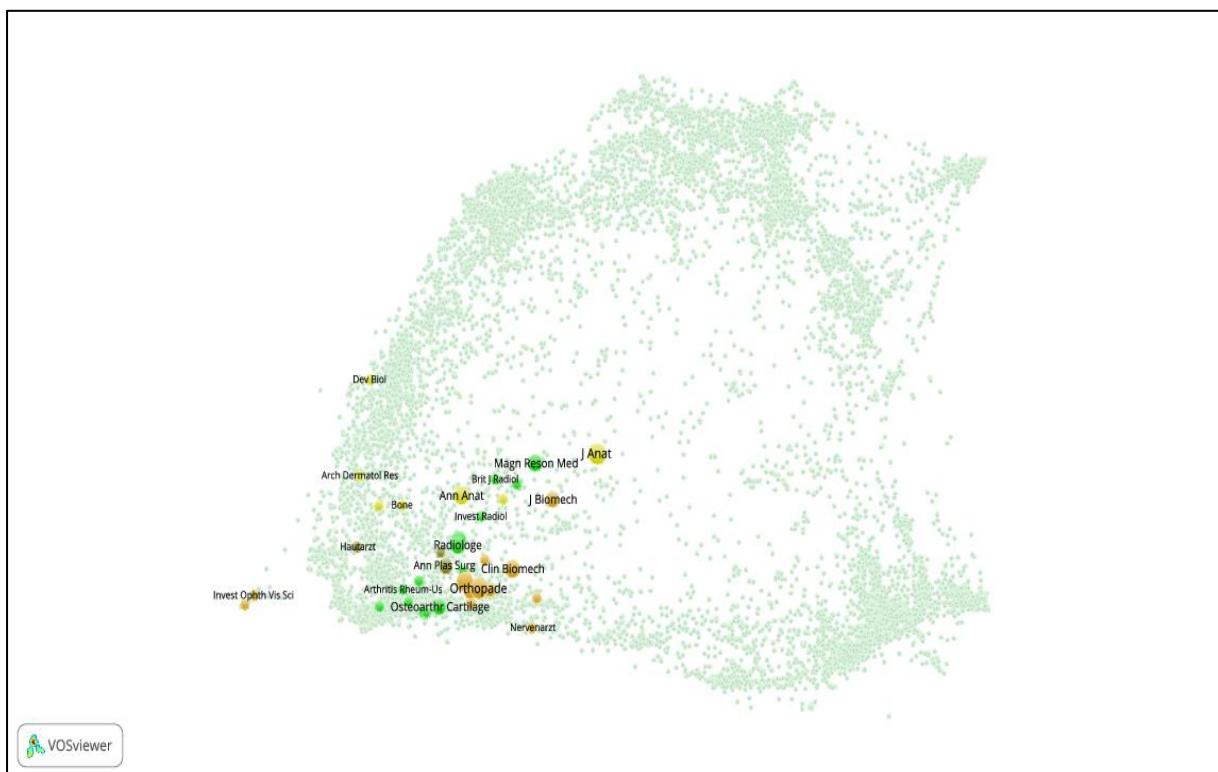


Figure 67. Journal overlay map of PM4's publications

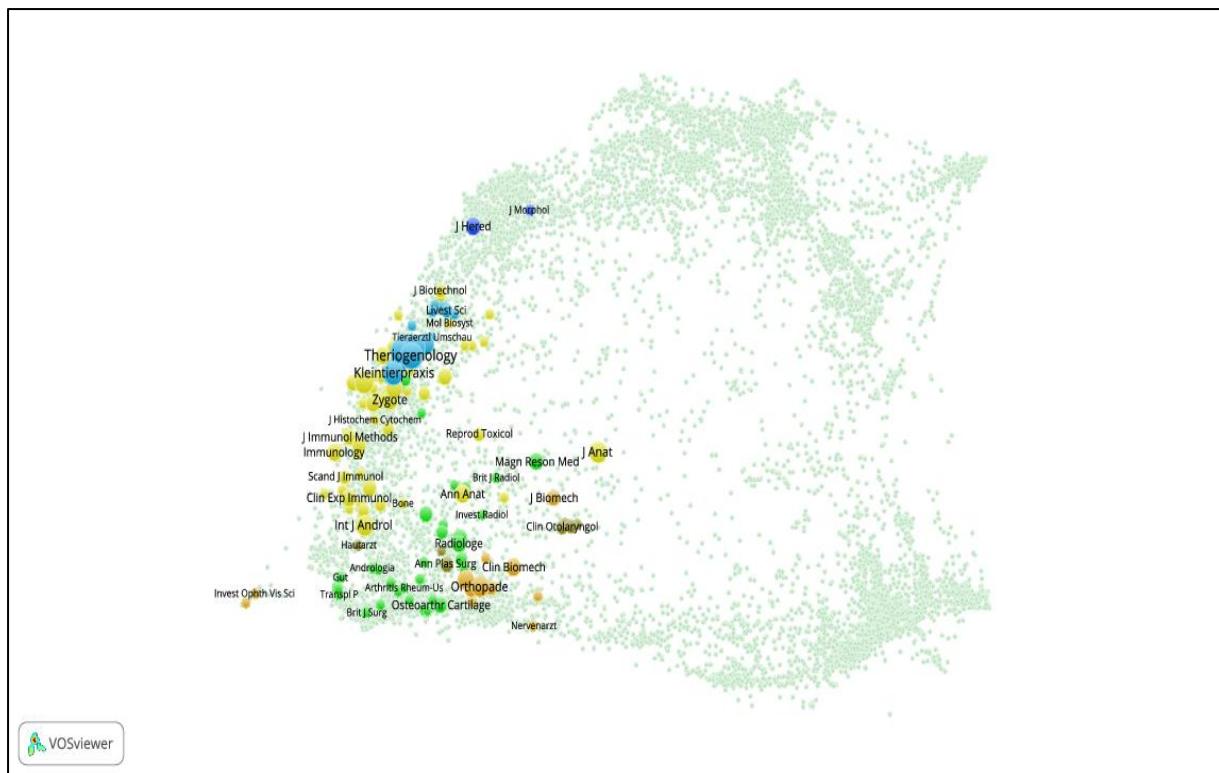


Figure 68. Journal overlay map of the panel's publications